Verifiable File Search on the Cloud

R. Varaprasad[1], B. Varalakshmi[2]
[1,2]Dept of Computer Science and Engineering
[1,2]G.Pullaiah College of Engineering and Technology

Abstract- *Presently, the world is steadily being blend with a potential volume of information that is available in the digital format. Hence, the amount of private, otherwise personal or sensitive information gathered and stored is continuously mounting. While handling with such scenarios, one of the most challenging issues is when businesses need to buy a heavy storage space and computational powers from untrusted third parties (service providers), such as cloud. In certain situations, these service providers (SPs) can alter or reproduce the confidential or sensitive data.*

Keywords- Verifiable File Search, Access Control, Cloud computing, Formal Language.

I. INTRODUCTION

They might provide feasible results without performing any computations over the data. Another is the use of devices (such as wireless sensors, security access cards, notebooks and cell-phones) that may also necessitate outsourcing of the expensive computations (like photo manipulation, data analysis) to the untrusted service providers. The process of producing this outsourcing data along with the verifiable property is known as verifiable computation (or verified computation). The concept of verifiable computation (VC) is absolutely relevant to several real-world situations. Volunteer computing is an application of distributed computing in which computer owners can donate their computational resources (storage and processing power) for computing the small units of projects.

II. VERIFIABLE FILE SEARCH

The basic mechanism is to split large data processing into small units, distribute these units to volunteers for processing, and merging the results through an easier mechanism. The Great Internet Mersenne Prime Search (GIMPS) was the first project that introduced the concept of volunteer computing. A complete middleware system for volunteer computing known as Berkeley Open Infrastructure for Network Computing (BOINC).BONCI is a software platform for volunteer computing includes a client, server, web components and apis for connecting other components. Cloud computing is another means, that provides the illusion of unlimited computing resources to both the organizations and the individual clients at the economical price. This prototype also concentrates on promoting the effectiveness of the shared resources. According to the National Institute of Standards and Technology (NIST), cloud computing has five fundamental characteristics, four deployment models, and three service models. The fundamental characteristics are: On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, Measured service. The deployment models are termed as:

1) Private cloud
2) Community cloud
3) Public cloud
4) Hybrid cloud.

PUBLIC CLOUD: In a public cloud, individual businesses share on premise and access to basic computer infrastructure (servers, storage, networks, development platforms etc.) provided by a CSP. Each company shares the CSP's infrastructure with the other companies that have subscribed to the cloud. Payment is usually pay-as-you-go with no minimum time requirements. Some CSPs derive revenue from advertising and offer free public clouds.

PRIVATE CLOUD: In a private cloud, a business has access to infrastructure in the cloud that is not shared with anyone else. The business typically deploys its own platforms and software applications on the cloud infrastructure. The business's infrastructure usually lies behind a firewall that is accessed through the company intranet over encrypted connections. Payment is often based on a fee-per-unit-time model.

HYBRID CLOUD: In a hybrid cloud, a company's cloud deployment is split between public and private cloud infrastructure. Sensitive data remains within the private cloud where high security standards can be maintained. Operations that do not make use of sensitive data are carried out in the public cloud where infrastructure can scale to meet demands and costs are reduced.

COMMUNITY CLOUD: Community clouds are a recent variation on the private cloud model that provide a complete cloud solution for specific business communities. Businesses share infrastructure provided by the CSP for software and

development tools that are designed to meet community needs. In addition, each business has its own private cloud space that is built to meet the security, privacy and compliance needs that are common in the community.

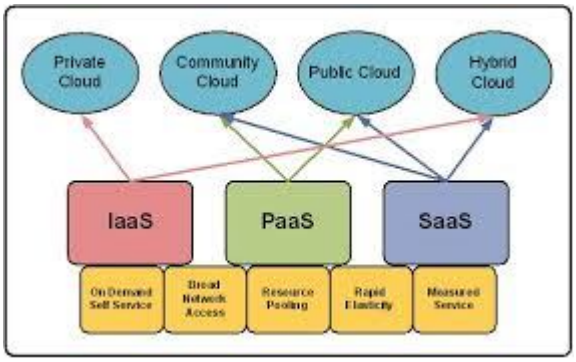
The service models or Deployment model include

1) Software as a Service (SaaS)
2) Platform as a Service (PaaS)
3) Infrastructure as a Service (IaaS)

SAAS: Cloud application services, or Software as a Service (SaaS), represent the largest cloud market and are still growing quickly. SaaS uses the web to deliver applications that are managed by a third-party vendor and whose interface is accessed on the clients' side. Most SaaS applications can be run directly from a web browser without any downloads or installations required, although some require plugins.

PAAS: Cloud platform services, or Platform as a Service (PaaS), are used for applications, and other development, while providing cloud components to software. What developers gain with PaaS is a framework they can build upon to develop or customize applications.PaaS makes the development, testing, and deployment of applications quick, simple, and cost-effective. With this technology, enterprise operations, or a third-party provider, can manage OSes, virtualization, servers, storage, networking, and the PaaS software itself. Developers, however, manage the applications.

IAAS: Cloud infrastructure services, known as Infrastructure as a Service (IaaS), are self-service models for accessing, monitoring, and managing remote datacenter infrastructures, such as compute (virtualized or bare metal), storage, networking, and networking services (e.g. firewalls). Instead of having to purchase hardware outright, users can purchase IaaS based on consumption, similar to electricity or other utility billing.



The characteristic on-demand services are beneficial for the businesses in the following ways:

1) No need to keep and maintain the high priced hardware.
2) Pay for what is used.
3) Provides the location independence.
4) Scalability is very easy.
5) It improves the performance of an organization.

Apart from the above advantages, the businesses may also have some of the issues such as:

1) The private data can be altered or duplicated.
2) The appropriate operations may not be performed on the specific data.
3) The results provided by the computations may not be definite.

In volunteer computing, for example, a fraudulent volunteer can unknowingly produce errors during computation process. In cloud services, a malicious worker could have some financial incentives for providing faulty results to the client. Furthermore, these services perform computation in a black-box manner, hence may invite in-traceable faults such as hardware problems, manipulation of data during storage or transmission, misconfiguration, and many more.The question that arises is:How can it be guaranteed to the clients that service providers have faithfully conducted the computation without revealing the confidential data? To answer the aforementioned question and to resolve these problems, a practical paradigm VC is introduced. The problem of malignant activities of workers/intruders to alter and forge the results motivated the formalization of the concept of security and verifiability of outsourced computation. In order to maintain the privacy of clients some of the conflicting properties such as integrity and anonymity, need to be overcome. Volunteer computing is shown in the below figure 1.

Fig 1: Volunteer Computing

III. TO OBTAIN A SOLUTION

To achieve the above properties after getting the results from outsourcing service providers the client could be capable to verify the accuracy of that outcome in a way that is:

(a) the client uses considerably fewer resources than that are actually needed to perform the computation from scrap.
(b) the service providers could not use any additional resources rather than what actually used for operating the computation.

The above all is a key factor that the private or sensitive information of the client is not compromised during the computation process. The proposal of Volunteer Computing is given by Goldwasser et al. and Babai who formalized the concept of interactive proofs by employing interactive and randomized verification mechanism in 1985.In Interactive proofs consider a scenario, in which the client communicates with the powerful but untrusted worker in order to verify the correctness of the computation results returned by the prover.An extension of interactive proofs known as zero-knowledge (ZK) proof systems that take into account a scenario, in which prover convinces a verifier that statement is true without leaking any related information.In order to improve the performance of the interactive proofs while reducing the computational power difference between prover and verifier. This work laid the foundation for probabilistically checkable proof (PCP) systems.Probabilistically checkable proofs study the power of probabilistically checkable proof systems under various restrictions of the parameters (completeness, soundness, randomness complexity, query complexity, and alphabet size).Proceeding to verifiable computation.

IV. ASSUMPTIONS BASED APPROACH

Many researchers and development, academia, IT industry and vendors of the information security community have recommended assumptions based solutions towards Volunteer Computing. Some among them focused on specially designed trusted hardware, while the others presented audit based solutions for providing security and verifying the correctness of computations. Further proposals suggested making the assumptions of either replication or attestation.

V. CLOUD SECURITY

Novel Protocols: We formalize the problem of verifiable file search over outsourced data. We propose two protocols: the first protocol enables a user to search the cloud data in a verifiable and privacy-preserving manner; the second protocol, besides including the functionality of the first protocol, supports accesscontrol inherently. Thus, the cloud cannot cheat a user by representing that an existing file does not exist, or a nonexistent file does exist.

Theoretical Foundation: We formally prove the correctness of our protocols and their security properties. We analyze the privacy of our designed protocol along with the overhead with respect to computation, storage, and communication. Our theoretical analysis contributes to the foundation and understanding ofthe verifiable file search problem and the design of secure and efficient protocols.

Empirical Validation: We have implemented both protocols, and using a real-world data set, we have conducted experiments to measure the computation, storage, and communication costs of the proposed protocols. Our experimental results validate the effectiveness and efficiency of the proposed protocols.

VI. CONCLUSION

The verifiable file search problem for cloud storage with single and multiple security levels, and addressed the problem by proposing two lightweight, efficient and secure protocols, i.e., VFS and DiffVFS. To be specific, the VFS protocol enables a group user to verify the correctness of a file search result from the cloud. VFS also protects filename privacy. Built on VFS, DiffVFS further enables user differentiation, meaning that different users can only access files that fit their security privileges. We have formally defined and then established the security for VFS and DiffVFS. We have also implemented a prototype for both protocols. Using a real-world data set, we have carried out experiments and measured the costs of our proposed protocols.

REFERENCES

[1] Alliance for Telecommunications Industry Solutions. Homepage URL: http://www.atis.org.
[2] Amazon S3 Availability Event: (2008). URL: http://status.aws.amazon.com/s3-20080720.html (Accessed on November 29, 2012).
[3] AOL Apologizes for Release of User Search Data (2006). URL: news.cnet.com/2010-1030_3-6102793.html. August 7, 2006.
[4] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinsky, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., &Zaharia, M (2009). Above the Clouds: A Berkley View of Cloud Computing. Technical Report No.UCB/EECS-2009-28, Department of Electrical Engineering and Computer Sciences, University of California at Berkley. February 10, 2009.
[5] http://ieeexplore.ieee.org/document/7506226/?reload=true
[6] https://apprenda.com/library/paas/iaas-paas-saas-explained-compared/

[7] 7.https://www.abacusnext.com/blog/whats-difference-between-public-private-hybrid-and-community-clouds