# Data Location Verification in Cloud Computing

Abhishek Vaish, PhD.
Indian Institute of Information Technology
Allahabad, India

Abhishek Kushwaha
Indian Institute of Information Technology
Allahabad, India

Rahul Das
Indian Institute of Information Technology
Allahabad, India

Chandan Sharma
Indian Institute of Information Technology
Allahabad, India

## ABSTRACT
Cloud computing has pioneered the area of On-demand services. The customer can choose the resources according to the current needs with the facility of incrementing and decrementing the resources in the future. It generally follows a pay-as-you-use model which has proven beneficial for enterprises and individual users alike. Since the services are hosted over the internet, one of the recent concerns that are rising among the users is about the location of their data. Sometimes it is necessary for the data to stay in a particular jurisdiction. Therefore, it may be required for the organization to verify the location of their data from time to time. Here in this paper we propose a mechanism based on remote attestation technology of trusted platform module. Remote attestation technique is used to validate the current location of the data, and the generated result is passed to the user/verifier. The very fact that the trusted platform module is tamper proof provides the basis for the accuracy of the result.

## General Terms
Cloud computing, Information security, Security compliance.

## Keywords
Trusted Platform Module, Data security, Remote attestation, Data location.

## 1. INTRODUCTION
Cloud computing is gaining momentum day by day. With the increasing popularity, the issues associated with it are also getting more attention. Once blindly trusted, the abstraction provided by cloud computing is also under scrutiny due to its unreliable nature. In today's cloud environment, it is difficult to identify where the data is actually located, merely relying on Cloud Service Provider (CSP) for the information of location cannot be considered a wise idea. The problem is complex because even if the information provided at some point is correct, the CSP can very well, later, move the data deliberately or as a part of routine maintenance process. Thus, it becomes necessary that some procedure must be put in place which can be relied upon for the location information at any point in time. Due to highly virtual nature of the cloud architecture, apart from normal security practices additional mechanisms are required to ensure security and privacy of data. Data location in cloud computing is becoming a significant issue due to various legislations that prevent local data from being processed outside their jurisdiction. Also, data disclosure fears prevent enterprises from relocating business sensitive data, for example, US PATRIOT Act allows US government to monitor and access all data stored within US boundaries. Researchers have recently started taking interest in the concept of location identification to bring out some reliable mechanism to ensure legislative compliance. Simple IP address location verification cannot be considered an effective means since it is easy to relay packets from one location to another over secure channels, which are difficult to identify. In this paper, we propose a location identification technique that is based on Trusted Platform Module (TPM). TPMs are considered tamper-proof, and this forms the basis of accuracy and reliability of the information. The TPM is configured with the location co-ordinates at the time of installation, and it compares with the co-ordinates obtained at real time from a secure GPS device. If the values are within the permissible error range, the information is passed on to the verifier.

The paper is organized as follows: Section II discusses the concept of cloud and its security issues and features of TPM. Section III discusses some related works on location identification and use of TPMs in clouds. In section IV, we briefly go through data movement and regulatory issues associated with the cloud. Section V describes the proposed architecture for location verification. Section VI covers a brief conclusion.

## 2. CLOUD COMPTING and TPM

### 2.1 Cloud computing and its security issues
Cloud computing is probably the fastest growing field in the IT industry. It fulfills the need for high data processing and storage in a cost effective manner. The cloud solutions, most of the time, are scalable on demand. This flexibility provided by the cloud is itself the source of many strengths and weaknesses of cloud computing. Some of the problems associated with cloud computing are discussed here:

**Confidentiality:** Only authorized parties and the system can gain access to the protected data or system. In cloud computing data compromise is higher because of the large number of access points and number of parties, applications and devices are involved [1].

**Data confidentiality:** Data confidentiality can be achieved by encrypting the data before storing in the cloud. But performing computations on encrypted data is a difficult task and has little functionality. For performing rich computations on cloud data need to be in plain text, so data confidentiality becomes important [2].

**Software confidentiality:** Software used for computing purpose takes or handles the users' data/personal data as an input. Software and applications used in a cloud environment

and interaction methods with the users' personal data must be certified, so that leakage of data can be prevented [1].

**Privacy:** In a cloud environment location of data matters a lot. The organizations have to follow the countries legal framework requirements. In a cloud environment data can be stored in multiple locations in different data centers available at different regions or countries, which increases the risk of data confidentiality and privacy [1].

**Integrity:** It is one of the pillars of information security. In cloud computing integrity means that the data will remain as it is and there are no unwanted changes in the data.

**Data Integrity:** The data available in the cloud environment should be protected from unauthorized or unwanted deletion or modification. By implementing strong authentication mechanism organizations can achieve a level of confidence to obtain data and system integrity.

**Software integrity:** Unauthorized deletion of software and modification in the lines of code must be prevented. The alteration in the code can be done intentionally or unintentionally. An employee can change the code of software and make it perform something that it is not intended for.

**Availability:** The authorized user must be able to access his or her data available in the cloud. All the resources like software, hardware and data must be available once the user has logged in successfully. The system must be ready to perform operations associated with it and cloud owner must ensure that information and information processing are available to the user.

**Why location of data is important in cloud?** In a cloud environment data can be stored in multiple locations or data centers. If a user wants to verify the location of his or her data in the cloud he or she can not verify this data location information immediately. They can only request to the service provider for verification of the data. The verification result is based on the audit conducted by service provider or some third party. The duration between the request and reply can be undesirably large. This proposed solution provides architecture for verification of the location of data in the cloud environment in a remarkably short time.

## 2.2 TPM and Remote Attestation

Computer Security undoubtedly is the most important issue today with new vulnerabilities and attacks emerging every day. Several approaches have been proposed the address the security issues including software and hardware based. "Trusted Computing" model proposes a hardware based approach to address the security issues of today. The Trusted Computing Group or TCG has created the Trusted Computing specifications. The main purpose is to develop, define and promote open, vendor-neutral industry specifications for trusted computing. TCG created the Trusted Platform Module (TPM) to protect the integrity of the platform. TPM is based on cryptographic keys which enforces certain principles and resist any unwanted and unauthorized changes [3]. The main functions of TPM are platform monitoring, secure storage, encryption operations, and authentication services [4].

Majorly TPM has the following features:

- Binding
- Sealed Storage
- Secure Boot Process
- Remote Attestation

Binding is an interesting feature of TPM. Binding ensures that the data is stored encrypted. The key is generated from the RSA key burnt in the chip or a key generated from this key. This provides data security. Sealed Storage is similar to binding, but it ensures better protection. The user can also mention the state a TPM should be in for the decryption process. Sealed storage ensures data integrity and confidentiality. It can be implemented at the hardware level as well as multiple software levels. The access to the file stored is restricted only to a limited number of subjects. TPM verifies and authenticates any subject before authorizing their access. Secure boot process protects a system from any malicious activity during the process. It ensures that the OS is not compromised.

The most notable feature of TPM is Remote Attestation. It provides an architecture for verification of any resource by a third party in a secure way. Major security incidents have occurred where it has been found that software and applications were modified with a malicious intent. Verification of these platforms, applications and softwares has become very necessary, but without compromising security of other related data. Attestation must ensure an integrity check by some trusted and protected party. Remote attestation addresses these issues of mutual suspicion. Remote attestation provides a secure way through which any running software or application's integrity. It enables attestation and verification of the platform from trusted third parties or bodies. An important feature of remote attestation is that can verify the integrity and trust worthiness of any platform without revealing the identity of the platform, thus ensuring privacy and security of other information related to the platform [5]. The five principles of remote attestation (i) Fresh information, (ii) Comprehensive information, (iii) Constrained disclosure, (iv) Semantic explicitness, (v) Trustworthy mechanism, as explained by George Coker et al. in [6] form an integral part of attestation procedure in our approach.

# 3. RELATED WORK

Recently, some services like Amazon Web Services and Windows Azure lets users specify preferred locations of their data in Service Level Agreements (SLAs). However, a note is also included that lists some of many "factors" that can cause services to be delivered from other nodes [7]. In that case users have no means to regularly verify whether the CSP is fulfilling its contractual obligations related to geographical locations. In [8], Peterson et al. present one of the earliest works on the issue of data location assurance. Their methodology is based on the combination of MAC based PDP (Provable Data Possession) and any network delay based measurement geo-location protocol. In [9], Aiiad Albeshri et al. present another solution to location assurance. Here they have combined POS (Proof of Storage) scheme with timing based distance bounding protocol. This appears somewhat similar to [8], but here a Third Party Auditor (TPA) is present which verifies data on behalf of user, unlike [8] where user performs most of the computations. Ali Noman [10] proposes a Data Location Assurance Service (DLAS). The complete process consists of four phases and involves encryption and decryption. It reflects (at most) 24 hours' earlier state of their data location. Thorsten Ries et al. [11], present a solution based on Virtual Co-ordinate Network System (VCS) coupled with Global Positioning Networks (GNP) landmarks and RTT measurements.

Some related work to TPM, incorporating it in a cloud environment is also present as in [12], [13]. In [14], Dongxi Liu et al. presented the concept of virtual TPMs on a cloud platform which shows the possibilities on part of TPM that it can provide and over the period of time become one of the much needed utility just like encryption today, for security purposes. Our new approach combines the two fields to bring out a much needed solution to the problem of location assurance. Our solution is simple and is more beneficial to the

user due to lesser computations and high trust facilitated through the TPM.

# 4. DATA LOCATION AND RELOCATION

One of the generic advantages of cloud computing is the mobility, providing access from anywhere and even on the go. But in recent years of legislation becoming stricter, this mobility has come under the scanner. In many cases, the location of data does not bother the user. For example, the social networking site Facebook. It implements cloud computing [15]. A user can upload photos, videos and send messages to others and is not concerned as to where his/her data is residing. Similarly, e-mail services like Yahoo, Outlook, Gmail, etc. the user is certainly concerned for security but not particularly data location. However, when the user is an enterprise and has some sensitive data stored on the cloud, they may very well want to know the location of data and again in many cases, they may want to restrict to a particular location, be it US, UK or EEA [16].

Generally, data centers are located across the geographies to provide global services. The location of data center is based on many factors like customer base, operational cost, regulatory environment, safety concerns among others. All data centers are connected for to and fro movement of data for purposes like backup, recovery and load balancing, etc. [17]. Data can also be moved due to lack of own resources, resource expansion, cheap pricing policy, efficiency among others [16]. A major factor for deciding the location of data center is the operational cost, one of which is power consumption bill. The choice of location and data movement can be easily justified by the pressure to reduce cost and curb emissions. Google built a data center in Belgium that relies entirely on ambient cooling and during warm weather the servers were shut down [18].

## 4.1 Regulatory Concerns

Cloud computing involves the virtualization of resources and enables access over the internet. Since virtualization can involve the movement of data across geographies, concerns arise regarding the jurisdiction over the data. In the legal scenario, jurisdiction is dependent on location. Also, jurisdiction may not be exclusive [19]. Several countries may go into a dispute for jurisdiction over a matter. The movement of data generally crosses trans-borders and different jurisdictions have their own requirements regarding such movements. Now we see some such requirements briefly:

In EU, Article 29 Working Party analyses all relevant issues for CSPs operating in EEA and their clients specifying all applicable principles from EU Data Protection Directive (95/46/EC) and e-privacy Directive 2002/58/EC (as revised by 2009/136/EC). Transfer of personal data is permitted only if recipient country provides an "adequate level of protection" [20]. The Data Protection Act (1998) of UK makes a foreign company, that uses equipment located in UK for processing personal data, comply with the UK data protection law, even if the company is not established or does not do business in UK [19]. In Canada, the Office of the Privacy Commissioner states via Guidelines that an organization must ensure a "comparable level of protection", to the transferred personal information, from the recipient [21]. Similarly in Australia, the organizations must comply with the trans-border data flow requirements in Information Privacy Principle 9 – fulfill the requirements similar to Information Privacy Act of Australia [22]. In USA, the laws that influence data protection and privacy are Electronic Communications and Privacy Act (1986), PATRIOT Act (2001), Stored Communication Act among others. Under these acts, the federal agencies have the power to demand any data stored on any computer within USA [16], [23].

The jurisdiction problems associated with the cloud computing are due to the lack of harmony in the regulations which are further fuelled by the global nature of cloud computing. This has led to organizations opening data centers in the local jurisdiction.

# 5. THE PROPOSED ARCHITECTURE

## 5.1 Basics

At present scenario of cloud computing market, the users have to rely solely on CSP for information regarding their data on the cloud. The highly virtualized nature of the cloud makes it easy for the CSP to overlook few constraints that may be crucial from user's point of view and data location is one of them. Though some CSPs declare, in contract, as to where the data will be stored, and processed [7], the violations may occur and may very well go unnoticed. To protect itself, the user utilizes remote attestation technique to verify the location of the data in the cloud, and if violations occur, the user can contact the CSP for immediate correction.

A simple scenario can be presented as follows: The user has stored some data on the cloud. Some days later he/she wishes to verify the current location of the data. User sends a request to the CSP for the verification of location. The CSP forwards the request to all the datacenters and collects the replies from all the datacenters. The collected replies are then forwarded to the user. All the processes from request handling to sending reply are considered automatic. Simply contacting the CSP and asking for the location cannot be considered an accurate approach, since it will become more resource intensive and will involve more human intervention. To overcome this, a better and more automated approach is used. TPM based location verification eliminates human intervention from the verification process and is more trusted.

## 5.2 Proposed Architecture

The proposed architecture of remote attestation based location verification is shown in the fig. 1.

**Verifier:** This is any user or any other entity that wishes to verify the current location of the data stored on the cloud. Verifier just needs to send a request to the CSP, and the process completes with minimum human intervention.

**Request Processor:** The request processor receives all the requests from all users. It is a centralized facility, and in our case it is concerned with only location verification requests. The request processor sends verification request to all the data centers where it is further processed.

**Verification Module:** The verification module is not centralized like request processor but local to each datacenter. The verification module checks for the data in the datacenter, since it is a local entity, it can check within its own datacenter. If the data is found residing in the datacenter, the verification module takes the metadata, related to user data, and the real-time co-ordinates from GPS device and transfers it to TPM for attestation. It also sends back the attested result to result aggregator. Since it is local to each datacenter, for each request all the verification modules at each datacenter perform their required processing, and if no data is found the attested negative reply is sent from verification module.

**GPS device:** It is a secure GPS device which calculates the real-time co-ordinates of the location and provides it to the verification module. It is assumed to be secure in a sense that it is tamper proof and always provides accurate data within error limits.

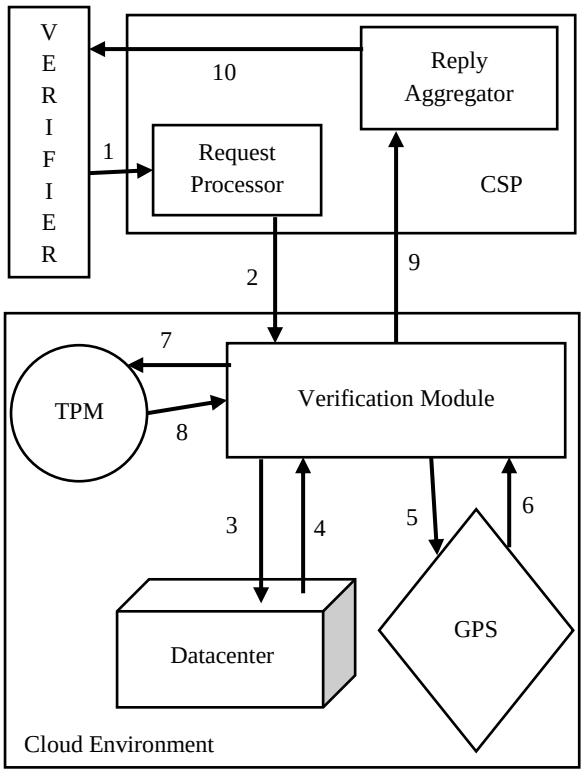


**Fig. 1. Data Location Verification System Architecture**

**Reply Aggregator:** It receives all the replies related to a particular request from all the datacenters through verification module. It combines the related replies and sends back to the verifier/user.

## 5.3 Location verification by remote attestation

The remote attestation of data location is described in detail in this section. Here, the cloud environment can be maintained by the CSP or can be provided by a third party, under contract. The process starts with a verifier who sends a request for the location verification.

(i). The verifier sends a request in a pre-defined format of CSP for verification. The request is handled by request processor.

(ii). The request processor forwards the request to all the datacenters of CSP located at various geographical locations. The request here id received by the verification module.

(iii). The verification module determines the files to be verified and then checks for them in the datacenter. If the data is found, the verification module takes the metadata related to the user data and obtains the location co-ordinates from the GPS device and concatenates them. The concatenated message is then sent to local TPM for attestation. If the files are not present, the attested negative reply is sent back from verification module. The purpose of attesting the negative reply is to make it trusted.

(iv). TPM receives the message from verification module and verifies the location co-ordinates. The TPM is itself configured with the location co-ordinates at the time of installation. If the co-ordinates provided with the message and those configured on the TPM match within permissible error limits then the TPM attests the message. It does not check for the metadata for here we are concerned with the integrity issue of the location and not the data itself. The attested message is returned to the verification module.

(v). The verification module returns the attested message to the reply aggregator which collects all replies related to a request from all datacenters and clubs them together.

(vi). The clubbed reply is returned to the verifier for analysis.

The functionality of verification module is critical for accurate measurements, and it has to be trusted. The CSP should take care to maintain the verification module's integrity, reliability and trust. The arrangements should also be in place to protect the replies from unauthorized disclosure. The use of TPM affords maximum accuracy and minimum time duration for the verification process.

# 6. CONCLUSION

Cloud computing is inevitable in the industry. Together with it, security and privacy are considered a major challenge to the cloud. Location verification eases the things a bit. Location assurance is required to comply with to comply with the SLAs and state legislations and to provide a heightened sense of security and privacy to the customer. In this paper, we proposed a remote attestation based location verification with more accuracy and in lesser time. The proposed architecture is designed to be suitable for real time cloud environments. With increasing use of TPM in enterprises, the architecture can prove to be most feasible in the cloud environment.

# 7. REFERENCES

[1]. Dimitrios Zissis and Dimitrios Lekkas, Addressing cloud computing security issues, Future Generation Computer Systems Vol. 28, No. 3, pp. 583–592, March 2012.

[2]. Krishana P.N.Puttaswamy, Christopher Kruegel and Ben Y.Zhao, "Silverline: Towards Data Confidentiality in Storage-Intensive Cloud Applications", Second ACM Symposium on Cloud Computing, Oct 2011.

[3]. Trusted Computing Group, (2013, February 2). Trusted Computing Group - Trusted Computing [Online] Available: http://www.trustedcomputinggroup.org/trusted_computing

[4]. Infineon Technologies AG. (2006, May 4). Infineon's Trusted Platform Module, [Online] Available: http://www.silicon-trust.com/trends/comp_tpm.asp

[5]. Bill Hewitt. (2013, February 2). Trusted Computing and the Trusted Platform Module: What All the Fuss Is About [Online] Available: http://www.cs.hmc.edu/~mike/public_html/courses/security/s06/projects/bill.pdf

[6]. George Coker et al., Principles of remote attestation, Int. J. of Info. Sec., Volume 10, No. 2, pp. 63-81, June 2011.

[7]. Windows Azure CDN Node Locations, [Online] Available: http://msdn.microsoft.com/en-us/library/windowsazure/gg680302.aspx

[8]. Zachary N. J. Peterson, Mark Gondree, Robert Beverly, "A Position Paper on Data Sovereignty: The Importance of Geolocating Data in the Cloud", HotCloud '11, Portland, OR, June 2011.

[9]. Aiiad Albeshri, Colin Boyd and Juan Gonzalez Nieto, GeoProof: Proofs of Geographic Location for Cloud Computing Environment, Proceedings of the 32nd International Conference on Distributed Computing Systems Workshops 2012, , Macau, China, pp. 506-514.

[10]. Ali Noman and Carlisle Adams, DLAS: Data Location Assurance Service for Cloud Computing Environments, Tenth Annual International Conference on Privacy, Security and Trust (PST), July 2012.

[11]. Thorsten Ries, Volker Fusenig, Christian Vilbois and Thomas Engel, Verification of Data Location in Cloud Networking, Fourth IEEE International Conference on Utility and Cloud Computing (UCC), Dec 2011.

[12]. Achemlal, M., Gharout, S., Gaber, C., Trusted Platform Module as an Enabler for Security in Cloud Computing, Conference on Network and Information Systems Security (SAR-SSI), pp. 1-6, 2011.

[13]. Bertholon, B., Varrette, S., Bouvry, P., Certicloud: A Novel TPM-based Approach to Ensure Cloud IaaS Security, IEEE International Conference on Cloud Computing (CLOUD), pp. 121-130, 2011.

[14]. Dongxi, Liu., Lee, J., Jang, J., Nepal, S., Zic, J., A Cloud Architecture of Virtual Trusted Platform Modules, 8th International Conference on Embedded and Ubiquitous Computing (EUC), pp. 804 – 811, 2010.

[15]. Tessa Finlayson, (2009, June 3) Cloud Computing & Data Protection [Online] Available: http://www.twobirds.com/English/News/Articles/Pages/Cloud_Computing_Data_Protection_030609.Aspx

[16]. Zaigham Mahmood, Data Location and Security Issues in Cloud Computing, International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), Sept 2011.

[17]. Confederation of Indian Industry, "The Indian Cloud Revolution", CII, New Delhi, 2012.

[18]. Will Knight, (2009, August 17), Energy-Aware Internet Routing [Online] Available: http://www.technologyreview.com/news/414771/energy-aware-internet-routing/page/2/

[19]. Francoise Gilbert. (2011, April 10) Server Location: A Significant Factor in Cloud Computing Services [Online] Available: http://www.francoisegilbert.com/2011/04/cloud-computing-legal-issues-data-location/

[20]. Data Protection Working Party, (2012 July) Article 29: Opinion 05/2012 on Cloud Computing [Online] Available: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf

[21]. Guidelines for processing personal data across borders, Office of the Privacy Commissioner of Canada, Canada, Jan. 2009.

[22]. Office of the Victorian Privacy Commissioner, 2011 May, Information Sheet: Cloud Computing, [Online] Available: http://www.privacy.vic.gov.au/privacy/web2.nsf/files/cloud-computing/$file/info_sheet_03_11.pdf

[23]. Zack Whittaker, (2011, April 26), Case study: How the USA PATRIOT Act can be used to access EU data [Online] Available: http://www.zdnet.com/blog/igeneration/case-study-how-the-usa-patriot-act-can-be-used-to-access-eu-data/8805