DATA SECURITY AND PRIVACY

Cloud computing has become an emerging technique due to its on demand service and scalability features. Most usage of cloud today is in data storage and big data or computation intensive applications. Thus data security and privacy has become the chief concern, especially for business level data. Data security mainly includes data confidentiality, availability and integrity.

Data privacy is to prevent identification of data stored in cloud. According to data security and privacy, issues in cloud exist during the data life cycle from generation, transfer, use, share, storage, archival until destruction. Traditional methods for data security usually rely on data encryption and access control.

Data encryption with AES or other encryption methods would prevent valuable information leakage although the adversary gets hold of the data. However, it has efficiency issue when dealing with oceans of data in cloud environment due to large encryption and decryption overhead in storage and computation.

Access control is to prevent unauthorized users to access data. However, in cloud computing, users do not have physical control over the machines they store data on, and also the same physical machine could be shared by multiple tenants with virtualization, adversary would be able to monitor the physical machine behavior to obtain valuable data from other tenants and also the cloud providers themselves are unreliable, they might accidentally or intentionally modify or leak the data stored to adversaries.

In public cloud environment, threats come from both the outsider and insider attack. The outsider attacks by malicious codes, DDoS attack, network eavesdropping etc.

There are three layers in cloud computing platform.

- In the infrastructure layer, each physical machine has multiple virtual machines (VMs) installed.
- The platform layer provides the platform for customers. Customers could have their own applications or software's and configurations installed.
- The software layer provides the software stacks by the cloud providers.

For the client side, a customer could either be a legal user or an attacker pretending as legal users. Network eavesdroppers could also sit in between to perform man in the middle attacks. Firewalls or Intrusion Detection Systems (IDS) could be installed to protect the entire cloud environment.

CLOUD DATA SECURITY CHALLENGES

As mentioned above, there would be more concerns on data security in cloud environment than in traditional single ma-chine which are in hold of users themselves. In this section, different concerns and possible attacks during cloud data usage and storage phases are summarized.

A. Data Use: -

Once data is migrated to cloud, cloud providers will clearly get hold of everything users transferred to cloud machine in-stances. Both adversaries and cloud providers themselves might misuse the data stored in cloud.

Thus some kinds of data transformation might be needed to prevent valuable information leakage. If cloud is only used for data storage and no further operations are needed, simple encryption is feasible. However, in most cases, further processing might be needed.

Users might need to apply certain processing over the data stored. For example, computation might be needed like matrix multiplication. Also data analytics methods such as ma-chine learning algorithms need to be applied for data classification.

B. Data Storage:-

In cloud environment, users' data are stored in remote virtual machine instances in possession of cloud providers. According to, there could be various outside attacks over virtual machines including malicious codes attack, compromising the corresponding Virtual Machine Monitor etc.

Besides the outside attack, users lack of physical control of their data. Insiders of cloud providers could clearly see what is stored in their virtual machine instances. It would be a catastrophe if the insiders of cloud providers collude with adversaries to intentionally modify or leak customers' data.

Data storage security includes confidentiality, integrity and availability. For data confidentiality, how to prevent information leakage and efficiently check data integrity over large amount of data stored in cloud remains a question. The goal here is to minimize the probability to recover the original data obtained from the compromised cloud storage system.

DATA CONFIDENTIALITY AND AVAILABILITY PROTECTION METHODS

A. **File distribution in multiple storages:**

Multiple storages are applied to minimize the information leakage when a single storage is compromised. With this method, encryption is not needed.

B. **Processing over encrypted data:**

In the above section, multiple storages are applied to secure data confidentiality by minimizing information contained in each storage node. However, in the case of data computation, the above method is not applicable. Since for data processing in cloud, each piece of data needs to contain computable information. A better solution might be computing while keeping data encrypted.

DATA PRIVACY PROTECTION METHODS

In data privacy against data mining is kept by distributing data to different cloud providers. Thus, data analytics based on each part in one cloud might be misleading. For example, prediction made on the overall data file could be different from that made on each part.

However, this kind of approach would not protect each individual's sensitive data. For example, a database contains columns of username and the corresponding income. If the file is simply divided by rows as in, each individual's income information is still leaked.