Cloud Computing (CS60118)

(Spring 2020-2021)

Federation, Presence, Identity and Privacy in Cloud

Dr. Sudip Misra
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/
Research Lab: cse.iitkgp.ac.in/~smisra/swan/

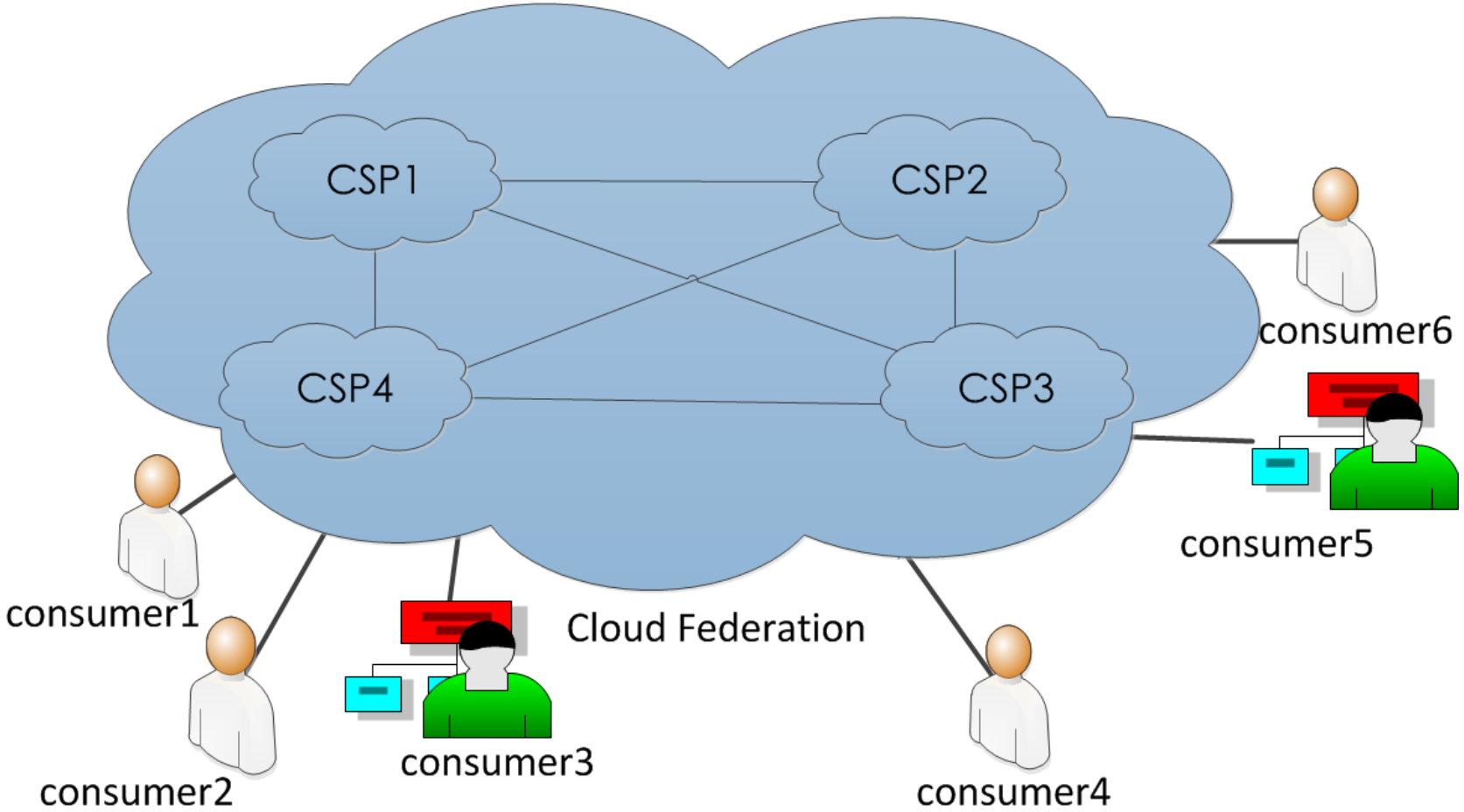
Contents

- Cloud Federation
- Privacy in Cloud

Cloud Federation

- Cloud federation is a geographically dispersed community, where many heterogeneous and autonomous cloud service providers cooperate and share resources to reach a common objective defined in the federation contract.
- The federation contract specifies the rules and policies that should be followed during the provision of services to consumers.
- It is an inter-cloud organization where multiple cloud service providers come under a single umbrella.

Cloud Federation (contd)



Objectives of Cloud Federation

- Dynamically expand resources to fulfill consumer demand.
- Resource lending.
- Integration of different types of cloud services in a single frame.
- Provision of reliable and required Quality of Service fulfilled services.
- Minimization of Service Level Agreement (SLA) violations.

Advantages of Cloud Federation

- Performance Guarantee – Resource lending helps to fulfill the demanded performance requirements of cloud service consumers.
- Service Availability Guarantee – Integration of different cloud service providers help to migrate services not only from one service provider, but also from disaster-prone area to a safe location.
- Convenience -- Cloud federation provides convenience and a unified view of services to the consumers.
- Dynamic Workload Distribution – Cloud federation distributes the service demand of the consumers into geographically distributed data centers of different cloud service providers.

Cloud Federation Models

- Semantics-based
- Market-oriented
- Reservoir
- Service-layers–oriented

Semantics-based Cloud Federation Model

- It is a theoretical federation model based on semantics and Infrastructure as a Service.
- It focuses on the interoperability of the components of different cloud service providers.
- Ontology is used to provide interoperability.

Market-oriented Cloud Federation Model

- This model focuses on the commercialization of infrastructure resources to fulfill the demand of the service market.
- Four components used in this model are as follows --
- Clouds – used for the provision of services.
- Application broker – a middleware interface for communication between consumers and cloud service providers.
- Cloud coordinator – a component located at each cloud in the federation and maintains the integrity of the federation.

Concentrator – it acts as the market of resources and services.

Reservoir

- It is a cloud federation project of IBM.
- It provides a cloud federation framework to provide Software as a Service to the cloud providers.
- The objective of this framework is to help the isolated cloud service providers in overcoming the difficulties faced during the provision of services to the consumers.
- The four functional aspects provided by Reservoir are –
  - Automatic and fast installation of different services and applications.
  - Elasticity.
  - Continuous optimization.
  - Independence of virtualization technologies.

Service-layers-oriented Cloud Federation Model

- This federation model focuses on the relationships of IaaS, Paas and SaaS of cloud.
- Services are isolated as layers.
- It integrates not only different heterogeneous resources, but also different services of clouds.
- It provides a framework for information flow and parameter translations among different cloud service layers.
- Additionally, provides a framework where brokers perform different service scheduling for providing services to the consumers.

Geneva Framework of Cloud Federation

- This framework is developed by Microsoft to focus the issues of cloud federation.
- It is a claim-based framework and provides a common framework for accessing applications and other systems seamlessly.
- Multiple cloud service providers use this framework to interact with each other.
- Developers use this model for developing different authentication protocols that will work on existing corporate identity systems such as Active Directory, LDAPv3-based directories etc.

XMPP and XCP for Cloud Federation

- Cloud federations are based on Internet Engineering Task Force (IETF) standard Extensible Messaging and Presence Protocol (XMPP) and interdomain federation using the Jabber Extensible Communications Platform (Jabber XCP).
- Jabber XCP is a highly scalable, extensible, available, and device-agnostic presence solution built on XMPP and provides a programmable platform for adding presence and messaging services to the existing applications and services for creating new presence-based solutions.
- Messages are exchanged within different XMPP servers.

Advantages of XMPP

- Decentralized – anybody can configure their own XMPP server.
- Uses open standard.
- Multiple implementation of clients and servers are present in XMPP.
- Secure -- Simple Authentication and Security Layer (SASL) and Transport Layer Security (TLS) are used to provide security.
- Flexible and extensible for incorporating different types of applications, services and systems.

Levels of Cloud Federations

- There are four levels of cloud federations based on the ability to exchange XML stanza and messages among XMPP servers.
- The four levels of cloud federations are ---
  - Permissive federation
  - Verified federation
  - Encrypted federation
  - Trusted federation

Permissive Cloud Federation

- In this federation, an XMPP server accepts connection request from other XMPP servers without verifying identity.
- DNS lookups or certificate checking is not present in this federation.
- Domain spoofing is the major problem in this federation.
- In domain spoofing an unauthorized third party can pretend to be another authorized domain.

Verified Cloud Federation

- In this federation, an XMPP server accepts connection request from other XMPP servers after verifying identity.
- It uses DNS lookups and domain specific key exchange to verify the identity of other XMPP servers.
- Domain spoofing is not present in this federation.
- The connection among different XMPP servers are verified but not encrypted.
- DNS poisoning attacks are the major problems in this federation.

Encrypted Cloud Federation

- In this federation, an XMPP server accepts connection request from other XMPP servers after verifying identity and connection requests are encrypted.
- Transport Layer Security (TLS) and digital certificates are used for providing security .
- XEP-0220 defined server Dialback protocol is used for identity verification.
- Server Dialback prevents the address spoofing present in the XMPP networking.

Trusted Cloud Federation

- In this federation, an XMPP server accepts the connection request from other trusted XMPP servers .
- Root certification authority (CA) makes trusted the XMPP server by providing a digital certificate.
- The authenticating server authenticates digital certificates before accepting the connection request.
- In this federation, trusted digital certificates provides strong security compared to other cloud federations.
- DNS poisoning attacks are avoided in trusted cloud federation.
- Trusted cloud federation is difficult to manage due to its complexity.

Future of Cloud Federation

- The objective of the cloud federation is to seamlessly interact between people, devices, information feeds, documents, application interfaces, and cloud service providers.
- It enables cloud service providers and software developers to integrate and deploy efficient, easily scalable, fault-tolerant and heterogeneous cloud services.
- The use of XMPP and XEP protocols help the individual cloud service providers to overcome the issues of scalability, failure of services and easy migration of services from one provider to another.

Privacy in Cloud

Difference Between Privacy and Security

- Privacy of the data ensures the appropriate use of personal data under different circumstances.
- Security is the set of practices used to ensure confidentiality, availability and integrity of data.
- Security techniques are used to ensure privacy of data.

Introduction

- Data privacy is the crucial aspect of any business organizations and individuals.
- Outsourcing of data in the cloud creates challenges in data privacy.
- Adoption of cloud services depends on the maintenance of data privacy in the cloud.
- Different business organizations refrain to adopt cloud services due to data privacy.
- Therefore, privacy in cloud is required to increase the chances of cloud service adoption.

Challenges of Cloud Privacy

- Consumers' perspectives---
  - Do not know the location of their data stored in the cloud.
  - Who can or can't access the data.
  - Who keeps data.
  - What is really happening when a request for data deletion sent to the provider.
  - Whether CSP sells their data or not.
- CSP Perspectives –
  - Due to support of multi-tenancy features, the cloud service provider has to ensure that data of one tenant is not accessed by other tenant.
  - . Public and hybrid cloud service providers are more prone to cloud privacy risk.

Cloud Privacy Laws and Legislations

- The different universally accepted laws and legislations published to manage the cloud privacy are ---
  - Fair Information Practices
  - European Directive 95/46/EC
  - USA Health Insurance Portability and Accountability Act (HIPAA)
  - USA Gramm–Leach–Bliley Act

Types of Private Data

- Personally identifiable information (PII) – used to identify an individual. It includes --
  - Key attributes -- name, phone number, social security or national identity number, email address and passwords.
  - Quasi-identifier – These attributes are used for linking anonymized dataset with other datasets and then identifying individuals such as ZIP code, date of birth, address etc.

- Sensitive Information --
  - Membership data – provides information about the membership into political, religious and other different communities.
  - Demographic characteristics -- nationality, gender, education level, job position, criminal record.
  - Finance data-- credit card number, account balance, financial transaction traces.
  - Health data -- medical record, diseases, diagnostics, medical images, prescriptions

Fair Information Practices (FIP)

- FIP is developed by USA to provide data protection and privacy.
- FIP defined data protection principles are ---
  - Data collection limitation
  - Purpose specification
  - Purpose use limitation
  - Individual participation
  - Visibility and transparency
  - Compliance of data
  - Accountability of data

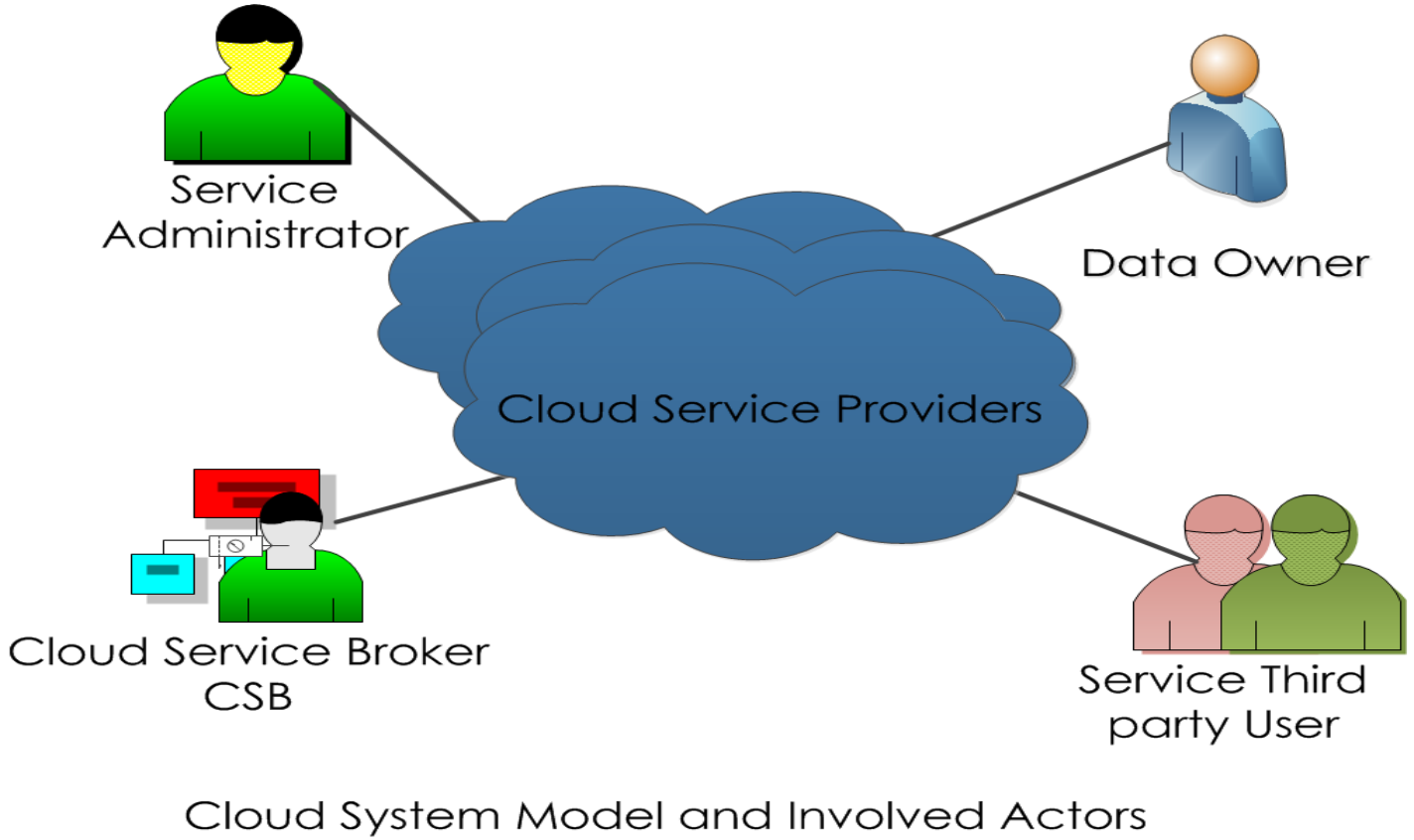
Privacy Issues in Cloud

- The privacy issues in cloud cover the following attributes –
  - Involved Actors
  - Lack of user control
  - Dynamic nature of the cloud environment
  - Compliance with laws and user's preferences
  - Accountability

Involved Actors

- The actors involved in handling data stored in cloud are ---
  - Data owner
  - Cloud service provider (CSP)
  - Cloud service broker (CSB) – it is an intermediate actor between CSP and data owner.
  - Cloud-based services – These includes application and programs deployed in cloud to provide different types of services such as customer relationship management, document management, cloud storage service etc.
  - Cloud-based service administrator – it is the owner of the service which can be CSP or other third party.
  - Cloud-based service third party

Involved Actors (contd)



Cloud System Model and Involved Actors

Source: shorturl.at/zKPQU

Lack of User Control

- In public cloud data is stored in the remote server of cloud service providers.
- Owner of data is unaware about data processing, data access, storage locations and any privacy violation occurring or not.
- Eg. Users of Dropbox and Mega do not know their actual data handling policy.
- Business data stored in cloud can be used for providing advertisement by CSP.
- The major concern of the data privacy is the lack of user control on the data stored in cloud.
- From provider's perspective revealing of data handling policy may creates threats of consumers data due to multi-tenancy aspect of cloud.

Lack of User Control (contd)

- The reported incidents of data privacy in cloud are ---
- In October 2007, a “Salesforce.com” employee became a victim of a phishing attack and leaked customer list.
- In March 2009, Google revealed documents of users to third parties who do not have the permission to explore the documents.
- In 2010, several Hotmail accounts were hacked.
- In 2011, Amazon customer services were unavailable for several days, and data were lost due to a logical flaw in the cloud storage design.

Dynamic Nature of Cloud Environment

- The dynamic nature of cloud creates problems in data privacy.
- Dynamic algorithms are used for storage and transferring of data in the cloud.
- Transborder cloud imposes several restrictions on data handling policy.
- In transborder cloud, data flows from one country to another.
- Different countries obeys different laws for data storage and processing which create issues in data privacy.
- Data replication causes data privacy issues in cloud.
- In data replication, several copies of the same data are made and stored into different servers which may or may not reside in the same country.

Compliance with laws and user's preferences

- Privacy Compliance (PC) is one of the major issue in data privacy in the cloud.
- Privacy Compliance (PC) depends on two factors –
  - Precise definition of the privacy policies
  - Capability of the used enforcement mechanisms
- The cloud service consumer is unaware of the data handling policy used by CSP.
- Cloud service consumers do not aware whether the data operations performed on their data are compliant with the privacy law enforced by the country where the data is actually stored.

Accountability

- According to the definition given by Galway project in the context of business data ---

  "Accountability is the obligation to act as a responsible steward of the personal information of others, to take responsibility for the protection and appropriate use of that information beyond mere legal requirements, and to be accountable for any misuse of that information."

- Auditing mechanism is used for accountability of data.

- In cloud, accountability of data is required due to dynamic storage, duplication and transfer of data from one server to another.

- Accountability helps the cloud consumer to know where data are actually located, who is the processor of data, who are accessing their data etc.

Techniques of Data Privacy Protection

- The techniques used for data privacy protection are –
  - Encryption
  - Processing encrypted data
  - Obfuscation
  - Sticky policy
  - Trusted platform module
  - Data segmentation
  - Trusted third party mediator (TTPM)

Encryption

- In encryption, cryptographic solutions are used to encrypt data in the cloud.
- Encryption can be done by either CSP or consumer or both.
- Encryption-decryption key management is one of the issues in encrypting data in cloud.
  - One approach is to keep encryption-decryption kept on the consumer side.
  - Key management is handled by a trusted third party.

Processing Encrypted Data

- Processing of encrypted data in cloud compromises the privacy of data, as data is decrypted before processing.
- To overcome data privacy compromise, processing of encrypted data is proposed.
- In Homomorphic encryption, data processing and query execution is done on the encrypted data.
- The result of Homomorphic encryption is also encrypted.
- The decrypted result gives the same output as if the processing is done on the plain text.
- IBM FHE (Fully Homomorphic Encryption) is one of the Homomorphic technique used in the cloud.

Obfuscation

- Data obfuscation is a privacy preserving technique.
- It is also known as data masking, where masking rules are used for maintaining the data privacy by replacing the actual data with new data. The new data looks like actual data but are unrelated.
- Data obfuscation techniques are ---
  - Data randomization
  - Data Swapping
  - Anonymization

Obfuscation (contd)

- Data randomization – Data is made fuzzy by adding random variables with data. Eg. multiplying a column of data with a secret factor, replacing person identity with pseudo-identity.
- Data swapping – In data swapping, data values are swapped by obeying a predefined rule such that original data can be recovered from the swapped data.
- Data anonymization–
  - Data owner identity is removed from the data and then cloud actors use the data.
  - Data anonymization is subject to linking attack where owner identity is linked with the removed data by using other known databases.

Obfuscation (contd)

- Data anonymization–
  - To avoid the linking attack, K-anonymization technique is proposed.
  - In K-anonymization technique, after removing person identity remaining data is classified into quasi identifier and sensitive attributes.
  - quasi identifier data is replaced with less specific data.
  - It is called k-anonymous as in this technique each record can not be differentiated with k-1 record in the same database.
- Data obfuscation is less secure than encryption but computationally efficient.

Sticky Policy

- Data is attached to a sticky policy.
- The sticky policy determines the authenticated processing request span and authorization.
- Policy enforcement is ensured through policy management components called Policy Decision Point (PDP).
- PDP evaluates the data processing request against predefined sticky policy and decides whether to grant the processing request or not.
- Stick policy is computationally inefficient as data processing request has to be evaluated for each request.

Trusted Platform Module

- Trusted Platform Module (TPM) is a tamper-resistant hardware component developed by the Trusted Computing Group (TCG).
- TPM is responsible to implement all data privacy preserving techniques.
- TPM provides a black box to store data securely.
- However, TPM does not provide secure data processing.
- The disadvantage of TPM based solution is that it is hardware based.
- CSP has to setup TPM hardware at its each data center.

Data Segmentation

- Data segmentation is used to maintain data confidentiality.
- It not only ensure privacy of data but also data associations.
- Data is divided into different sets of blocks and stored into different non-linkable storage .
- In cloud, data segmentation is used to ensure data privacy.
- Data segmentation is performed according to sensitive data associations.
- Data reassembly is also planned along with data segmentation in cloud.

Trusted Third Party Mediator (TTPM)

- TTPM acts as middleman between service consumer and the other cloud actors.
- It is responsible for policy enforcement and data auditing.
- Eg. In E-commerce applications, TTPM is used to build customer trust.
- Cloud TTPM hides the personal identity and sensitive data from other cloud actors.
- TTPM is a trusted third party which acts as trusted root authority in cloud.
- It also helps to manage encryption-decryption key in cloud.

References

- Cloud Computing, Authors: John W. Rittinghouse, James F. Ransome, CRC press, 2017.
- M. R. M. Assis, L. F. Bittencourt, and R. Tolosana-Calasanz, “Cloud Federation: Characterisation and Conceptual Model,” in Proc. of the IEEE/ACM 7th International Conference on Utility and Cloud Computing, December 2014, pp. 585–590.
- A. Ghorbel, M. Ghorbel, and M. Jmaiel, “Privacy in cloud computing environments: a survey and research challenges”, Journal of Supercomputing, June 2017, vol: 73, no: 6, pp. 2763 – 2800.

Thank You!!