Research on the Data Storage Security based on Cloud Computing

Xinqiang MA[1, 2, a], Yi HUANG[1, b*], Yongdan ZHANG[2, c], Youyuan LIU[1, d]

[1] Key Laboratory of Machine Vision and Intelligent Information System, Chongqing University of Arts and Sciences, Chongqing, 402160, China

[2] Colleges of Technology, Guizhou University, Guiyang, 550025, China

[a]email: xinqma@163.com, [b]* corresponding author email: cqhy@21cn.com, [c] email: 312865700@qq.com, [d] email: 39541385@qq.com

Keywords: Cloud Computing; Data Storage Security; Cloud Service; Software Security

Abstract. Cloud Computing has been envisioned as the next-generation architecture of IT Enterprise. Although envisioned as a promising service platform for the Internet, this new data storage paradigm in “Cloud” brings about many challenging design issues which have profound influence on the security and performance of the overall system. This work studies the problem of ensuring the integrity of data storage in Cloud Computing. One of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. To ensure cloud data storage security, it is critical to enable a third party auditor (TPA) to evaluate the service quality from an objective and independent perspective. We review Cloud Computing, cloud data storage and discuss data storage security based on analysis of cloud security treats. Extensive security and performance analysis show that the proposed schemes are highly efficient and provably secure.

Introduction

The development of virtualization technologies have made supercomputing more accessible and affordable. Powerful computing infrastructures hidden in virtualization software make systems to be like a true physical computer, but with the flexible specification of details such as number of processors, memory and disk size, and operating system. The use of these virtual computers is known as Cloud Computing [1], which has been one of the most robust Big Data techniques [2][3][4].

Cloud computing represents one of the most significant shifts in information technology many of us are likely to see in our lifetimes. It offers an innovative business model for organizations to adopt IT services without upfront investment. According to Gartner's Hype cycle, cloud computing has reached a maturity that leads it into a productive phase. This means that most of the main issues with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. This however does not mean that all the problems listed above have actually been solved, only that the according risks can be tolerated to a certain degree. Cloud computing is therefore still as much a research topic, as it is a market offering [5]. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. What is clear through the evolution of Cloud Computing services is that the CTO is a major driving force behind Cloud adoption. The major Cloud technology developers continue to invest billions a year in Cloud research and development (R&D); for example, in 2011 Microsoft committed 90% of its $9.6bn R&D budget to Cloud.

Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rack space, etc. are required for a variety of functions. With

cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications [6].

Cloud data storage is popularly used as the development of cloud technologies. We know that the network bandwidth capacity is the bottleneck in cloud and distributed systems, especially when the volume of communication is large. On the other side, cloud storage also lead to data security problems [7] as the requirements of data integrity checking. In order to solve the problem of data integrity checking, many schemes are proposed under different systems and security models. In Cloud Computing, the remotely stored electronic data might not only be accessed but also updated by the clients, e.g., through block modification, deletion, insertion, etc. Many schemes were proposed under different systems and security models [8] [9]. Unfortunately, the state of the art in the context of remote data storage mainly focus on static data files and the importance of this dynamic data updates has received limited attention so far [10][11] and so on.

Although the existing schemes aim at providing integrity verification for different data storage systems, the problem of supporting both public auditability and data dynamics has not been fully addressed. How to achieve a secure and efficient design to seamlessly integrate these two important components for data storage service remains an open challenging task in Cloud Computing [7].

It is a survey more specific to the different security issues and the associated challenges that has emanated in the cloud computing system [12].Cloud storage security is a complex issue, involving the different levels of the cloud, external and internal threats, and responsibilities that are divided between the user, the provider and even a third party. This paper provides knowledge of cloud computing and cloud security controls in terms of cloud computing based on analysis of cloud security controls and threats. In the same time, it proposes access control for ensuring the confidentiality and integrity of computations that are outsourced to varies services.

Cloud computing

Cloud computing is a computing term or metaphor that evolved in the late 2000s, based on utility and consumption of computer resources. Cloud computing involves deploying groups of remote servers and software networks that allow centralized data storage and online access to computer services or resources. Clouds can be classified as public, private or hybrid. Cloud computing frameworks (see Figure 1.) metaphor: For a user, the network elements representing the provider-rendered services are invisible, as if obscured by a cloud. Cloud computing is the result of evolution and adoption of existing technologies and paradigms. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and helps the users focus on their core business instead of being impeded by IT obstacles.

Cloud computing providers offer their services according to several fundamental models [6]: see Figure 2.

Infrastructure as a service (IaaS): In the most basic cloud-service model & according to the IETF (Internet Engineering Task Force), providers of IaaS offer computers physical or (more often) virtual machines and other resources. (A hypervisor, such as Xen, Oracle VirtualBox, KVM, VMware ESX/ESXi, or Hyper-V runs the virtual machines as guests. Pools of hypervisors within the cloud operational support-system can support large numbers of virtual machines and the ability to scale services up and down according to customers' varying requirements.) IaaS clouds often offer additional resources such as a virtual-machine disk image library, raw block storage, and file or object storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles. IaaS-cloud providers supply these resources on-demand from their large pools installed in data centers. For wide-area connectivity, customers can use either the Internet or carrier clouds (dedicated virtual private networks). To deploy their applications, cloud users install operating-system images and their application software on the cloud infrastructure. In this model, the cloud user patches and maintains the operating systems and the application software. Cloud providers typically bill IaaS services on a utility computing basis: cost reflects the amount of resources allocated and consumed.

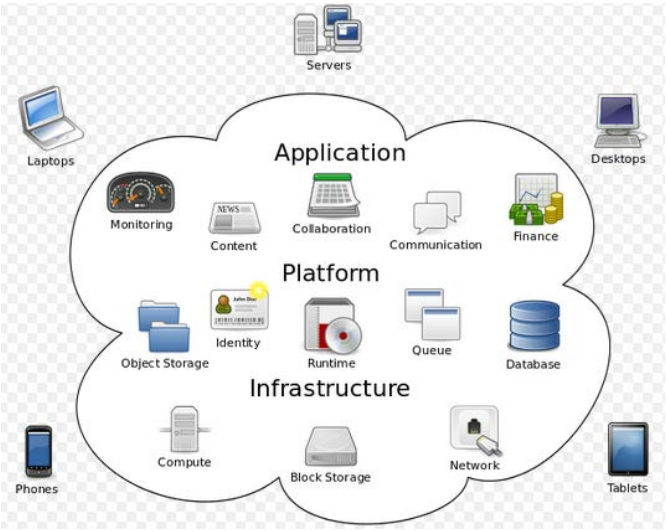

Fig.1. Cloud computing frameworks

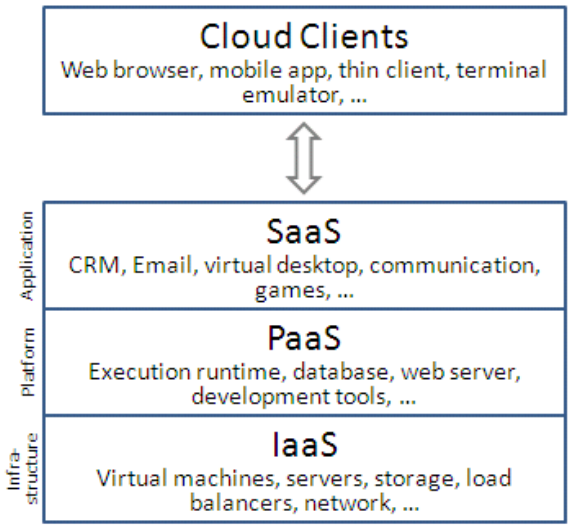

Fig.2. Cloud service models

Platform as a service (PaaS): In the PaaS models, cloud providers deliver a computing platform, typically including operating system, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers. With some PaaS offers like Microsoft Azure and Google App Engine, the underlying computer and storage resources scale automatically to match application demand so that the cloud user does not have to allocate resources manually. The latter has also been proposed by an architecture aiming to facilitate real-time in cloud environments. Even more specific application types can be provided via PaaS, e.g., such as media encoding as provided by services as bitcodin transcoding cloud or media.io.

Software as a service (SaaS): In the SaaS model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. Cloud users do not manage the cloud infrastructure and platform where the application runs. This eliminates the need to install and run the application on the cloud user's own computers, which simplifies maintenance and support. Cloud applications are different from other applications in their scalability—which can be achieved by cloning tasks onto multiple virtual machines at run-time to meet changing work demand. Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user, who sees only a single access point. To accommodate a large number of cloud users, cloud applications can be multitenant, that is, any machine serves more than one cloud user organization. Proponents claim SaaS allows a business the potential to reduce IT operational costs by outsourcing hardware and software maintenance and support to the cloud provider. This enables the business to reallocate IT operations costs away from hardware/software spending and personnel expenses, towards meeting other goals. In addition, with applications hosted centrally, updates can be released without the need for users to install new software. One drawback of SaaS is that the users' data are stored on the cloud provider's server. As a result, there could be unauthorized access to the data. For this reason, users are increasingly adopting intelligent third-party key management systems to help secure their data.

Cloud clients: Users access cloud computing using networked client devices, such as desktop computers, laptops, tablets and smartphones. Some of these devices – cloud clients – rely on cloud computing for all or a majority of their applications so as to be essentially useless without it. Examples are thin clients and the browser-based Chromebook. Many cloud applications do not require specific software on the client and instead use a web browser to interact with the cloud application. With Ajax and HTML5 these Web user interfaces can achieve a similar, or even better, look and feel to native applications. Some cloud applications, however, support specific client software dedicated to these applications. Some legacy applications are delivered via a screen-sharing technology.

Cloud computing also includes its architecture and types see Figure 3 and Figure 4 and so on.

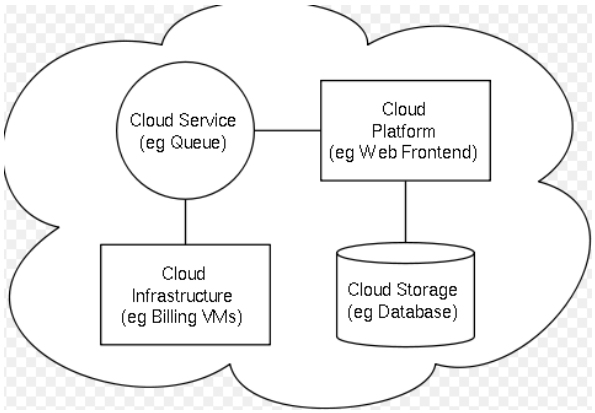

Fig.3. Cloud computing sample architecture

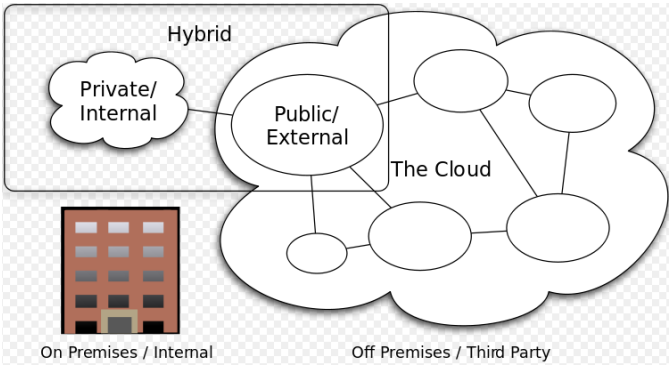

Fig.4. Cloud computing types

Cloud Data Storage

Cloud storage is a model of data storage where the digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data. It services may be accessed through a co-located cloud computer service, a web service application programming interface (API) or by applications that utilize the API, such as cloud desktop storage, a cloud storage gateway or Web-based content management systems.

Cloud storage is based on highly virtualized infrastructure (see Figure 5) and is like broader cloud computing in terms of accessible interfaces, near-instant elasticity and scalability, multi-tenancy, and metered resources. Cloud storage services can be utilized from an off-premises service (Amazon S3) or deployed on-premises (ViON Capacity Services). Cloud storage typically refers to a hosted object storage service, but the term has broadened to include other types of data storage that are now available as a service, like block storage. Object storage services like Amazon S3 and Microsoft Azure Storage, object storage software like Openstack Swift, object storage systems like EMC Atmos and Hitachi Content Platform, and distributed storage research projects like OceanStore and VISION Cloud are all examples of storage that can be hosted and deployed with cloud storage characteristics. Cloud storage is: Made up of many distributed resources, but still acts as one - often referred to as federated storage clouds [14]. Highly fault tolerant through redundancy and distribution of data. Highly durable through the creation of versioned copies. Typically eventually consistent with regard to data replicas.

Data Storage Security and privacy

Cloud security is an evolving sub-domain of computer security, network security, and, more broadly, information security. It refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing [6].

Identity management: Every enterprise will have its own identity management system to control access to information and computing resources. Cloud providers either integrate the customer’s identity management system into their own infrastructure, using federation or SSO technology, or provide an identity management solution of their own.

Physical security: Cloud service providers physically secure the IT hardware (servers, routers, cables etc.) against unauthorized access, interference, theft, fires, floods etc. and ensure that essential supplies (such as electricity) are sufficiently robust to minimize the possibility of disruption. This is normally achieved by serving cloud applications from 'world-class' (i.e. professionally specified, designed, constructed, managed, monitored and maintained) data centers.

Personnel security: Various information security concerns relating to the IT and other

professionals associated with cloud services are typically handled through pre-, para- and post-employment activities such as security screening potential recruits, security awareness and training programs, proactive security monitoring and supervision, disciplinary procedures and contractual obligations embedded in employment contracts, service level agreements, codes of conduct, policies etc.

Availability: Cloud providers help ensure that customers can rely on access to their data and applications; at least in part (failures at any point - not just within the cloud service providers' domains - may disrupt the communications chains between users and applications).

Application security: Cloud providers ensure that applications available as a service via the cloud (SaaS) are secure by specifying, designing, implementing, testing and maintaining appropriate application security measures in the production environment. Note that - as with any commercial software - the controls they implement may not necessarily fully mitigate all the risks they have identified, and that they may not necessarily have identified all the risks that are of concern to customers. Consequently, customers may also need to assure themselves that cloud applications are adequately secured for their specific purposes, including their compliance obligations.

Privacy: Providers ensure that all critical data (credit card numbers, for example) are masked or encrypted (even better) and that only authorized users have access to data in its entirety. Moreover, digital identities and credentials must be protected as should any data that the provider collects or produces about customer activity in the cloud.

Many of these regulations mandate particular controls (such as strong access controls [15] and audit trails) and require regular reporting. Cloud customers must ensure that their cloud providers adequately fulfill such requirements as appropriate, enabling them to comply with their obligations since, to a large extent, they remain accountable. Cloud providers have business continuity and data recovery plans in place to ensure that service can be maintained in case of a disaster or an emergency and that any data loss will be recovered. In addition to producing logs and audit trails, cloud providers work with their customers to ensure that these logs and audit trails are properly secured, maintained for as long as the customer requires, and are accessible for the purposes of forensic investigation (e.g., eDiscovery). In addition, there are considerations for acquiring data from the cloud that may be involved in litigation.

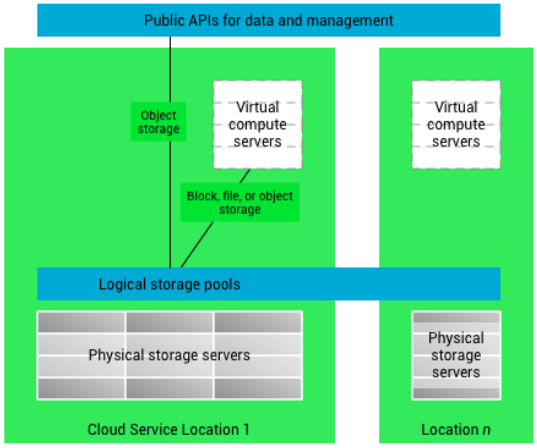

Fig.5. High level cloud storage architecture

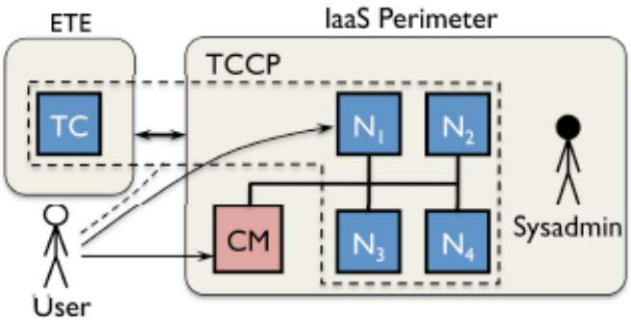

Fig.6. Changing the architecture of TCCP

Conclusion

Cloud Computing moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. Solutions to privacy include policy and legislation as well as end users' choices for how data is stored. Users can encrypt data that is processed or stored within the cloud to prevent unauthorized access. We present the trusted cloud computing platform (TCCP) [16] that provides a closed box execution environment by extending the concept of trusted platform to an entire IaaS backend.

Acknowledgement

In this paper, the research was sponsored by the Natural Science Foundation of CQ CSTC (Project No. cstc2014jcyjA40056 and cstc2013jcyjA40053) and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Project No. KJ1401112) and the Natural Science Foundation of Ycstc (Project No. 2013nb8001, 2014bf2001 and 2013ad2002) and the Youth Fund of Guizhou University (Project No. 201240).

References

[1] Borko Furht, Armando Escalante, Handbook of Cloud Computing, Springer, 2011.

[2] C.L. Philip Chen and Chun-Yang Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data[J], Information Sciences 2014 (275) 314-347.

[3] Sherif Sakr, Anna Liu, Daniel M. Batista, Mohammad Alomari, A survey of large scale data management approaches in cloud environments[J], IEEE Commun. Surv. Tutorials 2011 13(3) 311-336.

[4] Eric E. Schadt, Michael D. Linderman, Jon Sorenson, Lawrence Lee, Garry P. Nolan, Computational solutions to large-scale data management and analysis[J], Nat. Rev. Genet. 2010 11 (9) 647-657.

[5]Smith, David Mitchell. Hype Cycle for Cloud Computing [J], Gartner. 2013.

[6]http://en.wikipedia.org/wiki/Cloud_computing#cite_note-90, April 2015.

[7] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li, Enabling public auditability and data dynamics for storage security in cloud computing [J], IEEE Trans. Parallel Distrib. Syst. 2011 22 (5) 847-859.

[8] Qian Wang, Kui Ren, Wenjing Lou, Yanchao Zhang, Dependable and secure sensor data storage with dynamic integrity assurance, in: Proc. IEEE INFOCOM, 2009.954-962.

[9] Alina Oprea, Michael K. Reiter, Ke Yang, Space efficient block storage integrity, in: Proc. 12th Ann. Network and Distributed System Security Symp (NDSS 05) [C].2005.

[10] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z.Peterson, and D. Song, Provable Data Possession at UntrustedStores, Proc. 14th ACM Conf. Computer and Comm. Security (CCS'07) [C]. 2007. 598-609.

[11] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," Proc. Fourth Int'l Conf. Security and Privacy in Comm. Networks (SecureComm '08) [C]. 2008.1-10.

[12] Kuyoro S. O., Ibikunle F. , and Awodele O. Cloud Computing Security Issues and Challenges[J], International Journal of Computer Networks, 2011 3(5) 247-255.

[13] Hassan, Qusay. Demystifying Cloud Computing [J], The Journal of Defense Software Engineering (CrossTalk) 2011 (2) 16–21.

[14] Vernik, Gil, et al. Data On-boarding in Federated Storage Clouds[C]. Proceedings of the 2013 IEEE Sixth International Conference on Cloud Computing. IEEE Computer Society, 2013.

[15] Yi. Huang and Xinqiang Ma. An access control model based on Trusted Computing[J], Journal of Chongqing University of Arts and Sciences, 2010 29(3) 54-57.

[16] Huang Yi, Ma Xinqiang, Liu Youyuan, Li Danning. Research of the Issues based on Trusted Cloud Security [J], Advanced Materials Research, 2014(Vols. 989-994) 5000-5003.