

Security and Privacy of Data in Cloud Computing



Amar Buchade, Rajesh Ingle

Abstract: The cloud computing paradigm is being used because there is no need to setup additional IT infrastructure such as hardware and software, its low up-front cost. Security and privacy of data is important in day to day life especially for applications that uses cloud computing such as social media. Customer information that is stored at Cloud is crucial that needs to be protected against potential intruders. There is threat to maintain the data in transit and data at cloud due to different possible attacks. Due to this there is growing need of privacy and security of data. In this paper, the privacy and issues, privacy preservation techniques are addressed. In addition to this, in order to protect the data, the secret sharing algorithm is implemented and analyzed. The shamir's secret sharing (k,n) algorithm is used to split the data into n partial shares which can be distributed in cloud. The user collects at least k partial shares to reconstruct the complete data. It is observed that if the file size is increased, the data recovery time is also increased. The paper concludes with privacy preservation guidelines.

Index Terms: privacy, security, access control, identity management, homomorphic encryption, anonymity

I. INTRODUCTION

Now a day, e-government services are accessible to the citizens from his locality through common service delivery interfaces. This is possible by the use of Cloud computing technology. The advantage of using this technology is that the complete IT infrastructure need not be set up by the government. Cloud enable environments can handle large number of transactions due to its elasticity in nature. For example, simple tasks like booking a rail tickets, birth certificate, and death certificate online through the cloud service transparently.

Many software industries use Cloud for data storage and computation at Cloud service provider. Because it's on demand self service, rapid resource elasticity, relief of burden for storage management, universal data access with independent geographical locations, cheap, doesn't require installation/maintenance, requires no environmental conditions and doesn't require energy for power or cooling. Cloud platform can be deployed as private cloud, public, community and hybrid cloud. Cloud has three service

models: Software as Service, Platform as service and Infrastructure as service [1].

Figure 1 illustrates control of cloud service provider (CSP) and cloud consumer (CC) for each service models. The arrows at the left and right of the figure indicate range of the cloud providers and cloud consumers in the cloud environment for each service model respectively. Resources layer includes computers, network, storage components and other physical computing infrastructure elements. e.g. In SaaS, CSP has full control over on resources, hypervisor, application development environment as well as application but CC has only control on application layer. From figure 1, it is also observed that cloud consumer no longer control on the storage of data. So in traditional approach there was complete control of data to the user side but now data and code control with cloud service provider. CSP may reuse allocated storage space data of consumer for financial benefit. There is an increasing trend towards the use of social networks such as facebook, linkedin etc. The data of these social networks stored in cloud data centre. Recently users which were using facebook, their important data is leaked [2]. There is need of privacy and security of data and only intended user should access the data.

Rest of the paper is organized as defining privacy, security and privacy preservation in section II, privacy preservation techniques in section III, Shamir's secret sharing algorithm in section IV. The section V describes results and analysis of Shamir's secret sharing algorithm. Finally the paper is concluded with guidelines of privacy preservation in Section VI.

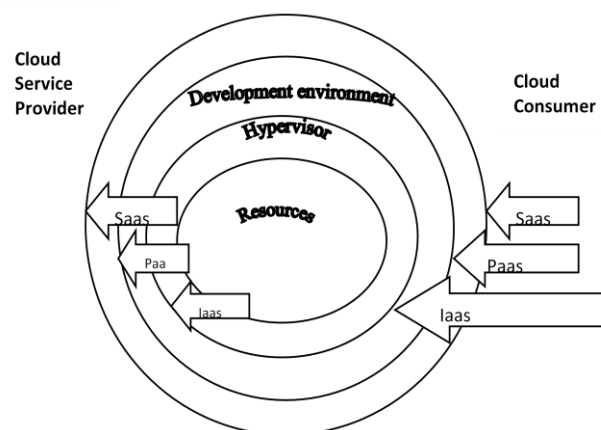


Figure 1: Cloud Computing Model

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Amar Buchade*, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India.

Rajesh Ingle, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

II. PRIVACY, SECURITY AND PRIVACY PRESERVATION

The privacy is the ability of an individual or group to keep apart from others about themselves or information. e.g with respect to wealth information, privacy is defined as the right of an individual to keep his/her individual wealth information from being disclosed. Privacy encompasses controlling who is authorized to access wealth information. Security is defined as mechanism in place to protect the privacy of wealth information as well as safeguard information from unauthorized disclosure, alteration, loss or destruction.

Privacy preservation is the technique for not disposing the information to others. e.g personal information, user identity, sensitive data, financial information, location of user's data, device identity, shares market data.

There are various reasons by which privacy is not preserved such as identity theft, internal employee problem at cloud service provider side, laws and regulation, data storage redundantly in multiple physical locations, security through third party, exposure of data to third party, web browser attack, weak password, tampering of data, sensitive data leak, hypervisor attack, multi-tenancy, loss of control of data, lack of trust, keeping data on public cloud,, e.g. Amazon's S3 online storage service has experienced significant downtime [3], Google App Service outage [4], hacker used Amazon's Elastic Computer Cloud, or EC2, service to attack Sony's online entertainment systems[5], criminals attacked Salesforce.com and steal customer emails and addresses using phishing attack.

In multitenancy cloud environment, tenants share single instance of software (SaaS) or infrastructure resources such as hardware, compute server and data storage device. Risks associated with the multitenancy must be addressed [6].

III. TECHNIQUES FOR SECURITY AND PRIVACY OF DATA

A. Access control policy

Access control to resources should be based on the policy and user's data protection mostly required in multitenant environment such as cloud. Intruders getting unauthorized access to personal data in cloud hence there is need of strong access control policy to access the data in cloud [7].

B. Identity and access management (IAM)

It gives the protection of cloud resources through rules and policy. IAM includes authentication, authorization and auditing. Authentication is the validation of credentials provided by the user. (e.g ID, Password). Credential management, strong authentication (One time password, multifactor authentication), ID privacy required to protect the resources in the cloud. One such algorithm is anonymous authentication based on public key cryptography technique without revealing identity of consumer to untrusted computing node [7][8].

Authorization is the access of specific resources decided by access control policy which is after successful authentication. Privacy preserving authorization based on Policy to protect the privacy of users data is discussed in [9]. The challenge is the authorization in the multitenant environment of cloud.

Auditing is the process of examining authorization and authentication record to detect the security breach. One approach is the use of external audit party deals with verifying data in cloud and log [10].

C. Homomorphic encryption

In this approach, user can store the data in encrypted form. User is able to carry any arbitrary computation on without any understanding from other cloud users

```
e.g .C1=Encrypt(15,"country") // rdjcign
      C2=Encrypt(15,"asia") //phxp
      C3=concatenation (C1,C2) // rdjcignphxp
      Var plaintext=decrypt(13,C3); //countryasia
```

Here 15 is the key which is applied to string for shifting the character by 15 times. C1, C2 are sent to cloud by the consumer, C3 is result of concatenation of C1, C2 that is done at cloud side and sent to the user.

Homomorphic encryption scheme allows the transformation of ciphertexts of message m, C(m) to ciphertext of a computation/function of message m, C(f(m)) without disclosing the message.

Fully homomorphic encryption scheme:

$E(m1 \oplus m2) \leftarrow E(m1) \oplus E(m2); m1, m2 \in M$

Where M is the set of plaintext, \oplus represents any arbitrary function and \leftarrow means computation is done without the plaintexts being decrypted [11].

D. Anonymity

It is often necessary to publish personal information for research purposes. For example hospital may release patients diagnosis records so that researchers can study the characteristics of various diseases. The raw data also called micro data contains e.g names of individuals and some information which should not be released to protect their privacy [12].

Other example College has student data in the office. Such each record contains name, age, address, branch and year of admission. Consider the tuples in Table I.

College anonymized the student record on attributes branch and year of admission called quasi identifiers before data can be released to the cloud service provider. Suppose 2-anonymity is required. Table II shows 2 anonymous release of records with respect quasi identifiers (name, branch and year of admission). And then college anonymized the student records on attributes student name, address and year of admission.

TABLE I: STUDENT RECORDS

Student Name	Age	Address	Branch	Year of admission
Sagar	20	Sion (W)	Computer Technology	2007
Suhass	21	Dadar(E)	Computer Management	2006
Amrit	22	Pawai(E)	Electronics and Telecommunication	2005
Yatin	20	Vikroli(W)	Electronics	2007

TABLE II : A 2-ANONYMOUS RELEASE OF TABLE I WITH RESPECT TO QUASI ATTRIBUTES(NAME,BRANCH,YEAR OF ADMISSION)

TABLE III : A 2-ANONYMOUS RELEASE OF TABLE I WITH RESPECT TO QUASI ATTRIBUTES(NAME,ADDRESS, YEAR OF ADMISSION)

Student Name	Age	Address	Branch	Year of admission
Sagar	20	Sion	Computer Technology	2007
Suhas	21	Dadar	Computer Management	2006
Amit	22	Pawai	Electronics and Telecommunication	2005
Yatin	20	Vikroli	Electronics	2007

Now cloud service provider contains two Tables II and III. But suppose University want this information of student. It can take these two tables and analyze it that Sagar with age 20 having computer technology branch, year of admission 2007 staying at Sion(W). So when we publish anonymous data, we should consider multiple quasi-identifiers (QI) attributes for different agencies.

E. Client based approach privacy preservation

By this approach, user can control their sensitive information to be sent to cloud [13]. There are techniques such as obfuscation, preference setting, data access, feedback and personae.

Encryption and Decryption – It is done by the user having key. Encryption and decryption is done at the client side (obfuscate). The cloud service provider and attacker are unable to deobfuscate the data because of non availability of key Preference setting – Privacy policy along with data is sent with encryption and decrypted at receiver side.

Data access – Cloud service provider may able to make the access of specific data to the user for audit purpose decided in service level agreement.

Feedback – This approach can give feedback of personal information and data usage in the cloud.

Personae – This feature allows the user to choose multiple personae when interacting with cloud services like anonymous user or full disclosure of identity.

F. Implementation of fault free API

Cloud services are accessed through third-party access by exposing application programming interfaces. Developers and customers do not use properly API functions to access the cloud and their data. Cloud and Web service developers must first follow best practices in opening up their APIs to third parties. In return, third-party developers need to handle the API functions in a secure manner. Attacker may use these keys and can cause a denial of service [14].

G. Middleware based approach for privacy preservation

Active bundle approach – Active bundle [15] is middleware agent that composes sensitive data, identity data, privacy policies, integrity check metadata, access control metadata, server id, trust level threshold to access data in active bundle. Active bundle can be sent from source host to destination host. When arriving at destination host, an active bundle ascertains host’s trust level by approaching to third party and decides whether host is eligible to access or part of data. If data reaches to unintended destination, it destroys sensitive data included in the active bundle.

Student Name	Age	Address	Branch	Year of admission
Sagar	20	Sion (W)	Computer	2007
Suhas	21	Dadar(E)	Computer	2006
Amit	22	Pawai(E)	Electronics	2005
Yatin	20	Vikroli(W)	Electronics	2007

H. Hardware based approach for privacy preservation

Trusted platform module (TPM) – TPM [16] is cryptoprocessor chip ensures that only intended and authorized users/systems access the information while limiting accessibility of system administrators, root processes along with the other users.

I. Threat model

There are following threats that needs to be considered to preserve privacy in the cloud such as attackers, Identity theft, Tampering of data, repudiation, Information disclosure, denial of service, privilege theft by unauthorized user. Threat model is discussed in [17].

J. Privacy preserving audit of digital content by third party

With the cloud, users can remotely store their data into the cloud. The situation is that users no longer have physical possession of outsourced data makes the data integrity protection in very challenging for the users with limited resources and capabilities. Some researchers have given third party auditor with maintaining privacy of user to periodically verify the data stored by a service [18].

Some researchers also work on public verifiability without help of a third party auditor with guaranteed that protocol does not leak any private information to third party verifiers.

In figure 2, user transfer certain number of blocks with metadata information such as its size for calculation of MAC (message authentication code) by the third party auditor [TPA] and TPA also get the certain number of blocks and calculates the MAC.If both the MAC are same, data is intact at cloud service provider side.

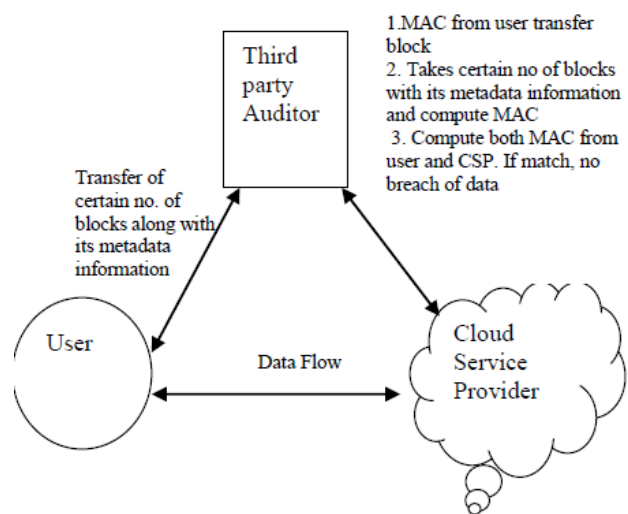


Figure 2: Auditing of data in cloud

K. Monitoring of Service level agreement (SLA)

As cloud is emerging technology, more and more companies deploying applications and data at cloud. Hence service-level agreement is very important aspect for providing service to the user. It is negotiated agreement between company/customer and CSP. Enterprise customers are outsourcing its critical data at cloud. There is need of monitoring of SLA for effective provision of service by CSP. In paper [19], mechanism for managing SLA in cloud environment using the Web Service Level Agreement (WSLA) framework developed for SLA monitoring and SLA enforcement in a Service Oriented Architecture (SOA) is given. Third party support feature of WSLA is used to delegate monitoring and enforcement tasks to other entities in order to solve the trust issues.

L. Obfuscation with Noise Injection approach

Along with the real request, noise requests are sent to cloud service provider to confuse about real request. The paper [20] provides a new historical based noise generation strategy so that requests including noise and real one reach about same occurrence probability. This is causing to service provider not able to distinguish between real and noise request.

M. Security and privacy issues in virtualization

There are several attacks possible such as SQL injection, Guest OS user to run code on host or another Guest OS e.g vulnerability in Microsoft Virtual PC and Microsoft Virtual Server attack [21], two VMs communicating with each other and sharing data, DDOS attack on VM, Data leakage, Data remanence. The paper [22] proposes the architecture based on security and reliability monitor units at VM side (VSEM and VREM) and at Hypervisor (HSEM and HREM).

N. Combing various approaches for securing data in cloud

In the paper [23], user calculates sensitivity rating [SR] of data and SR is used to allocate the data to one of sections in Cloud i.e Public or Private or Owner's limited access. Data is stored in encrypted form on Cloud. Message authentication code is used to check the integrity of data. For retrieval of data by any user double authentication by owner and cloud as well as verification of digital signature of owner. Thus this solves issues like tampering of data, unauthorized access.

O. Session key negotiation for fast and secure scheduling of scientific applications in cloud computing

Data generated by scientific applications is very important. In cloud there can be server instances. Thus there is need to exchange the key between cloud controller (CLC) and server instance. The paper [24] gives authenticated key exchange protocol and uses the symmetric key algorithm to encrypt the data at each server instance.

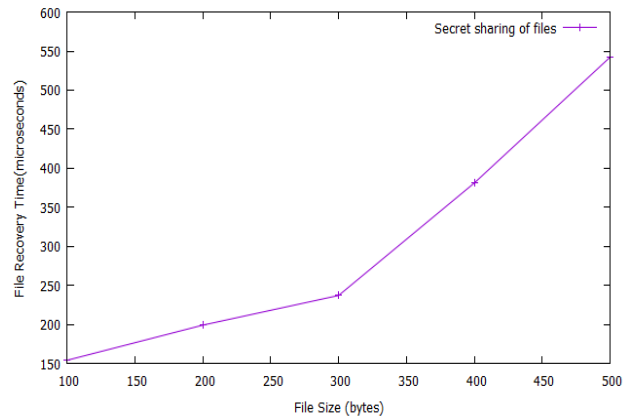
IV. SHAMIR'S SECRET SHARING ALGORITHM

Shamir's secret sharing (k,n) algorithm [25] is used to split the data into multiple shares of data. It is threshold cryptography algorithm. In this, k shares out of n splits can be grouped together to form the complete data. These multiple shares can be stored in virtual machines of cloud computing environment. Even if attacker grabs the control on (k-1) virtual machines, he/she is unable to get the complete data. At least k number of virtual machines needs to be compromised to get the k data splits from the cloud computing

environment. Data owner splits the data into multiple shares and stores in cloud environment to be accessed by the users. Users in turn accesses at least k shares from the cloud environment and reconstruct the data.

V. RESULTS OF SHAMIR'S SECRET SHARING ALGORITHM

The shamir's secret sharing algorithm (k,n) is applied to data of various sizes. The file is splitted into 3 shares. The value of k is taken as 2. Users take at least k shares to recover the complete data. From the analysis, it is observed that as the size of the file increases, file recovery or reconstruction time is also increased.



VI. GUIDELINES FOR PRIVACY PRESERVATION

The guidelines for privacy preservations are as below.

- Minimize personal information sent to and stored in the cloud.
- Personal information has to be protected from any lost or theft created by intruders. Employees or third parties should only be given access to information they need to fulfill their business purpose.
- Assign appropriate control policy.
- Users must be allowed with a choice whether they want to share their information or not.
- Limit the purpose of data usage- When the information is loaded into the cloud, it must be limited to the preferences and conditions set by user or organization. Data usage has to be restricted to the uses specified purpose.
- Provide feedback - Cloud applications should be user friendly and clear indicate privacy functionality by using icons, providing tutorials, help documents. Applications need to be designed in a way that users are provided with feedback, allowing them to make decision in terms of privacy.

VII. CONCLUSION

Customer information is very sensitive and important that needs to be protected from attackers. Hence there is need of privacy. In this paper, various security and privacy preservation techniques mentioned to minimize the risks by the cloud service provider and attackers. The user can use one of the techniques as per requirement and security level.

The Shamir's secret sharing (n,k) algorithm is used to protect the data from the attackers. The attacker has to compromise at least k number of virtual machines of cloud computing environment. It is also observed that as the data size increases, data recovery time is also increased.

REFERENCES

- Wayne Jansen, Timothy Grance, "Guidelines on Security and Privacy in Public Cloud Computing", src.nist.gov/publications/nistpubs/800-144/SP800-144.pdf, December 2011
- Glaser, "Facebook data breach", <https://slate.com/technology/2019/04/facebook-data-breach-540-millions-users-privacy.html>, April 2019.
- Amazon Web Services Outage, "https://istheservicedown.in/problems/amazon-web-services-aws/history", 2019
- App Engine Team, "App engine outage", groups.google.com/group/google-appengine/browse_thread/thread/a7640a2743922dcf, Google-appengine Group, 2010
- Pavel Alpeyev, Sony Attack, "Amazon.com Server Said to Have Been Used in Sony Attack", <http://www.bloomberg.com/news/2011-05-13/sony-network-said-to-have-been-invaded-by-hackers-using-amazon-com-server.html>
- Robby Higgins, "Securing a multi-tenant environment", <http://searchcloudsecurity.techtarget.com/tip/Securing-a-multi-tenant-environment>, Juniper networks 2012
- Cloud Security Alliance, "Domain 12: Guidance for Identity & Access Management V2.1", <https://cloudsecurityalliance.org>, Apr 2010
- Safwan M. Khan and Kevin W. Hamlen, "{AnonymousCloud}: A Data Ownership Privacy Provider Framework in Cloud Computing", in Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) for the coming June 2012
- David W. Chadwick, "A privacy preserving authorisation system for the cloud" in Journal of Computer and System Sciences, 2011.
- Tengfei Tu, Lu Rao, Hua Zhang, Qiaoyan Wen, and Jia Xiao, "Privacy-Preserving Outsourced Auditing Scheme for Dynamic Data Storage in Cloud", Volume 2017, Pages 1-17, 2017
- Craig Gentry, "Fully Homomorphic Encryption Using Ideal Lattices", in STOC'09, May 31–June 2009
- ZakariaeEl Ouazzania, HananEl Bakkali, "A new technique ensuring privacy in big data: K-anonymity without prior value of the threshold k", Procedia Computer Science, 2018
- Miranda Mowbray, "client based privacy manager for cloud computing", COMSWARE'09, June 16–19, 2009, Dublin, Ireland
- Robert Lemos, "Insecure API Implementations Threaten Cloud", <http://www.darkreading.com/cloud-security/167901092/security/application-security/232900809/insecure-api-implementations-threaten-cloud.html> - API, 23 apr 2012
- Rohit Ranchal, "An Approach for Preserving Privacy and Protecting Personally Identifiable Information in Cloud Computing"
- Jeffrey Naruchitparame, "Enhancing data privacy and integrity in the cloud" in International Conference on High Performance Computing & Simulation - HPCS, 2011
- Priya Metri, "Privacy Issues and Challenges in Cloud computing", in International journal of advanced engineering sciences 2011
- Cong Wang, "Privacy-Preserving Public Auditing for Data Storage", in Security in Cloud Computing 2010 IEEE
- G. Justy Mirobi, L. Arockiam, "Service Level Agreement in cloud computing: An overview", 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies
- Gaofeng Zhang, Yun Yang, "A historical probability based noise generation strategy for privacy protection in cloud computing", in Journal of Computer and System Sciences 2012.
- Michael Pearce, Sherali Zeadally, Ray Hunt, "Virtualization: Issues, Security Threats, and Solutions", ACM Computing Surveys, Vol. 45, No. 2, Article 17, February 2013
- Farzad Sabahi, "Secure Virtualization for Cloud Environment using Hypervisor based Technology", in International Journal of Machine Learning and Computing, Vol. 2, No. 2, February 2012.
- Sandeep K. Sood, "A combined approach to ensure data security in cloud computing", in Journal of Network and Computer Applications 2012.
- Chang Liu, Xuyun, Chi Yang, "CCBKE-Session key negotiation for fast and secure scheduling of scientific application in cloud computing", in Future Generation Computer Systems 2012

- Adi Shamir. How to share a secret. Communications of the ACM, 22(11):612–613, 1979

AUTHORS PROFILE



Amar Buchade, Ph.D. in Computer Engineering from College of Engineering, Pune. He has completed B.E. and M.E. Computer Engineering from Walchand College of Engineering, Sangli. His research area is distributed system, computer network, cloud security and internet of things. He has published more than 35 papers in national, international journals and conferences. He is IEEE member and Life member ISTE. He is secretary, IEEE Pune Section. He has completed NPTEL courses such as cryptography and network security, cloud computing, blockchain usecases and design, intellectual property rights, patent design for beginners and internet of things.



Rajesh Ingle is a Dean, Head, and Professor, Department of Computer Engineering at Pune Institute of Computer Technology. He has received a PhD in Computer Science and engineering from Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Powai, Mumbai. He has received the BE Computer Engineering from Pune Institute of Computer Technology, University of Pune, and ME in Computer Engineering from Government College of Engineering, Savitribai Phule Pune University. He has also received an MS in Software Systems from BITS, Pilani, India, in 1994. He is a senior member of the IEEE, IEEE Communications Society, and IEEE Computer Society.