Cloud Computing and Security Issues based on The Combination of Encryption: Survey Paper

[1]Karuna Raikwar [2]Sumit Nema
[1]Research Scholar, Department of Computer Science & Engineering
[2]Associate professor, Department of Computer Science & Engineering
Global Nature Care Sangathan's Group of Institutions, Jabalpur

***Abstract*—** Data Security and consumer data privacy are the key challenges in cloud computing era. The appropriateness and privacy of data stored in cloud may be compromised because of limited security for data owners. This paper presents an extensive survey on privacy preservation, data and storage security challenging issues in cloud computing. The Security of cloud data is further analyzed in terms of data integrity, access control and attribute based encryption. The survey analyzes each category of work in detail. A comparison table is also presented along with the strength and weakness of each approach.

Keywords*—*Cloud Computing, Data Security, Data Classification, File Splitting Security.

I. Introduction

Now a day's Cloud Computing Technology is the most promising technology comes in the real world. Users are more aware of the advantages of the cloud computing and they start using it. Cloud Computing is the next generation system which provides an easy and customizable way of managing data in the Internet. It provides the user various services of accessing and work with the application of Cloud. Users can upload their data in the Cloud Storage and can access through anywhere through any devices like laptop, mobile, desktop etc. Whenever the discussion of data comes some of the properties of data emerges and some of them are as follows Accuracy, Completeness and Consistency.

Computer in its evolution form has been changed multiple times, as learned from its past events. First from the beginning when mainframes (i.e. state-of-the-art computers for mission critical tasks, named in mid of 1960s referred to their main CPU cabinets) were predicted to be the future of computing. Indeed mainframes and large scale machines were built and used, and in some circumstances they are used similarly today. The trend, however, turned from bigger and more expensive, to smaller and more affordable commodity PCs and servers which are tired together to construct the so called Cloud Computing System, denoted as Cloud in short, due to their same capability in providing services, say storage, computation, and management and so on. Moreover, Cloud has advantages in offering more scalable, fault- tolerant services with even higher performance. Cloud computing can provide infinite computing resources on demand due to its high scalability in nature, which eliminates the needs for Cloud service providers to plan far ahead on hardware provisioning. As Cloud service providers can eliminate an up-front commitment, they are able to start from small companies and increase hardware resources only when there is an increase in need. Because of the capability of renting hardware from Cloud Computing providers, they are charged in terms of computing resources usage on a short-term basis (e.g., processors by the hour and storage by the day), and can release computing resources as they need, which is the so called utility computing. Therefore, some of the Cloud users enjoy the scalability of Cloud to provide services (e.g. Data as a Service (Daas), Software as a Service (SaaS), and Platform as a Service (PaaS)) to a larger amount of end users above Cloud Computing systems. Under this computing ecosystem, Cloud Computing is developing at an amazing pace.

Many companies, such as Amazon, Google, and Microsoft and so on, accelerate their paces in developing Cloud Computing systems and enhancing its services providing to a larger amount of users. As consisted of hundreds of thousands of commodity PCs and servers (say low prices), Cloud is more capable for competitions between companies. The successes of the above companies, say Google, Amazon and so on, are great examples and encourage an amount of other companies to step into the Cloud, such as MediaTemple, Mosso, Joyent, Flexicale, and so on. Lots of services, such DaaS, SaaS, PaaS, IaaS, etc. get into practice and provide to millions of users. On the other hand, more and more users are considering Cloud Computing is important and start to setup applications in the Cloud Computing system or adopt the services provided by it. According to a survey conducted by Forrester [1] over a large number of firms, which evaluate the importance of using Software as a Service (SaaS) in terms of their points of view, more and more firms are thinking it is important. Fig. 1 gives it in detail. 15% of the firms view it is important and another 5% of the firms consider it is very important. In the survey [1], it also claims that a typical organization today might have 5 to 15 applications in the Cloud. As Cloud Computing has advantages for both providers and users, it is developing in an amazing pace and predicted to grow and be adopted by a large amount of users in the near future [2]. Thus, Cloud Computing is becoming a well-known buzzword nowadays.

However, security and privacy issues present a strong barrier for users to adapt into Cloud Computing systems. According to an IDC survey in August 2008, which is conducted of 244 IT executives/CIOs and their line-of-business (LOB) colleagues about their companies' use of and views about IT Cloud Services, security is regarded as the top challenge of nine [3]. Fig. 2 shows the nine challenges in detail. Security is the top one concern, say users of Cloud Computing worry about their businesses' information and critical IT resources in the Cloud Computing system which are vulnerable to be attacked. Nevertheless, concerns on performance and availability are below the security. Furthermore, Cloud Computing becomes a hot topic at the RSA security conference in San Francisco in April 2009. Cisco CEO Chambers said that Cloud computing was inevitable, but that it would shake- up the way that networks are secured on that conference. Again, Forrester [1] also said data protection, operational integrity vulnerability management, business continuity (BC), disaster recovery (DR), and identity management (IAM) are top concerns of security issues for Cloud Computing and privacy is another key concern. Security and privacy of Cloud Computing system become a key factor for users to adapt into it.

- Moreover, many security and privacy incidents are also observed in today's Cloud Computing systems. We list a few of them blew:
- Google Docs found a flaw that inadvertently shares user's docs in March 2009.
- A Salesforce.com employee fell victim to a phishing attack and leaked a customer list, which generated further targeted phishing attacks in October 2007.
- Epic.com lodged a formal complaint to the FTC against Google for its privacy practices in March 2009. EPIC was successful in an action against Microsoft Passport.
- Steven Warshak stops the government's repeated secret searches and seizures of his stored email using the federal Stored Communications Act (SCA) in July, 2007.

II. Literature Review

We all know that cloud storage provides various advantages like better accessibility means make data to access from anywhere using internet. No need for the users to depends upon the physical storage for carry and retrieving of data.

Ji Hu, Andreas Klein [1], focus on the security of the data in the place of use of data. Here they make a Benchmark which analyze the basic requirement of applications in the cloud for the security of the e-commerce data. In this different approach of the encryption is discussed for securing the data during transactions as the data in the transaction is critical.

In Cloud computing preserving integrity of the data is the most important challenge, method for checking the integrity of the data without download the data. In normal the user has to download the data and check after that they get the doubt of the illegally modification of the data. But they doesn't have any evidence to proof for the same. An approach is made which know the integrity of the data without downloading the data [2].

Ashish Singh, Kakali Chatterjee [3], according to the characteristics of the cloud, the user should not to know the location of the data and which security is used for securing the data. Not only the outer attacks, insider attacks are also be taken care. For avoiding the insider attack, CSP uses authentication scheme. Here in this paper new Two-Tier Authentication is made which is more secure and provide advancement in the security in cloud.

We all know that encryption is the basic security mechanism in networking and AES-512 is the most secured algorithm which can provide high security but due to its time consumption and more computations, it is not in use in everywhere, but Abidalrahman Moh'd, Yaser Jararweh, Lo'ai Tawalbeh [4] , proposed a new Advanced AES-512 Encrytion algorithm which makes the data more secure to attacks. In these, they use 512 key size as well as 512 bit block size. Using their new algorithm, AES-512 can easily be used anywhere where high security to data is required.

Cloud Security Alliance (CSA) prepared a document in 2010, [5] in which they discuss the threats which are harmful to the data present in the cloud. Some of the issues they figure out are Corrupt and Abuse use of Cloud Computing, Insecure Programming Interface, Insider Attacks, Vulnerabilities in shared environment, Data Loss and Leakage and so on. In the Cloud Storage environment, all the data present in the cloud storage are get encrypted with single security algorithms without considering the whether it requires or not. So Rizwana Shaikh, Dr. M. Sasikumar, [7] proposed a technology which classifies the data into different category based upon the properties of the data like Access Control, Content, Storage etc. After the classification they provide the security accordingly, by using permissions to data like Restricted, Moderate and Mandatory.

While implementing the Cloud environment, the Service providers faces many issues regarding the security and other aspects. R. Velumadhava Raoa, K. Selvamani [8],shows the major security challenges which the services providers has to face during the implementation of the cloud computing platform are Integrity of Data, Security, Locality, Access, Breaches, Segregation and also some solutions for the issues like Encryption, Calculation of Hash Value etc.

Microsoft Trustworthy Computing, Frank Simorjay [9], the document which declared only for the information regarding Cloud computing. In this document it is discussed that the data in the cloud exists in 3 states: at Rest, at Process and at Transmit. And according to the document the data should remain status in all three states. Security to the data also depends on the states of data.

In cloud computing there are many challenges the providers face, 2 important challenges are Confidentiality and Privacy. In [10], the author provide the solution for the above two problems, a framework proposed by the author where different tasks are executing before the actual use of the data by the user. Task like Key Management, Encryption Mechanism, Verification Mechanism at client side and at server side tasks are Authorization Mechanism and Integrity Mechanism.

Efficient Cloud Storage Confidentiality to Ensure Data Security: L. Arockiam et al[11], we know, for the protection of the data in the cloud Encryption is used but sometimes only encryption doesn't provide high security so, author proposed new technique for providing more security. Here in this paper, two technique used Encryption and Obfuscation. Encryption converts the readable text to unreadable form while Obfuscation makes the numeric content confusing.

High-Throughput Encryption for Cloud Computing Storage System: Yaser Jararweh et al[12] , various methods and algorithms were proposed to make the data secure efficiently. In this paper a Symmetric Block Algorithm (CHiS-256) proposed which encrypt the which encrypt the cloud data efficiently and metadata based Cloud Storage which stores the data into small parts. There are various security algorithm presents but only few are used for the high security. Some of the security algorithms are as follow [13] RSA, SHA1 and MD5. In cloud computing data first get encrypted with security algorithms and then uploaded in the cloud storage. In cloud storage the encrypted data is stored and here author compares three algorithms in terms of security, efficiency and other factors like using different key size to judge the performance of the algorithm in best key size.

A Secure Cloud Computing Model based on Data Classification: Lo'ai Tawalbeh et al[14], here they find out the problem that treating all the data in the same manner and providing same level of security. As a solution for the above problem they proposed a framework which classify the data into different categories like Basic, Confidential and High Confidential and providing the different security techniques according the requirement they provide that type of security to that category.

Security Issues in Cloud Computing: A Survey, Rizwana Shaikh et al[15], this paper is the survey paper which surveyed the security issues in cloud computing. They proposed the different security concerns, what are the security issues the Client and the Providers facing in the Cloud Computing. For some of the issues they also focus on the solutions to be taken.

Anup Mathew [16], proposes a survey paper focus on the various Security, Data Privacy and Data Confidentiality Issues. Here in the safety and privacy concern covers like Location Independent Services, Communication and Infrastructure etc. also classify the security issues into two category: Access Security and Service Security.

Trust Model for Measuring Security of Cloud Computing Service: Dr. M. Sasikumar et al[17], due to increase in the Cloud Computing Technology, now a day's there are various CSP's in the market. There are some of the tools and measurements presents which the providers are using for measuring the security of their services. This paper presents a measurement for cloud service. This measurement is based on the trust model. The trust model calculates a value called Trust Value and this value comprises of many parameters like Identity, Management, Authentication, Authorization etc.

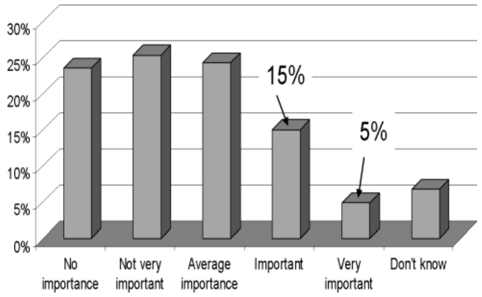
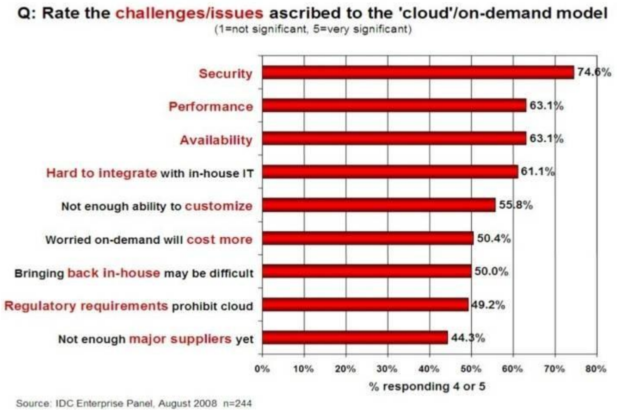


Fig. 1.Rating the Importance of Using SaaS in Terms of firms' Points of view



Source: IDC Enterprise Panel, August 2008 n=244

Fig. 2. Rate the Challenges/Issues Ascribed to the Cloud On-demand Model

However, the government argues the Fourth Amendment doesn't protect emails at all when they are stored with an ISP or a webmail provider like Hotmail or Gmail. That's to say, many Cloud Computing systems in the real world do have security and privacy problems. In this paper, we investigate the security and privacy concerns of current Cloud Computing systems provided by an amount of companies. As Cloud Computing referred to both the applications delivered as services over the Internet and the infrastructures (i.e., the hardware and systems software in the data centers) that provide those services [2], we present the security and privacy concerns in terms of the diverse applications and infrastructures. Based on the investigation to those Cloud Computing systems, we find that the security and privacy concerns provided by companies nowadays are not adequate, and consequently result in a big obstacle for users to adapt into the Cloud Computing systems. Hence, more concerns on security issues, such as availability, confidentiality, data integrity control, audit and so on, should be taken into account. New strategies are able to be deployed into Cloud Computing systems to make them even more secure. We present a few such strategies in terms of the five aspects in the Cloud Computing literature. Cloud Computing system usually has a special relationship between users and providers (i.e. three parties), which will be introduced in Section III. The special relationship results in many privacy protection acts not applicable in the Cloud Computing scenarios. We investigate a few privacy acts to illustrate that they are out of date. Data storage in the Cloud Computing system which is located in multi regions (locations) to make the system more tolerant may also raise the privacy problems. We present a brief introduction to this latter. We admit the prosperity of Cloud era do to be coming after those issues on security and privacy being resolved.

The rest of the paper is organized as follows. Security concerns should be added to the current systems, which are presented in Section II, while those privacy concerns are presented in Section III. We conclude our paper in Section IV.

III. SECURITY ON DEMND

Cloud services are applications running somewhere in the Cloud Computing infrastructures through internal network or Internet. For users, they don't know or care about the data where to be stored or services where to be provided.

Cloud computing allows providers to develop, deploy and run applications that can easily grow in capacity (scalability), work rapidly (performance), and never (or at least rarely) fail(reliability), without any concerns on the properties and the locations of the underlying infrastructures. The penalties of obtaining these properties of Cloud Computing are to store individual private data on the other side of the Internet and get service from other parties (i.e. Cloud providers, Cloud service providers), and consequently result in security and privacy issues. Then, what kind of security is sufficient for users? Basically, we say the Cloud Computing systems are secure if users can depend on them (i.e. DaaS, SaaS, PaaS, IaaS, and so on) to behave as users expect. Traditionally, it contains 5 goals; say availability, confidentiality, data integrity, control and audit, to achieve adequate security. The five goals are integrated systematically, and none of them could be forfeited to achieve the adequate security. Nevertheless, few Cloud Computing systems can achieve the five goals together nowadays.

A. *Availability*

The goal of availability for Cloud Computing systems (including applications and its infrastructures) is to ensure its users can use them at any time, at any place. As its web-native nature, Cloud Computing system enables its users to access the system (e.g., applications, services) from anywhere. This is true for all the Cloud Computing systems (e.g., DaaS, SaaS, PaaS, IaaS, and etc.). Required to be accessed at any time, the Cloud Computing system should be severing all the time for all the users (say it is scalable for any number of users). Two strategies, say hardening and redundancy, are mainly used to enhance the availability of the Cloud system or applications hosted on it.

Many Cloud Computing system vendors provide Cloud infrastructures and platforms based on virtual machines. For example, Amazon Web Services provide EC2, S3 entirely based on the virtual machine called Xen [4], and Skytap [5] offers virtual lab management application relaying on hypervisors, including VMware [6], Xen and Microsoft Hyper-V [7], and so on. Let's take the virtual machine Xen provided by Amazon as an example, it is capable in providing separated memory virtualization, storage virtualization, CPU/machine virtualization and etc., which are hosted on a large number of commodity PCs. That is the reason why Cloud service provider can rend resources (e.g., CPU cycles, storage capacity, memory) from Amazon on demand at the expense of usage in terms of a single unit. Hence, the virtual machine is the basic component to host these services. Fig. 3 shows this service provision architecture in detail. As shown in Fig. 3, hosted services reside on the virtual machine, which is combined with or shared from a set of CPUs, memory, storage on demand, and is regarded as services' infrastructures or platforms running on. Clearly, virtual machines have the capability in providing on demand services in terms of users' individual resource requirement for a large amount of users. In certain sense, users can use them as on-premises systems and can upgrade at any time they want to. On the other hand, Cloud system vendors depends on the virtual machine to tie commodity personal computers or servers together and to provide a scalable, robust system. Thus, this virtual machine is always available to use. Furthermore, current Cloud system vendors who are providing infrastructures and platforms based on virtual machine (e.g. Amazon, Skytab) offer the ability to block and filter traffic based on IP address and port only to secure their systems, but these facilities are not equivalent to the network security controls in most enterprises. These security control strategies are hardened into to their virtual machine, which in turn enhances availability of the provided infrastructure.

As for redundancy, large Cloud Computing system vendors (e.g., Amazon, Google) offer geographic redundancy in their Cloud syst em s , hopefully enabling high availability on a single provider. For example, Amazon builds data canters in multiple regions (e.g., USA, Europe) and various availability zones within those regions. Availability zones are distinct locations that are engineered to be insulated from failures in other availability zones and provide inexpensive, low latency network connectivity to other availability zones in the same region. Using instances in separate availability zones, one can protect applications from failure of a single location. That's to say, Cloud system has capability in providing redundancy to enhance the high availability of the system in nature. As estimated, Google owns more than 1 million machines which are distributed in 36 data centres across the world. Similar to Amazon, Google offers geographic redundancy in its systems. At the mean time, Google file system (GFS)

[8] developed by Google set 3 as the default number of replications for each objects it stores. That's to say, each file stored in the GFS is replicated at 3 places, which further enhances the availability of the system.

In a word, Cloud Computing systems are able to provide available services in nature through hardening and redundancy strategies. Strategies are hardened into to their virtual machine, which in turn enhances availability of the provided infrastructure.

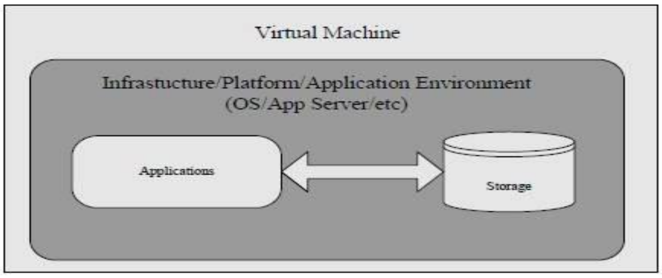


Fig. 3. Virtual Machine as Infrastructure/Platform

As for redundancy, large Cloud Computing system vendors (e.g., Amazon, Google) offer geographic redundancy in their Cloud systems, hopefully enabling high availability on a single provider. For example, Amazon builds data centers in multiple regions (e.g., USA, Europe) and various availability zones within those regions. Availability zones are distinct locations that are engineered to be insulated from failures in other availability zones and provide inexpensive, low latency network connectivity to other availability zones in the same region. Using instances in separate availability zones, one can protect applications from failure of a single location. That's to say, Cloud system has capability in providing redundancy to enhance the high availability of the system in nature. As estimated, Google owns more than 1 million machines which are distributed in 36 data centres across the world. Similar to Amazon, Google offers geographic redundancy in its systems. At the mean time, Google file system (GFS) [8] developed by Google set 3 as the default number of replications for each objects it stores. That's to say, each file stored in the GFS is replicated at 3 places, which further enhances the availability of the system.

In a word, Cloud Computing systems are able to provide available services in nature through hardening and redundancy strategies.

B. *Confidentiality*

Confidentiality means keeping users' data secret in the Cloud systems. The confidentiality in Cloud systems is a big obstacle for users to step into it, as many users said "My sensitive corporate data will never be in the Cloud" in the article named "Above the Cloud" [2]. Currently, Cloud Computing system offerings (e.g., applications and its infrastructures) are essentially public networks, say the applications or systems are exposed to more attacks when comparison to those hosted in the private data centres. Therefore, keeping all confidential data of users' secret in the Cloud is a fundamental requirement which will attract even more users consequently. Traditionally, there are two basic approaches (i.e., physical isolation and cryptography) to achieve such confidentiality, which are extensively adopted

by the Cloud Computing vendors. As discussed, Cloud system offerings (e.g., data, services) are transmitted through public networks. That's to say no physical isolation could be achieved. Alternatively, Virtual Local Area Networks, and network middleboxes (e.g. firewalls, packet filters) should be deployed to achieve the virtual physical isolation [2]. For example, CohesiveFT releases VPN-Cubed [9] to provide a security perimeter for the IT infrastructure whether it is inside a single Cloud or multiple Cloud or hybrid Cloud-datacenter ecosystem. Moreover, Vertica [10] deploys its database on the Amazon EC2 and provides VPN and firewall to secure its database, as shown in Fig. 4 [10]. When a Vertica database instance is provisioned by the Amazon EC2, it provides users full root access so users can secure the system as they see it. They chose to create a VPN between their enterprise users and their Vertica for the Cloud instance and set up a firewall to the outside world. Aside from the VPN port and software, they blocked off all external communication. Encrypted storage is another choice to enhance the confidentiality. For example, encrypting data before placing it in a Cloud may be even more secure than unencrypted data in a local data center; this approach was successfully used by TC3 [11], a healthcare company with access to sensitive patient records and healthcare claims, when moving their HIPAAcompliant application to AWS [11].

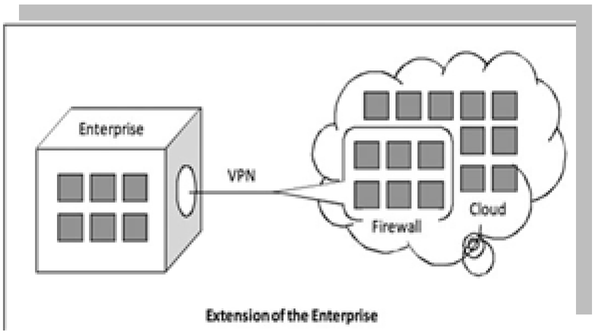


Fig. 4. Vertica Provides VPN and Firewall to Secure Its Database

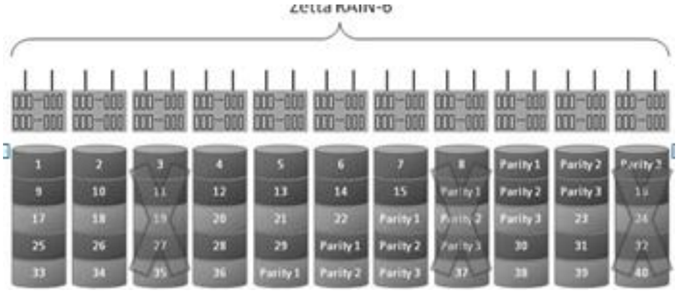


Fig. 5. Zetta RAIN-6 system architecture

C. Data Integrity

Data integrity in the Cloud system means to preserve information integrity (i.e., not lost or modified by unauthorized users). As data is the base for providing Cloud Computing services, such as Data as a Services, Software as a Service, Platform as a Service, keeping data integrity is a fundamental task.

Furthermore, Cloud Computing system usually provides massive data procession capability. Herein, massive data means many Tera Bytes (TB) data or even Peta Bytes (PB) data in volume. The challenges for data integrity associated with data storage in the Cloud Computing system are as follows. Firstly, in terms of the current development of state for hard disk drivers (or solid state disks or tapes), their capacity increases are not keeping pace with the data growth [12]. Therefore, to scale up the data storage in the Cloud Computing systems, vendors need to increase the population of hard drives (or solid state disks or tapes). This may consequently result in increased high probability of either node failure or disk failure or data corruption or even data loss. Secondly, disk drives (or solid state disks) are getting bigger and bigger in terms of their capacity, while not getting much faster in terms of data access. In this section, we give a brief introduction to the Zetta system [13] provided by Zetta which mainly focus on data integrity for Cloud Computing services, which has a similar idea to RAID systems [14].

Zetta [13] provides Zetta system for storage service on demand mainly considering on the data integrity. There, data integrity means that the system won't corrupt or data won't lose, even at tremendous large scale and over long periods of duration, regardless of the corruption vector. Zetta implements RAIN-6 (Redundant Array of Independent Nodes-6) in its Zetta system for the primary data hosting service, mainly considering on the data integrity. It is called RAIN-6 because it has a similar implementation to RAID-6 (Redundant Array of Independent Disks-6), and result in similar capability considering the data integrity. Therefore, RAIN-6 is not only able to tolerant hard drive failure and bit errors, but also to recover from node failure and bit errors for any causes (e.g., network failure, power supply shortage, memory or a hard drive corruption and etc.). This data integrity property is achieved by data placement in terms of node striping. Additionally, Zetta system uses an N+3 implementation comparing to RAID system [14], which means that it is able to tolerate three simultaneous failures (e.g., a three disks failure or even a three entire nodes failure) in a given stripe. Thus, Zetta system provides rather high availability and data integrity. Fig. 5 [13] shows the RAIN-6 architecture in Zetta system. The standard for Zetta system encoding is an 8+3 encoding (i.e., 8 pieces of primary data are encoded into 11 chunks, which are further distributed across 11 independent storage nodes, and each of which contains redundant network connections). This node level processing capacity allows Zetta to deliver data with high integrity (i.e., without corruption).

Digital signature is a commonly used technique for data integrity testing. The widely adopted distributed file systems (e.g., GFS [8], HDFS [15]) usually divide data in large volumes into a set of blocks, each of which has a default size (e.g., 64MB, 128Mb). When a block of the data is physically stored on, a digital signature is attached to it. This digital signature is useful for future integrity testing. Herein, digital signature is able to test the integrity of the data, and recover from corruption.

Hence, data integrity is fundamental for Cloud Computing system, and it is hopeful to be achieved by techniques such as RAID-liked strategies, digital signature and so on.

C. Control

Control in the Cloud system means to regulate the use of the system, including the applications, its infrastructure and the data. Cloud computing system always involves distributed computation on multiple large-scale data sets across a large number of computer nodes. Even more, every Internet user is able to contribute his or her individual data to the Cloud Computer systems which are located on the other side of the Internet, and make use of them. For example, a user's click stream across a set of webs (e.g., Amazon book store, Google search web pages, etc.) can be used to provide targeted advertising. Future healthcare applications may use an individual's DNA sequence (which is captured by hospitals) to develop tailored drugs and other personalized medical treatments. When all these personal data are stored in the Cloud Computing system environment, users of Cloud Computing systems may face many threats to their individual data. For example, let's consider a medical patient who is deciding whether to participate in a health-care study or not. Firstly, he or she may concern the careless or malicious usage of his or her data, and consequently results in the exposure of his or her individual data. For instance, by writing his or her individual data into a world wide readable file which may further be indexed by a search engine. Second, he or she may be concerned that even if all computations are done correctly and securely. However, the study result itself (e.g., the aggregate health-care statistics computed as part of the study) may leak sensitive information about his or her personal medical information. Performing distributed computation in the Cloud Computing systems on such sensitive individual data raises serious security and privacy concerns. Control of the distributed computation in the Cloud Computing systems over those individual data is essential in need.

In the Cloud Computing literature, Airavat [16] integrates decentralized information flow control (DIFC) [17] and differential privacy [18], [19], [20] to provide rigorous privacy and security control in the computation for the individual data in the MapReduce framework [21]. Airavat integrate DIFC into the MapReduce framework (it uses Hadoop [22] instead in the implementation). Therefore, it is able to pay particular attention to the division of labor between the MapReduce framework, the distributed file system (i.e., Hadoop Distributed File System [15]) and the operating system (e.g., Linux). Airavat uses DIFC to ensure that the system is free from unauthorized storage access. For example, it prevents Mappers to leak data over unsecured network connections or leave the intermediate result data in unsecured local files. By providing several trusted initial mappers and trusted reducers, Airavat is able to carry out privacy-preserving computations in the MapReduce framework, eventually allowing users to insert their own mappers while dynamically ensuring differential privacy.

Fig. 6 [16] shows the architecture of Airavat. MapReduce computations start with an Airavat supplied mapper. The initial mapper must be trusted because it reads the data schema. The initial mapper also provides a sampling of input data to allow multiple queries to be composed in parallel, which reduces the amount of noise needed to achieve differential privacy. So long as a user computation starts with a trusted mapper and uses trusted reducers, the result of the computation can use differential privacy. The user can supply arbitrary mapper stages that do not need to be audited. This creates a powerful and flexible computation platform that retains the provable guarantees of differential privacy.

Hence, efficient and effective control over the data access in the Cloud Computing system and regulate the behaviors of the applications (services) hosted on the Cloud Computing systems will enhance the security of systems.

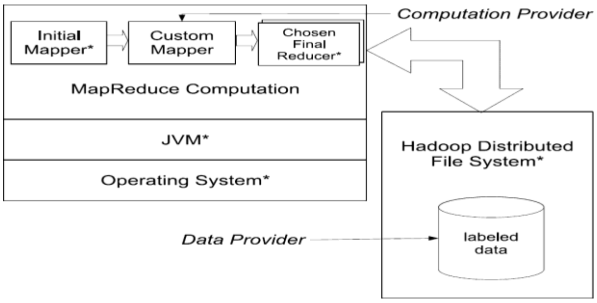


Fig. 6. High level architecture of, trusted components are starred

D. Audit

Audit means to watch what happened in the Cloud system. Auditability could be added as an additional layer above the virtualized operation system (or virtualized application environment) hosted on the virtual machine to provide facilities watching what happened in the system. It is much more secure than that is built into the applications or into the software themselves, since it is able watch the entire access duration. For such kind of scenarios, three main attributes should be audited:

· Events: The state changes and other factors that effected the system availability.

· Logs: Comprehensive information about users' application and its runtime environment.

· Monitoring: Should not be intrusive and must be limited to what the Cloud provider reasonably needs in order to run their facility.

Such a new feature (i.e., audit ability added as an additional layer in the virtual operation systems) reinforces the Cloud Computing developers to focus on providing virtualized capabilities instead of specific hardware to being provided. That's to say they have the capability in auditing the entire Cloud Computing system in technique perspective. Another related concern is that many nations have laws requiring Cloud Computing providers (or SaaS providers) to keep customer data and copyrighted material within national boundaries, which make the audit ability hopefully in the law issue perspective. However, some businesses do not like the ability of a country to get access to their data via the court system; for example, a European customer might be concerned about using Cloud Computing system in the United States given the USA PATRIOT Act [23].

IV. PRIVACY ON DEMAND

As Cloud Computing system usually offers services (e.g. DaaS, SaaS, IPMaaS, PaaS, and so on) on the other side of the Internet in terms of its users, the secret information of individual users' and business' are stored and managed by the service providers, and consequently results in privacy concerns. Privacy issues exist for a long time in the computing literature, and many law acts have been published to protect users' individual privacy as well as business secret. Nevertheless, these acts are out of date and inapplicable to the new scenarios, where a new relationship between users and providers (i.e. three parties) raises .In this subsection, we investigate a few privacy acts to illustrate those acts are not applicable in the new environment, in the subsection III- A. Data storage in the Cloud Computing system which is located in multi regions (locations) to make the system more tolerant may also raise the privacy problems. We present a brief introduction latter in the subsection III-B.

A. Legal Issues

Cloud computing systems (including applications and services hosted on them) has significant implications for the privacy of personal information as well as for the confidentiality of business and governmental information. That's because any information is shifted from local computers to the Cloud Computing systems at the Cloud Computing era,

including email, word processing documents, spreadsheets, videos, health records, photographs, tax or other financial information, business plans, PowerPoint presentations, accounting information, advertising campaigns, sales numbers, appointment calendars, address books, and more. Furthermore, the entire contents of a user's originally stored on local device may be shifted to a single Cloud provider or even to many Cloud providers. Whenever an individual, a business, a government agency, or other entity shares information in the Cloud, privacy or confidentiality questions may arise.

B. Multi Location Issues

Cloud system means to offer huge computer resource to users, including infrastructure, platform, services (e.g., storage, computing power, and so on). Hence, a business has to trust the Clouds system vendor and store its private data to the Cloud system. That's to say the business's data are stored in someone else's computer (say in someone else's facility). However, many things can go wrong, if data are stored in someone else's devices. For example, the Cloud service provider may go out of business or may decide to hold the data hostage if there is a dispute. Furthermore, large Cloud system vendors have their Cloud mirror sites in many other countries. For example, Amazon has its EC2 in multi-locations, and currently one in USA and the other in Europe.

Google App Engine locates in may countries too (i.e., it has 36 data centers across the world), such as USA, China, and so on. We list a few problems that may occur, if the private data are stored in multi-locations [23].

V. Conclusion & Future Work

Cloud Computing becomes a buzzword nowadays. More and more companies step into Cloud and provide services above on it. However, security and privacy issues impose strong barrier for users' adoption of Cloud systems and Cloud services. We observed the security and privacy concerns presented by an amount of Cloud Computing system providers in this paper. Nevertheless, those concerns are not adequate. More security strategies should be deployed in the Cloud environment to achieve the 5 goals (i.e. availability, confidentiality, data integrity, control and audit) as well as privacy acts should be changed to adapt a new relationship between users and providers in the Cloud literature. We claim that prosperity in Cloud Computing literature is to be coming after those securities and privacy issues resolved.

References

[1] C. Wang, "Forrester: A close look at cloud computing security issues, "http://www.forrester.com/securityforum2009, 2009.

[2] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, et al., "Above the clouds: A Berkeley view of cloud computing," University of California, Berkeley, Tech. Rep, 2009.

[3] IDC, "It cloud services user survey, pt.2: Top benefits & challenges," http://blogs.idc.com/ie/?p=210, 2008.

[4] GNU, "Xen," http://www.xen.org/, 2008.

[5] Skytap, "Skytap," http://www.skytap.com/, 2008.

[6] E. Corporation, "Wmware," http://www.vmware.com/, 2008.

[7] Microsoft, "Hyper-v," http://www.microsoft.com/windowsserver2008/en/us/hyperv-main.aspx, 2008.

[8] S. Ghemawat, H. Gobioff, and S. Leung, "The Google file system," in Proceedings of the 19th Symposium on Operating Systems Principles (OSDI'2003), 2003, pp. 29–43.

[9] CohesiveFT, "VPN Cubed," http://www.cohesiveft.com/vpncubed/, 2008. [10] Vertica, "Vertica for the Cloud," http://www.vertica.com/cloud, 2008. [11] T. Healthcare, "TC3 Healthcare," http://www.tc3health.com/, 2008.

[10] J. F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlicht- ing, and A. Toncheva, "The diverse and exploding digital universe," IDC Future Report, 2008.

[11] Zetta,"Zetta: Enterprise cloud storage on demand," http://www.zetta.net/, 2008.

[12] P. Chen, E. Lee, G. Gibson, R. Katz, and D. Patterson, "RAID: High- performance, reliable secondary storage," ACM Computing Surveys (CSUR), vol. 26, no. 2, pp. 145–185, 1994.

[13] Yahoo!,"Hadoop distribted file system architecture,"http://hadoop.apache.org/common/docs/current/hdfs design.html, 2008.

[14] I. Roy, H. Ramadan, S. Setty, A. Kilzer, V. Shmatikov, and E. Witchel, "Airavat: Security and Privacy for MapReduce."

[15] M. Krohn, M. Brodsky, M. Kaashoek, and R. Morris, "Information flow control for standard OS abstractions," ACM SIGOPS Operating Systems Review, vol. 41, no. 6, pp. 321–334, 2007.

[16] C. Dwork et al., "Differential privacy," LECTURE NOTES IN COM- PUTER SCIENCE, vol. 4052, p. 1, 2006.

[17] C. Dwork, "Differential privacy: A survey of results," Lecture Notes in Computer Science, vol. 4978, p. 1, 2008.

[18] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in Proceedings of the 48th Annual Symposium on Foundations of Computer Science, 2007.

[19] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," in Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation-Volume 6 table of contents, 2004, pp. 10–10.

[20] Yahoo!, "Hadoop," http://hadoop.apache.org, 2008.

[21] J. Bardin, "Security Guidance for Critical Areas of Focus in Cloud Computing," www.cloudsecurityalliance.org/guidance/csaguide.pdf, 2009.