Introduction

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction[1]. Basically the cloud computing model is composed by three service models:

- *Cloud Software as a Service(SaaS)* provides computation capacity by running software on a cloud infrastructure. `These applications are remotely accessible by users who do not have to manage the cloud infrastructure hardware but only utilize the services. Different cloud consumers' applications are organized in a single logical environment on the SaaS cloud to achieve economies of scale and optimization with regard to issues like speed, security, availability, disaster recovery, and maintenance[1, 2]. Basically those cloud storage services which this research mainly aims to, Dropbox and Google Drive for instance, are representative examples of SaaS.
- *Cloud Platform as a Service(PaaS)* is a development framework and service hosting platform which allows users to develop cloud service by providing runtime environment, R&D toolkits, SaaS application APIs, storage and middleware/OS support. A typical instance of PaaS is Google Application Engine.
- *Cloud Infrastructure as a Service(IaaS)* refers to on-demand provisioning of infrastructural resources on the cloud, usually in terms of VMs[3]. The capability provided to the consumer is to provision processing, storage management, network configurations, and other fundamental computing resources so that the consumer is able to deploy and run any software on the cloud[1, 3]. Amazon EC2 is a representational IaaS cloud platform.

Cloud storage refers to the service on the cloud which is delivered as remote storage infrastructure. It can be accessed from any location connected to the Internet, and offer additional and data-related functionalities with regard to data management and maintenance[4]. Technically when the core tasks that a cloud computing system processes are relevant to big data storage and maintenance, it becomes a cloud storage system in which most hardware/software deployment/configuration are optimized aiming for providing data storage service. Just as it is a service that provide interface to users to do data manipulations but not making use of certain physical storage devices directly, cloud storage is a typical application of SaaS.

For personal users, cloud storage is a good extension of local storage with high cost performance, an ideal replacement of portable storage device like flash drive or portable hard drive by which stable data storage is not guaranteed, also a more reliable local disk backup given the its essence of good cost efficiency, remote accessibility and reliability. Besides, from the enterprise-level user perspective, cloud storage reduces the investment they should have made in terms of design, management and maintenance of device clusters which providing systematic big data storage service. What's more, the data sharing, which local storage could never provide but cloud storage services do given the characteristics of Internet accessibility, is another thing that boost up data usage efficiency via

facilitating the real time data exchange and decreasing the workload about data collection especially in terms of frequently reusable data resource.

As a highly developed commercial-level cloud storage product and the ground breaker in this area, Dropbox is becoming the industry leader with a market share of 14.14% in 2011[5]. Google Drive, another popular cloud storage service which is built based on the mature application framework of Google and well integrated with other Google services (e.g. Gmail as a major way of user or system activity notification) is also increasing dramatically. It is an industrial routine for cloud storage service vendors to provide multiple ways of getting access to the service from different platforms like desktop and portable devices to radically achieve the goal of a ubiquitous usage.

Motivation

As same as the high speed development of cloud storage industry, the security anxieties in the age of cloud are rising as well. The cloud storage technology, especially its security features, is always compared with traditional ways like local disk or RAID. To personal users, the reasons of using or not using the cloud storage service are usually simple and straight forward. For example, would people put the top security sensitive personal data like bank account information and passcode on the cloud? What is more, even if most service carriers claim that they deploy highly secure encryption scheme throughout, is it definitely sure that the employees from the service carriers will abide strictly by the professional integrity and never try to access users' data unconsciously by technical approaches which could be performed easily from internal of the system? In addition, some expert personal users would consider about if the twenty-four-seven on service cloud storage server which is a big hacking target at the same time will face vast number of hacking trials is essentially even more security risky than local storage? Last but not least, the personal users, particularly those who are using the cloud storage service for free, are facing the problem that if the carrier will take full responsibility of the loss which brought about by the security issues. Such problems with regard to security confidence make personal users, even though the cloud storage service provider could be Google or Dropbox, put only unimportant data on the cloud as redundancy and utilize the its feature of portability. It seems paradoxical that people usually do redundancy for important data but the security concerns stop them from doing that.

To the enterprise-level users, besides the same concerns that might be taken into consideration as a personal user, there are more significant problems to be dealt with. According to the survey hosted by Intel in May 2012 [6], the concerns of IT professionals in terms of public cloud (Opposite to the private cloud and traditional IT storage solutions like local disk or RAID. The storage service provided by Google Drive and Dropbox talking about in the paper are typical instances of public cloud) could be concluded as follows: First, the lack of measurement of security capabilities of cloud storage service. Despite the security methods deployed are claimed to be secure, the providers usually remain the details under cover and the only approach that is transparent to users is usually and simply the identity based access control, which is apparently not enough to gain their users confidence towards the storage system in terms of security. 57% of the survey participants think it is the most significant problem concerned when

using such service among all the concerns [6]. 55% of the participants treat the lack of control over data as another major concern because of the shared infrastructure in the cloud with no visibility into and operability of these abstracted resources [6]. Only 36% of the participants think the cloud storage service lack of transparency/ability to perform audits, which indicates that the continuous improved comprehensive query interfaces with regard to storage service itself make it acceptable by gaining the level of transparency[6]. Compliance with regulatory mandates is another key problem concerned by an average of 78% participants.[6] Due to the opaque security and operability of cloud storage, the security of storing legislation sensitive data might not be fully measurable which may leads to compliance problems. Such violation judged based on laws varies from country to country but the advised way to make fully use of cloud storage for enterprise data is classifying the data with different priority and keep them in internal infrastructure and external storage service respectively.

The goal of the project is to investigate the feasibility of gaining the confidence towards security of cloud storage service by utilization of asymmetric encryption mechanism in local host and evaluate the advantage of information sharing based on asymmetric cryptology over the symmetric one which is commonly used currently. To achieve this, it was decided to develop a client on Linux platform which performs local file management and cryptology manipulations and a server, which works as public key infrastructure, user management system and a file sharing access interface. The combination of two components will simulate the entire procedure as a prototype.

On completion of the project, the production would be demonstrate to experts in security area and also several testers who has not a background of computer science would be extensively invited to perform the user acceptance test to assess the availability and user experience.

This Report

This report will contain a review of the state of the art in cloud storage, cloud security approaches and other potential technologies as they relate to this project. Some of the mainstream referable approaches and how they relate to this project will be discussed as well.

Some important design decisions which are made during the procedure of this project will then be explained, so would the reasoning behind these decisions. It would be followed by a detailed outline of both the research and development, also the implementation of key components of the product software. A system analysis of the software prototype would be given based on not only the expert in certain area but also the testers who use potential users. It will focus on the theoretical correctness, availability and quality of the project based on the theoretical proof and feedback from beta users.

Additionally, the feasibility of implementing such app on portable devices in terms of both network and computation capacity would be discussed as possible follow ups to be carried out.

At last the conclusion of the whole project would be presented.

Bibliography

[1]. Mell, P., & Grance, T. (2011). The NIST definition of cloud computing (draft).*NIST special publication*, *800*, 145.

[2]. Dillon, T., Wu, C., & Chang, E. (2010, April). Cloud computing: issues and challenges. In *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on* (pp. 27-33). Ieee.

[3]. Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications*, *1*(1), 7-18.

[4]. Cachin, C., Keidar, I., & Shraer, A. (2009). Trusting the cloud. *ACM SIGACT News*, *40*(2), 81-86.

[5]. OPSWAT(Dec 2011). Security Industry Market Share Analysis

[6]. Intel Corporation(May 2012). What's holding back the cloud: Intel Survey on Increasing IT Professionals' Confidence in Cloud Security. 327311-001