Review of Privacy Preserving Architectures in Cloud Computing

Navdeep Kaur Khiva
Deptt. Of Computer Science and Engineering
GZS PTU Campus
Bathinda, Punjab, India
navdeep07khiva@gmail.com

Amandeep Singh
Deptt. Of Computer Science andEngineering
GZS PTU Campus
Bathinda, Punjab, India

Pankush Singla
Deptt. Of Computer Science and Engineering
GZS PTU Campus
Bathinda, Punjab, India

Abstract— *Cloud computing is a technique in which data has to be transmitted across the internet from one point to another. Since the data is free of the transmission, there is a high possibility that the data sent across may be lost or be attacked by intruders, so the privacy among the data is to be maintained for secure transmission.*

Privacy in cloud computing is the ability of a user or a business to control what information they reveal about themselves over the cloud or to a cloud service provider, and the ability to control who can access that information. Numerous existing privacy laws impose the standards for the collection, maintenance, use, and disclosure of personal information that must be satisfied by cloud providers. Cloud Service Providers can store information at multiple locations or outsource it, then it is very difficult to determine, how secure it is and who has access to it.

Cloud computing is a revolutionary computing paradigm which enables flexible, on demand and low-cost usage of computing resources. Those advantages, ironically, are the causes of privacy problems, which emerge because the data owned by different users are stored in some cloud servers instead of under their own control. The privacy problem of cloud computing is yet to be solved effectively.

In this paper, I reviewed many research papers to check how many privacy preserving architectures exists and how they are preserving the data.

INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

The major issues of concern to Cloud Computing are the Security and Privacy issues. If the responsibility of providing services to the user is one task, the other important task is to ensure that these services are being provided while also ensuring the important features of privacy and security. Organizations and governments which try to implement their services using a cloud will have to implement it while considering the cloud as a social infrastructure.

Examples of cloud computing applications are Amazon's Simple Storage Service (S3), Elastic Computing Cloud (EC2) for storing photos on Smugmug and on line photo service, and Google Apps for word-processing.

Privacy in cloud computing is the ability of a user or a business to control what information they reveal about themselves over the cloud or to a cloud service provider, and the ability to control who can access that information. Numerous existing privacy laws impose the standards for the collection, maintenance, use, and disclosure of personal information that must be satisfied by cloud providers.

The nature of cloud computing has significant implications for the privacy of personal, business and governmental information. Cloud SPs can store information at multiple locations or outsource it, then it is very difficult to determine, how secure it is and who has access to it.

Cloud Features[1]

A. Identification & Authentication

In Cloud computing, depending on the type of cloud as well as the delivery model, specified users must firstly be established and supplementary access priorities and permissions may be granted accordingly. This process is targeting at verifying and validating individual cloud users by employing usernames and passwords protections to their cloud profiles.

B. Authorization

Authorization is an important information security requirement in Cloud computing to ensure referential integrity is maintained. It follows on in exerting control and privileges over process flows within Cloud computing. Authorization is maintained by the system administrator in a Private cloud.

C. Confidentiality

Confidentiality plays a major part especially in maintaining control over organizations data situated across multiple distributed databases. It is a must when employing a Public cloud due to public clouds accessibility nature. Asserting confidentiality of user's profiles and protecting their data, that is virtually accessed, allows for information security protocols to be enforced at various different layers of cloud applications.

D. Integrity

The integrity requirement lies in applying the due diligence within the cloud domain mainly when accessing data. Therefore ACID (atomicity, consistency, isolation and durability) properties of the cloud's data should without a doubt be robustly imposed across all Cloud computing deliver models.

E. Non-repudiation

Non-repudiation in Cloud computing can be obtained by applying the traditional e-commerce security protocols and token provisioning to data transmission within cloud applications such as digital signatures, timestamps and confirmation receipts services (digital receipting of messages confirming data sent/received).

F.Availability

Availability is one of the most critical information security requirements in Cloud computing because it is a key decision factor when deciding among private, public or hybrid cloud vendors as well as in the delivery models. The service level agreement is the most important document which highlights the trepidation of availability in cloud services and resources between the cloud provider and client.Maintaining the Integrity of the Specifications

PRIVACY

The American Institute of Certified Public Accountants (AICPA) and Canadian Institute of Charted Accountants (CICA) define that, "Privacy is the right and obligation of individuals and organizations with respect to the collection, use, retention, and disclosure of personal information".

The first problem with privacy is the disclosure of sensitive private information when exchanging data through the cloud service. And the sensitive private information includes: Personally identifiable information, Usage data, unique device identities and so on. The second problem is that people getting inappropriate or unauthorized access to personal data in the cloud by taking advantage of certain vulnerabilities, such as lack of access control enforcement, security holes and so on [2]. The third problem is that: because the feature of cloud computing is that it is a dynamic environment, in that service interactions can be created in a more dynamic way than traditional e-commerce scenarios. Services can potentially be aggregated and changed dynamically by service providers can change the provisioning of services. In such scenarios, personal sensitive data may move around within an organization or across organizational boundaries, so adequate protection of this information must be maintained despite the changes.

PRIVACY REQUIREMENTS

Pearson [3] mentions some key privacy principles that have to be met:

a) *Notice, openness and transparency:* during the collection of personal information it is important to tell users how their information is used. Once this way of using information is changed, the user should be notified about this.

   Privacy policies must be made available to the users.

b) *Choice, consent and control:* the user must have a choice about which information is collected. Data subjects must give their approval for collecting, using and disclosing their personal data.

c) *Scope/minimization:* only data that is required during the processes should be collected. The amount of data collected should be minimized.

d) *Access and accuracy:* users must have the ability to check which data is held and check its accuracy. All the personal information has to be accurate.

e) *Security safeguards:* safeguards must prevent unauthorized access, disclosure, copying, use and modification of personal data.

f) *(Challenging) compliance:* customers must be able to challenge the privacy process. All the operations that are performed on the data have to be compliant to privacy legislation.

g) *Purpose:* there must be a clear purpose for the collection of personal data. Data subjects should be told why their data is collected.

h) *Limiting use - disclosure and retention:* collected data should be used only for the purpose for which it was collected. Personal data should be stored as long as necessary.

i) *Accountability:* an organization must appoint someone who will be responsible for ensuring that privacy policies are followed.

PRIVACY TOOLS

A) PRIME:

Privacy and Identity Management for Europe (PRIME) provides privacy-preserving authentication using anonymous credentials. The user-side component uses protocols for getting third party (IdP) endorsements for claims to *relying parties* (RPs). Anonymous credentials are provided using an identity mixer protocol (based on the selective disclosure protocol) that allows users to selectively reveal any of their attributes in credentials obtained from IdP, without revealing any of their information. The credentials are then digitally signed using a public key infrastructure. A major limitation of PRIME is that it requires both user agents and SPs to implement the PRIME middleware, which hinders standardization.

B) Windows CardSpace

Windows CardSpace is a plug-in for Internet Explorer 7, in which every digital identity is a security token. A security token consists of a set of claims, such as a username, a user's full name, address, SSN etc. The tokens prove that the claims belong to the user who is presenting them. When a CardSpace-enabled application or website wishes to authenticate a user, it requests a particular set of claims from the user. The user selects an InfoCard to use among the ones visually presented to him, and the CardSpace software contacts an IdP to obtain a digitally signed XML token that contains the requested information, which is communicated to the requesting application.

The CardSpace framework is criticized due to its reliance on the user's judgment of the trustworthiness of an RP. Most users do not pay attention when asked to approve a digital certificate of an RP, either because they do not understand the importance of the approval decision or because they know that they must approve the certificate in order to get access to a particular website. RPs without any certificates at all can be used in the CardSpace framework (given user consent). Even if an RP presents a higher-assurance certificate, the user still needs to rely on an IdP providing that certificate to the RP, thus the user needs to trust the IdP. Another drawback is that, in a case where a single IdP and multiple RPs are involved in a single working session, (which we expect to be a typical scenario) the security identity metasystem within the session will rely on a single layer of authentication, that is, the authentication of the user to the IdP. If a working session is hijacked or the password is cracked the security of the entire system is compromised.

C) Open ID

OpenID is a decentralized authentication protocol that helps cloud users in managing their multiple digital identities with greater control over sharing of their PII. A user has to remember one username and password—an OpenID. She can log onto websites with this OpenID. She interacts with an RP that provides means to specify an OpenID for the authentication. The user has previously registered an OpenID with an OpenID provider (a TTP). Upon being discovered by the RP, the OpenID provider authenticates (commonly by prompting a password) and asks the user whether the RP should be trusted to receive the necessary identity details for the service. If she accepts, she is redirected to the RP along with her credentials, which need to be confirmed by the RP to provide service. After the OpenID has been verified, authentication is considered successful, and the user is considered logged in to the RP under the identity specified by the given OpenID. OpenID has been termed "phishing heaven" due to its susceptibility to phishing attacks and social engineering. A malicious attack can be easily set up to lure users into entering their authentication information at a website that poses as an OpenID provider website.

LITERATURE SURVEY

1. WANG et al. 2012 [4] explained, with cloud storage services, it is common place for data to be not only

stored in the cloud, but also shared across multiple users. However, public auditing for such shared data - while preserving identity privacy- remains to be an open challenge. Wang proposes the first privacy-preserving mechanism shown in figure 1 that allows public auditing on shared data stored in the cloud. There is an exploitation of ring signatures to compute the verification information needed to audit the integrity of shared data. With the mechanism, the identity of the signer on each block in shared data is kept private from a third party auditor (TPA), who is still able to verify the integrity of shared data without retrieving the entire file. ORUTA, the first privacy-preserving public auditing mechanism for shared data in the cloud was proposed. With ORUTA, the TPA is able to efficiently audit the integrity of shared data, yet cannot distinguish who is the signer on each block, which can preserve identity privacy for users.

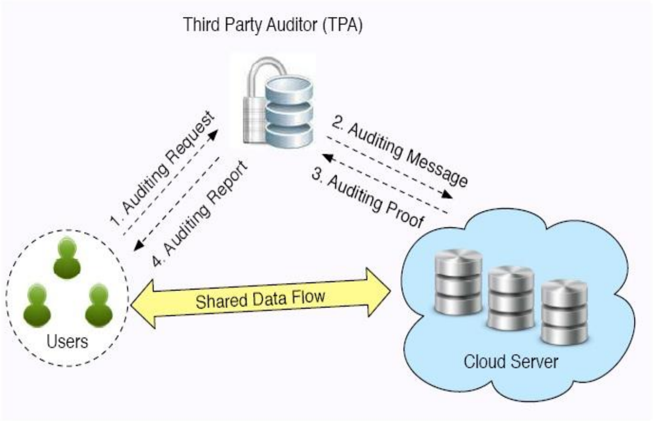

Figure 1: System Model includes the cloud server, the third party auditor and users.

2. BUGIEL et al. [5] explains Cloud Computing promises a more cost effective enabling technology to outsource storage and computations. Existing approaches for secure outsourcing of data and arbitrary computations are either based on a single tamper-proof hardware, or based on recently proposed fully homomorphic encryption. The hardware based solutions are not scalable, and fully homomorphic encryption is currently only of theoretical interest and very inefficient. Bugeil proposed architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud as shown in figure 2. In this approach, the user communicates with a trusted cloud (either a private cloud or built from multiple secure hardware modules) which encrypts and verifies the data stored and operations performed in the untrusted commodity cloud. They split the computations such that the trusted cloud is mostly used for security-critical operations in the less time-critical setup phase, whereas queries to the outsourced data are processed in parallel by the fast commodity cloud on encrypted data.

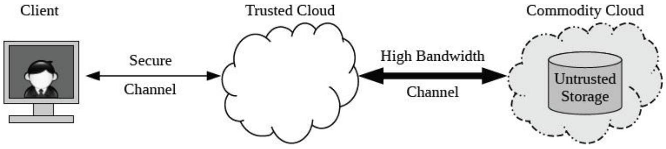

Figure 2: Trusted Cloud Architecture

3. HSUEH et al. [6] explains that care coordination services bring together a multitude of providers to deliver continuity of care outside clinical settings. The coordinated services improve wellness management and operational outcomes but pose challenges on privacy when integrating multiple sources of personal health data and providing a data access and sharing mechanism to third party providers. Hsueh particularly address the privacy challenges associated with data integration and sharing in a multi-tenant cloud environment for healthcare and present three care coordination use cases and detailing of the functional requirements across different stages of a personal data service cycle as shown in figure 3. Additionally, reflecting on technical challenges associated with privacy-preserving data integration and sharing, there is an introduction of a set of common data services to handle these issues, which ultimately lend support to the development of accountable coordinated care services. The recent paradigm shift from provider-centric to patient-centric services has resulted in a heightened level of privacy intrusion activity and in turn incurred protection initiatives across the world. Hsueh revisited the design of a cloud based platform, review its use in three major care coordination use cases, discuss requirements, and introduce protection mechanisms in three different stages of data lifecycle: data collection, data sharing and data integration for analytics.

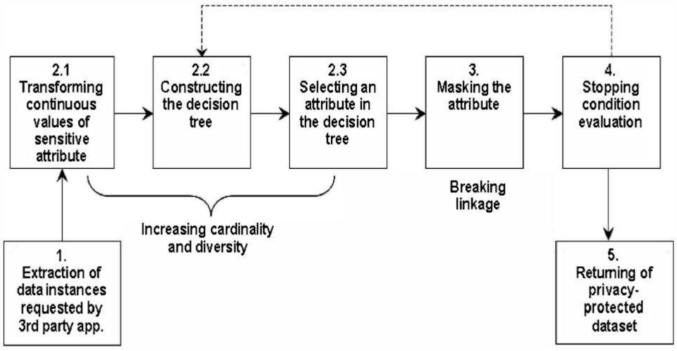
2.1 Transforming continuous values of sensitive attribute
2.2 Constructing the decision tree
2.3 Selecting an attribute in the decision tree
3. Masking the attribute
4. Stopping condition evaluation
Breaking linkage
Increasing cardinality and diversity
1. Extraction of data instances requested by 3rd party app.
5. Returning of privacy-protected dataset

Figure 3: Proactive re-identification protection mechanisms

4. CAO et al. 2011 [7] defines and solve the problem of privacy-preserving query over encrypted graph-structured data in cloud computing (PPGQ), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. The work utilizes the principle of "filtering-and-verification". Cao prebuild a feature-based index to provide feature-related information about each encrypted data graph, and then choose the efficient inner product as the pruning tool to carry out the filtering procedure(figure 4). To meet the challenge of supporting graph query without privacy breaches, they propose a secure inner product computation technique, and then improve it to achieve various privacy requirements under the known-background threat model. Thorough analysis investigating privacy and efficiency of their scheme is given, and the evaluation shows their scheme introduces low overhead on computation and communication.

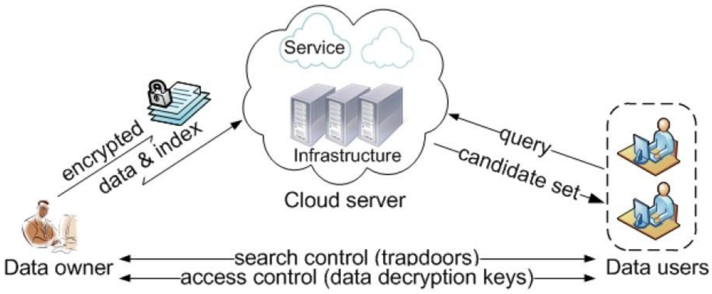

Figure 4: Architecture of graph query over encrypted cloud data

5. DROSATOS et al. 2012 [8] present a privacy-preserving system architecture for participatory sensing contexts which relies on cryptographic techniques and distributed computations in the cloud. Each individual is represented by a personal software agent, which is deployed on one of the popular commercial cloud computing services. The system enables individuals to aggregate and analyze sensor data by performing a collaborative distributed computation among multiple agents. No personal data is disclosed to anyone, including the cloud service providers. The distributed computation proceeds by having agents execute a cryptographic protocol based on a homomorphic encryption scheme in order to aggregate data.

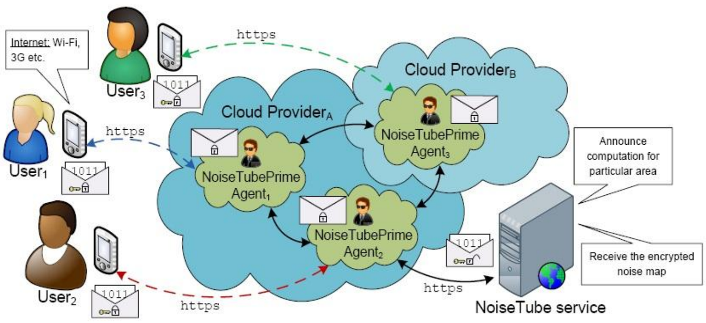

Figure 5: The general architecture of the system

Drosatos's show formally that their architecture is secure in the Honest-But-Curious model both for the users and the cloud providers. Their approach was implemented and validated on top of the NoiseTube system (figure 5) which enables participatory sensing of noise. In particular, they repeated several mapping experiments carried out with NoiseTube, and show that their system is able to produce identical outcomes in a privacy-preserving way. They experimented with real and simulated data, and present a live demo running on a heterogeneous set of commercial cloud providers. The results show that the approach goes beyond a proof-of-concept and can actually be deployed in a real world setting. To the best of their knowledge, this system is the first operational privacy-preserving approach for participatory sensing. While validated in terms of NoiseTube, their approach is useful in any setting where data aggregation can be performed with efficient homomorphic cryptosystems.

6. JAYALATCHUMY et al. 2010 [9] explains that Cloud computing is an On-demand self-service Internet infrastructure (figure 6) where a customer can pay and use only what is needed, managed by an API.

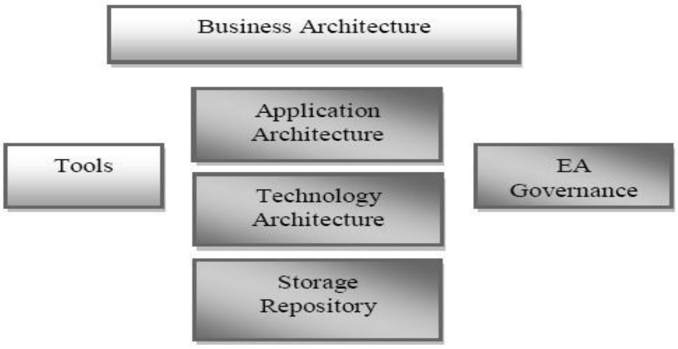

Figure 6: Cloud Computing Architecture

The SP plays an active role in transmitting information across the cloud. Privacy for the information through authentication is being considered important. Providing security requires more than user authentication with passwords or digital certificates. The discretion algorithm has been designed and the IDS provide passive security solution. Since the context data is stored by the service provider the control of the data propagate to the whole cloud chain.

7. RAHAMAN et al. 2012 [10] explains the widespread focus on the Cloud Computing which has necessitated the corresponding mechanisms to ensure privacy and security. Various attempts have been made in the past to safeguard the privacy of the individual or agency trying to utilize the services being provided by the cloud.

The most challenging task is to provide services to the users while also preserving the privacy of the user's information. Rahaman proposes a model that incorporates three-level architecture named "Preserving cloud computing Privacy (PccP)" (figure 7) which aims to preserve privacy of information pertaining to cloud users. The Consumer Layer deals with all the aspects which relate to enabling the user of the cloud to access the cloud services being provided by the cloud service provider.

The Network Interface Layer creates an appropriate mapping between the original IP addresses of the users with a modified IP address, and thereby ensuring the privacy of the IP address of the users. The Privacy Preserved Layer utilizes the functionality of the Unique User Cloud Identity Generator for which an algorithm is proposed to generate an unique User Service Dependent identity(USID) with privacy check by establishing mapping among the existing user identity(ID), if any to ID's available in a pool of User ID's to enhance the privacy of sensitive user information. A Privacy check method based on information privacy is being proposed which contributes significantly in maintaining user control over the generated user identities (USID's).

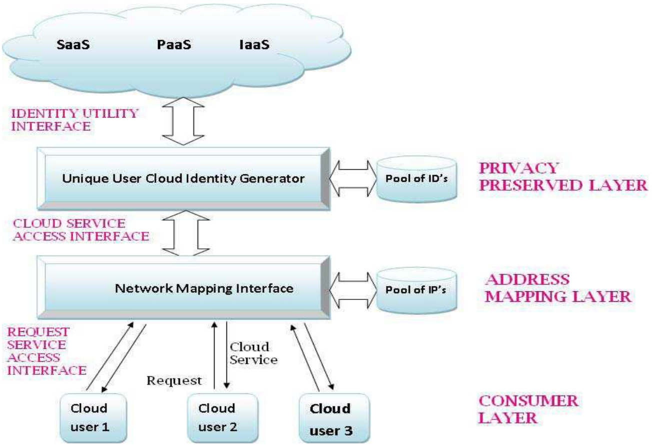

Figure 7: Preserving cloud computing Privacy (PccP) Model

8. KARUPPANAN et al. 2012 [11] explains about the Mobile social networks' (MSN) that diverse security concerns have immensely compromised users' personal details leaving them vulnerable to cybercrimes. karuppanan proposes an adaptive privacy architecture which provides content, identity and location privacy against disclosure of information that the user intends to keep private.

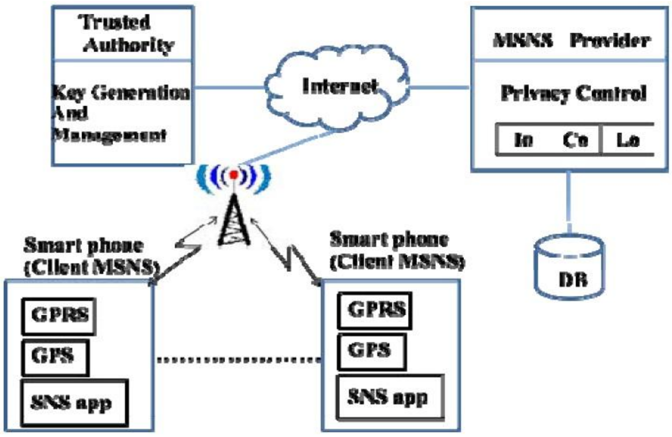

Figure 8: System Architecture of IoCoLo

The architecture (figure 8) implements multiple servers with the global social graph cached in the front-end server and sub-graph of each user across several servers. Content privacy is enforced through a role based access control model (RBAC) that relies primarily on the relationship and the trust between the users. Identity privacy is preserved through pseudonym generation. Location obfuscation is performed to safeguard the mobile user's location information. This ensures that the privilege and control is given to the user rather than a central authority namely social network service provider as suggested in the existing systems.

9. ANGIN et al. [12] tells that entities (e.g., users, services) have to authenticate themselves to service providers (SPs) in order to use their services. An entity provides personally identifiable information (PII) that uniquely identifies it to an SP. In the traditional application-centric Identity Management (IDM) model, each application keeps trace of identities of the entities that use it.

In cloud computing, entities may have multiple accounts associated with different SPs, or one SP. Sharing PIIs of the same entity across services along with associated attributes can lead to mapping of PIIs to the entity. again proposes an entity-centric approach for IDM in the cloud. (Figure 9) The approach is based on: (1) active bundles:- each including a payload of PII, privacy policies and a virtual machine that enforces the policies and uses a set of protection mechanisms to protect themselves; (2) anonymous identification to mediate interactions between the entity and cloud services using entity's privacy policies. The main characteristics of the approach are: it is independent of third party, gives minimum information to the SP and provides ability to use identity data on untrusted hosts.

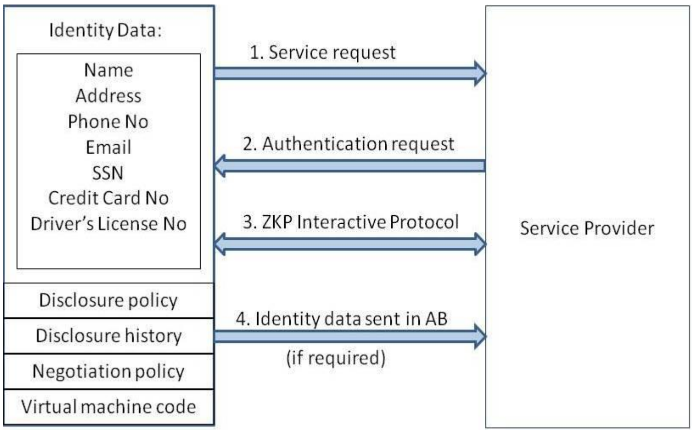

Figure 9: Structure of IDM Wallet

10. XIAO 2010 [13] states that Cloud computing provides a novel computing paradigm for enterprises to store programs and data in the Cloud in a transparent manner, which poses the challenge of security and privacy.

In this paper, based on homomorphic cryptography and Zero-Knowledge Proof, Xiao present a novel privacy-preserving scheme for Cloud publish/subscribe service, which achieve efficient privacy-preserving authentication, data integrity, and publish-subscribe confidentiality. (Figure 10) The performance evaluation and security analysis demonstrate the practice and validity of the proposed scheme.

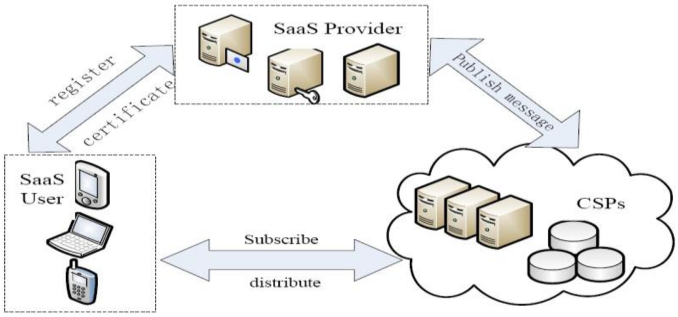

Figure 10: Message Publishing and Subscribing Architecture

CONCLUSION

Preserving privacy is one of the biggest challenges in Cloud Computing. Cloud data must be protected not only against external attackers, but also corrupt insiders. With the immense growth in the popularity of cloud computing, privacy have become important concerns for both the public and private sectors. It is very likely that users end up having multiple identities in cloud service providers, security repositories, as well as multiple credentials and multiple access permissions for different services provided by different CSPs. There is a strong need for an efficient and effective privacy-preserving system.

REFERENCES

[1] S Ramgovind, MM Eloff, E Smith, "The Management of Security in Cloud Computing", 2010 IEEE.

[2] D Jayalatchumy, P Ramkumar, and D Kadhirvelu, "Preserving Privacy through Data Control in a Cloud Computing Architecture using Discretion Algorithm", Third International Conference on Emerging Trends in Engineering and Technology, 2010 IEEE, DOI 10.1109/ICETET.2010.103.

[3] Siani Pearson. "Taking account of privacy when designing cloud computing services" , Cloud '09, May 2009.

[4] Boyang Wang, Baochun and Hui Li Wang, "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud", 2012 IEEE Fifth International Conference on Cloud Computing, 2012 IEEE, DOI 10.1109/CLOUD.2012.46.

[5] Sven Bugiel, Stefan Nurnberger, Ahmad-Reza Sadeghi, Thomas Schneider, "Twin Clouds: An Architecture for Secure Cloud Computing".

[6] Pei-Yun Hsueh, Tyrone Grandison, Liangzhao Zeng, XinXin Zhu, Ci-Wei Lan, lenHao Hsiao, and Henry Chang, "Privacy Protection for Personal Data Integration and Sharing in Care Coordination Services", A Case Study on Wellness Cloud.

[7] Ning Cao, Zhenyu Yang, Cong Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Query over Encrypted Graph-Structured Data in Cloud Computing", 2011 31st International Conference on Distributed Computing Systems, 2011 IEEE, DOI 10.1109/ICDCS.2011.84.

[8] George Drosatos, Pavlos S. Efraimidis, Ioannis N. Athanasiadis, Ellie D'Hondt and Matthias Stevens, "A privacy-preserving cloud computing system for creating participatory noise maps", 2012 IEEE 36th International Conference on Computer Software and Applications, 2012 IEEE DOI 10.1109/COMPSAC.2012.78.

[9] D Jayalatchumy, P Ramkumar, and D Kadhirvelu, "Preserving Privacy through Data Control in a Cloud Computing Architecture using Discretion Algorithm", Third International Conference on Emerging Trends in Engineering and Technology, 2010 IEEE, DOI 10.1109/ICETET.2010.103.

[10] Syed Mujib Rahaman and Mohammad Farhatullah, "PccP: A Model for Preserving Cloud Computing

Privacy", 2012 International Conference on Data Science & Engineering (ICDSE), 2012 IEEE.

[11] Komathy Karuppanan, K AparnaMeenaa, K Radhika, and R Suchitra, "Privacy Adaptation for Secured Associations in a Social Cloud" **,** 2012 International Conference on Advances in Computing and Communications, 2012 IEEE, DOI 10.1109/ICACC.2012.45.

[12] Pelin Angin, Bharat Bhargava, Rohit Ranchal, Noopur Singh, Lotfi Ben Othmane, Leszek Lilien and Mark Linderman, "An Entity-centric Approach for Privacy and Identity Management in Cloud Computing".

[13] Yanping Xiao, Chuang Lin, Yixin Jiang, Xiaowen Chu and Fangqin Liu, "An Efficient Privacy-Preserving Publish-Subscribe Service Scheme for Cloud Computing", 2010 IEEE.