A Literature Survey on Data Privacy/Protection Issues and Challenges in Cloud Computing

Abdul Wahid Khan, Siffat Ullah Khan, Muhammad Ilyas and Muhammad Ilyas Azeem

Abstract: Cloud computing is a technique to deliver software, storage and processing. It increases system’s capability without changing the existing infrastructure, educating new people or taking license for the softwares. It improves the existing software capabilities and extends the Information Technology resources. In recent years, cloud computing has grown up rapidly and boosted the business concept in IT industry. Despite of all the achievements in cloud computing, security is still a critical challenge in cloud computing paradigm. These challenges include user’s secret data (like health and financial data) loss, leakage and disclosing of privacy. We have studied literature and discussed various model in cloud computing, it shows that privacy/protection in cloud is still immature.

Keywords: Cloud computing, Data privacy, Data protection, Security issues.

1. Introduction:

Now minute and average business organization are realizing that simply exchange to the cloud can get access to excellent business claims and increase up their infrastructure assets in a very low-cost, Internet on an as-needed basis [1]. This new and exciting paradigm has generated significant interest in the marketplace and the academic world [2], resulting in a number of notable commercial and individual cloud computing services, e.g., form Amazon, Google, Microsoft, Yahoo, and Sales force [3]. Also, top database vendors, like Oracle, are adding cloud support to their databases.

The providers are enjoying the superficial opportunity in marketplace but they should ensure that they possess the right security features. The cloud provide facilities like fast development, lower cost on pay-for-use, quick provisioning, quick flexibility, everywhere network contact, hypervisor defense against network vulnerability, economical failure recovery and data storage solution, on-request security checks, synchronized detection of system altering and rapid re-construction of services. The cloud provides this compensation, until some of the risks are better understood.

The basic concept of the cloud, based on the services they offer, from application service provisioning, grid and service computing, to Software as a Service [1, 4, 5]. Despite of the specific architecture, the dominant concept of this computing model is that customers’ data, which can be of individuals, organizations or enterprises, is processed remotely in unknown machines about which the user not aware. The ease and efficiency of this approach, however, comes with privacy and security risks [3, 6, 7]. Confidentiality of data is the main hurdle in implementation of cloud services.

A huge data centers are established in cloud computing, but the deployment of data and services are not trustworthy. These create various new security challenges. These challenges are vulnerabilities in accessibility, virtualization and web?? such as SQL injection, cross site scripting, physical access issues, privacy and control issues happening from third parties having physical control of data, issues related to identity and credential, issues related to data confirmation, changing and privacy, data loss and theft, issues related to integrity and IP spoofing.

1.1. Characteristics of cloud computing

Cloud computing exhibit five essential characteristics as defined by NIST (National Institute of Standards and Technology)[8].

1. On-demand self-service. A consumer can unilaterally provide computing capabilities.
2. Broad network access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms.
3. Resource pooling. The provider’s computing resources are pooled to serve multiple consumers, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
4. **Rapid elasticity.** Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in.

5. **Measured service.** Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service.

### Service Models

NIST has identified three “service models” through which cloud computing is offered. They are:

1. **SaaS** The concept in “Software as a Service” is the simple use of the cloud provider’s applications running on a cloud infrastructure. The user does not manage or control the underlying cloud infrastructure such as the network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

2. **PaaS** The next layer of complexity in cloud computing, “Platform as a Service”, as far as the user is concerned, is to deploy onto the cloud infrastructure consumer-created or acquired applications using programming languages and tools supported by the provider. As in the case of a SaaS model, the user does not manage or control the underlying network, servers, operating systems, or storage, but in the case of a PaaS model, the user does have the ability to control the deployed applications and potentially application hosting environment configurations.

3. **IaaS** The most comprehensive model of cloud computing is known as “Infrastructure as a Service”. In this model, the provider supplies the required processing, storage, networks, and other fundamental computing resources and the user is able to deploy and run any software that it may require, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems; storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

### Deployment Models

Finally, NIST identifies four different types of deployment models for the foregoing service models. These deployment models are:

1. **Private cloud.** The cloud infrastructure is operated solely for one organization. It may be managed by the organization or a third party and may exist on premise or off premise. Arguably this may be the most secure type of infrastructure, depending on the nature of the controls deployed and the diligence of the operator.

2. **Community cloud.** In this model, the cloud infrastructure could be shared by several organizations and supports a specific community or interest group that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).

   It may be managed by the organizations or a third party and may exist on premise or off premise.

3. **Public cloud.** The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

4. **Hybrid cloud.** The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

This paper explores the various issues, especially data security issues of cloud computing and its solution. This paper is structured as follows: Section 2 describes frequent security issues of cloud model. “Software as Service” (SaaS), “Platform as Service” (PaaS) and “Infrastructure as Service” (PaaS). Section 3 describes security threats and its solutions. Section 4 describes policy-driven framework, in which we first briefly analyze existing techniques to be adopted by each component of the framework and then introduce data protection models to be deployed in the cloud. Finally, Section 5 describes the conclusion and research directions.

## II. Security Issues in Service Model:

Cloud computing having three delivery models through which services are delivered to end users. These models are SaaS, IaaS and PaaS which provide software, Infrastructure and platform assets to the users. They have different level of security requirements.

2.1 **Security issues in SaaS:**

   Software as service is a model, where the software applications are hosted slightly by the service provider and available to users on request, over the internet. In SaaS, client data is available on the internet and...
may be visible to other users, it is the responsibility of provider to set proper security checks for data protection. This is the major security risk, which create a problem in secure data migration and storage.

The following security measures should be counted in SaaS application improvement process such that Data Security, Data locality, Data integrity, Data separation, Data access, Data confidentiality, Data breaches, Network Security, Authentication and authorization, Web application security, Identity management process.

The following are the basics issues through which malicious user get access and violate the data security, store at the SaaS dealer such that Cross-site scripting, SQL Injection flaw, Cross-site request forgery, Insecure storage, Insecure configuration.

2.2 Security issues in PaaS
PaaS is the layer above the IaaS. It deals with operating system, middleware, etc. It provides set of service through which a developer can complete a development process from testing to maintenance. It is complete platform where user can complete development task without any hesitation.

In PaaS, the service provider give some command to customer over some application on platform. But still there can be the problem of security like intrusion etc, which must be assured that data may not be accessible between applications.

2.3 Security issues in IaaS
IaaS introduce the traditional concept of development, spending a huge amount on data centers or managing hosting forum and hiring a staff for operation. Now the IaaS give an idea to use the infrastructure of any one provider, get services and pay only for resources they use. IaaS and other related services have enable set up and focus on business improvement without worrying about the organization infrastructure.

The IaaS provides basic security firewall, load balancing, etc. In IaaS there is better control over the security, and there is no security gap in virtualization manager. The main security problem in IaaS is the trustworthiness of data that is stored within the provider’s hardware.

III. Cloud Computing Security Threats and solution

3.1. Top Seven Security Threats
Top seven security threats to cloud computing discovered by “Cloud Security Alliance” (CSA) are [9]:

i. Abuse and Nefarious Use of Cloud Computing.
Abuse and nefarious use of cloud computing is the top threat identified by the CSA. A simple example of this is the use of botnets to spread spam and malware. Attackers can infiltrate a public cloud, for example, and find a way to upload malware to thousands of computers and use the power of the cloud infrastructure to attack other machines.
Suggested remedies by the CSA to lessen this threat:
• Stricter initial registration and validation processes.
• Enhanced credit card fraud monitoring and coordination.
• Comprehensive introspection of customer network traffic.
• Monitoring public blacklists for one’s own network blocks.

ii. Insecure Application Programming Interfaces.
As software interfaces or APIs are what customers use to interact with cloud services, those must have extremely secure authentication, access control, encryption and activity monitoring mechanisms - especially when third parties start to build on them.
Suggested remedies by CSA to lessen this threat:
• Analyze the security model of cloud provider interfaces.
• Ensure strong authentication and access controls are implemented in concert with encrypted transmission.
• Understand the dependency chain associated with the API.

iii. Malicious Insiders.
The malicious insider threat is one that gains in importance as many providers still don't reveal how they hire people, how they grant them access to assets or how they monitor them. Transparency is, in this case, vital to a secure cloud offering, along with compliance reporting and breach notification.
Suggested remedies by CSA to lessen this threat:
• Enforce strict supply chain management and conduct a comprehensive supplier assessment.
• Specify human resource requirements as part of legal contracts.
• Require transparency into overall information security and management practices, as well as compliance reporting.
• Determine security breach notification processes.
iv. **Shared Technology Vulnerabilities.**

Sharing infrastructure is a way of life for IaaS providers. Unfortunately, the components on which this infrastructure is based were not designed for that. To ensure that customers don't thread on each other's "territory", monitoring and strong compartmentalization is required.

Suggested remedies by CSA to lessen this threat:
- Implement security best practices for installation/configuration.
- Monitor environment for unauthorized changes/activity.
- Promote strong authentication and access control for administrative access and operations.
- Enforce service level agreements for patching and vulnerability remediation.
- Conduct vulnerability scanning and configuration audits.

v. **Data Loss/Leakage.**

Be it by deletion without a backup, by loss of the encoding key or by unauthorized access, data is always in danger of being lost or stolen. This is one of the top concerns for businesses, because they not only stand to lose their reputation, but are also obligated by law to keep it safe.

Suggested remedies by CSA to lessen this threat:
- Implement strong API access control.
- Encrypt and protect integrity of data in transit.
- Analyze data protection at both design and run time.
- Implement strong key generation, storage, and management, and destruction practices.
- Contractually demand providers to wipe persistent media before it is released into the pool.
- Contractually specify provider backup and retention strategies.

vi. **Account, Service & Traffic Hijacking.**

Account service and traffic hijacking is another issue that cloud users need to be aware of. These threats range from man-in-the-middle attacks, to phishing and spam campaigns, to denial-of-service attacks.

Suggested remedies by CSA to lessen this threat:
- Prohibit the sharing of account credentials between users and services.
- Leverage strong two-factor authentication techniques where possible.
- Employ proactive monitoring to detect unauthorized activity.
- Understand cloud provider security policies and SLAs.

vii. **Unknown Risk Profile.**

Security should be always in the upper portion of the priority list. Code updates, security practices, vulnerability profiles, intrusion attempts – all things that should always be kept in mind.

Suggested remedies by CSA to lessen this threat:
- Disclosure of applicable logs and data.
- Partial/full disclosure of infrastructure details (e.g., patch levels, firewalls, etc.).
- Monitoring and alerting on necessary information.

3.2. **Other Security Threats [10, 11]**

a. **Failures in Providers Security.** Cloud providers control the hardware and the hypervisors on which data is stored and applications are run and hence their security is very important while designing cloud.

b. **Attacks by other customer.** If the barriers between customers break down, one customer can access another customer's data or interfere with their applications.

c. **Availability and reliability issues.** The cloud is only usable through the Internet so Internet reliability and availability is essential.

d. **Legal and Regulatory issues.** The virtual, international nature of cloud computing raises many legal and regulatory issues regarding the data exported outside the jurisdiction.

e. **Perimeter security model broken.** Many organizations use a perimeter security model with strong security at the perimeter of the enterprise network. The cloud is certainly outside the perimeter of enterprise control but it will now store critical data and applications.
Integrating Provider and Customer Security Systems. Cloud providers must integrate with existing systems otherwise the bad old days of manual provisioning and uncoordinated response will return.

3.3. Existing Solutions for Security Threats

3.3.1. Mirage Image Management System [12]

The security and integrity of VM images are the foundation for the overall security of the cloud since many of them are designed to be shared by different and often unrelated users. This system addresses the issues related to secure management of the virtual-machine images that encapsulate each application of the cloud.

The overall architecture of Mirage Image Management System. Mirage Image Management System consists of 4 major components:

1. **Access Control.** This framework regulates the sharing of VM images. Each image in the repository has a unique owner, who can share images with trusted parties by granting access permissions.

2. **Image Transformation by Running Filters.** Filters remove unwanted information from images at publishing and retrieval time. Filters at publish time can remove or hide sensitive information from the publisher’s original image. Filters at retrieval time may be specified by the publisher or the retriever.

3. **Provenance Tracking.** This mechanism tracks the derivation history of an image.

4. **Image maintenance.** Repository maintenance services, such as periodic virus scanning, that detect and fix vulnerabilities discovered after images are published.

**Advantages.** Filters mitigate the risk in a systematic and efficient way. The system stores all the revisions which allows the user to go back to the previous version if the current version not fulfill the requirements. The default access permission for an image is private so that only owner and system administrator can access the image and hence distrusted parties cannot access the image.

**Limitations.** Huge performance overheads, both in space and time. Filters cannot be 100% accurate and hence the system does not eliminate risk entirely. Virus scanning does not guarantee to find all malware in an image. "The ability to monitor or control customer content" might increase the liability of the repository provider (For detailed explanation about Mirage Image Management System please refer to [13]).

3.3.2. Client Based Privacy Manager [14]

Client based privacy manager helps to reduce the risk of data leakage and loss of privacy of the sensitive data processed in the cloud, and provides additional privacy related benefits.

The overall architecture of the privacy manager have main features of the privacy manager are:

- **Obfuscation.** This feature can automatically obfuscate some or all of the fields in a data structure before it is sent off to the cloud for processing, and translate the output from the cloud back into de-obfuscated form. The obfuscation and de-obfuscation is done using a key which is chosen by the user and not revealed to cloud service providers.

- **Preference Setting.** This is a method for allowing users to set their preferences about the handling of personal data that is stored in an un-obfuscated form within the cloud. This feature allows the user greater control over the usage of his data.

- **Data Access.** The Privacy Manager contains a module that allows users to access personal information in the cloud, in order to see what is being held about them, and to check its accuracy. This is an auditing mechanism which will detect privacy violations once they have happened.

- **Feedback.** The Feedback module manages and displays feedback to the user regarding usage of his personal information, including notification of data usage in the cloud. This module could monitor personal data that is transferred from the platform.

- **Personae.** This feature allows the user to choose between multiple personae when interacting with cloud services.

**Advantages.** This solution solves many practical problems such as Sales Force Automation Problem, Customized End-User Services Problem and Shared Portfolio Calculation problem.
Disadvantages. If the service provider does not provide full cooperation, the features of the Privacy Manager other than obfuscation will not be effective, since they require the honest cooperation of the service provider. The ability to use obfuscation without any cooperation from the service provider depends not only on the user having sufficient computing resources to carry out the obfuscation and de-obfuscation, but also on the application having been implemented in such a way that it will work with obfuscation.

3.3.3. Transparent Cloud Protection System (TCPS) [15]
TCPS is a protection system for clouds aimed at transparently monitoring the integrity of cloud components. TCPS is intended to protect the integrity of guest Virtual Machines (VM) and of the distributed computing middleware by allowing the host to monitor guest VMs and infrastructure components. Figure 1 shows the architecture of TCPS. TCPS is a middleware whose core is located between the Kernel and the virtualization layer. By either actively or passively monitoring key kernel or cloud components TCPS can detect any possible modification to kernel data and code, thus guaranteeing that kernel and cloud middleware integrity has not been compromised and consequently no attacker has made its way into the system.

![Figure 1. TCPS Architecture.](image)

All TCPS modules reside on the Host and Qemu, are leveraged to access the guest. Suspicious guest activity can be noticed by the Interceptor and they are recorded by the Warning Recorder into the Warning Queue where the potential alteration will be evaluated by the Detector component. TCPS can locally react to security breaches or notify the distributed computing security components of such an occurrence. In order to avoid false positives as much as possible, an administrator can notify TCPS of the new components' checksum.

Advantages. This system is effective in detecting most kind of attacks. This system is able to avoid false-positives (Guest maintenance tolerance). The system minimizes the visibility from the VMs (Transparency). The system and the sibling guests are protected from attacks of a compromised guest. The system can be deployed on majority of the available middleware. The system can detect an intrusion attempt over a guest and, if required by the security policy, takes appropriate actions against the attempt or against the compromised guest and/or notify remote middleware security-management components.

3.3.4. Secure and Efficient Access to Outsourced Data [16]
Providing secure and efficient access to outsourced data is an important component of cloud computing and forms the foundation for information management and other operations.

Problem. Figure 2 shows the typical owner-write-user read scenario. Only the owner can make updates to the outsourced data, while the users can read the information according to access rights. Since the data owner stores a large amount of information on the untrusted service provider, the owner has to encrypt the outsourced data before putting on the server. The outsourced data will be accessed by different end users from all over the network and hence computationally expensive operations on the data blocks (smallest unit of data) should be avoided and the amount of data stored in the end users must be reduced. Right keys should be provided to the end users to control their access.
Solution. Fine-grained access control should be provided for the outsourced data by encrypting every data block with a different symmetric key.

IV. Data Protection Model

As the awareness of the cloud privacy issues is going to increase, but still small work has been done in this area. Recently, Pearson et al. has proposed accountability mechanisms to address privacy concerns of end users [17] and then develop a simple solution, a privacy manager, relying on obfuscation techniques [18]. Their basic idea is that the personal data of the user’s is in an encrypted form on the cloud, and only the encrypted data is processed there.

In this section, the characteristics of cloud services, and policy-driven framework for protecting data privacy is present.

The data protection framework consists of three major components: policy ranking, policy integration, and policy implementation.

4.1 Policy Ranking Models

Policy ranking use to rank the service provider with the most frequent privacy policies contrasted to the users’ privacy checks (or policies). Privacy and efficiency requirements are two important factors to be considered. Based on different importance on the two factors, we proposed three policy ranking models: (i) User-oriented ranking model; (ii) Service-provider-oriented ranking model; and (iii) agent based ranking model.

4.1.2 User-oriented Ranking Model

In this model, users have processing and storage capacity and responsible for policy comparison. First, users need to collect privacy policies from service providers who provide the required services. The second option is that users show their service needs in the cloud and the corresponding service providers will then send their privacy policies to the users. After the users getting policies from providers, the users can start ranking and select the most suitable one.

Advantages:
Privacy preservation is high at user end.

Disadvantages:

i. Processing cost (end-user and overall) is high.
ii. Communication cost (end-user and overall) is high.
iii. Privacy preservation at service provider side is low.
iv. Additional requirements required at user end and having high processing capability.

4.1.3 Service-provider-oriented Ranking Model

This model deploys the policy comparison task. Users need to show their service needs as well as, privacy requirements. Service providers who are interested in attracting more users will run the same ranking program to compare the privacy policies and return the similarity scores back to the user. Then users can select their preferred service providers. Since policy comparison is executed at the service provider side, so the service provider will be responsible for privacy of policies. The more similar the policies are there, fewer efforts are estimated to integrate policies.

Advantages:

i. Processing cost at user end is medium.
ii. Communication cost (user end and overall) is medium.
iii. Low processing capability at user end.
iv. Privacy preservation at service provider side is high.
Disadvantages:
Overall processing cost is high.

4.1.4 Agent-based Ranking Model
The last two models show that, either the users need to have some additional ability to avoid releasing their privacy policies to all service providers, or the service providers may need to carry out this problem and resolve by using certified third parties, which can provide brokers in-between users and service providers. The agent assembles the policies of service providers and gives them to users according to their requirements along with the ranking list.

Advantages:
i. Processing and communication cost at user end and overall cost is low.
ii. Privacy preservation at user end and service provider side is medium.

Disadvantages:
As a additionally a certified broker or third-party is required.

4.2 Policy Integration Models
Successful selection of service provider by the customers, the next step is to reach agreement on both parties’ data privacy concerns. Policy integration model gets all privacy requirements as input and helps to generate policies, which should be implemented by expecting celebrities.

4.2.1 A Policy Integration Approach
The following are some properties for designing good policy integration

P1: In the cloud each service provider and sub contractor having its own set of privacy policies. The policy integration approach is used to solve the variance and get agreement of all requirements.
P2: The policy integration second approach is to create or generate automatically actual policies as output. This property of the policy integration approach is very important in real-life situations.
P3: The policy integration approach should be elastic to make possible policy updates and minimize maintenance cost and it should not require re-executing the policy change every time.

4.2.2 Common point approach
In cloud computing different parties involved at different level at a single service. Here the problem is that which party policies to be integrated, such that the privacy requirements are fulfilled and overall performance is not disturbed. For that purpose the focus is on the cost and the policies without loss of its simplification. These problems can be handling through binary expression, where the leaf nodes present the policies to be integrated and the internal node present the integration operations.

4.2.3 Shared approach
In this approach the adjacent parties, who have a direct contact, can integrate polices. These integration cost be equally shared among the contributing parties. It is also difficult to achieve the common optimal solution compared to the common point approach.

4.3 Policy implementation Model
Once polices are created and it is properly integrated then the next step is to start these policies. But before carrying out polices it must satisfy the conditions and avoid the issues raised during its implementation. It must satisfy:

P1: Polices implementation must satisfy the integrity, availability and confidentiality of data and policies. The integrity data should be available to authorized service provider on need basis. Confidential data and policies can only be accessible to authorized data.
P2: The policies can be customizing for the party owning the data but it should be uniquely identified.
P3: The authorize parties should be modified polices, who have the changing rule of service providers according to the user requirements.
4.3.1 **Tight coupling**

This technique can be used to enforce or intact the policy in their original format. Confidentiality of data and policies will always be assured using tight coupling technique. Some of the approaches [19] to apply sticky policies to data have been assumed, no satisfactory explanation has been proposed.

4.3.2 **Loose coupling**

This technique is used to update the policies dynamically according to data access. Due to this technique the policies are stored at the remote trusted location. These policies can only be modified by the authorized service provider(s).

V. **Conclusion**

This paper, explored the importance of the cloud computing; but still there are a number of risks associated with the cloud computing procedure and process. This paper also illustrated the data privacy problem in cloud computing environment. Different data protection models and techniques have been defined that show their contribution in cloud computing. This paper will provide a base for future research work in the field of data security of cloud computing system. The defined models in the paper have a lot of challenges and issues, which open a new way for more research in this area.

**References:**


