Framework for Solving Data Loss/Leakage Security Threat in Cloud Environment

Shehu M. Studu, Anas Tukur B, Khalifah, Zalihat
Department of computer science
Sokoto State University, Sokoto

Abstract- Cloud computing is a flexible, cost-effective, and proven delivery platform for providing business or consumer IT services over the Internet. However, Cloud computing has the potential to change how organizations manage information technology and transform the economics of hardware and software at the same time [9]. Cloud computing have capability of motivating small and modern entrepreneurs to initialize their venture with little cost on IT infrastructure. Despite the benefits above this emerging Technology has a data loss or leakage issues which is a very serious. Therefore, it became necessary for an individual or organization to approach it carefully. In this paper, we discuss the causes of data loss or leakage issues of cloud computing and possible secure framework that will prevent data loss or leakage of individual or organizational in cloud environment.

Keywords: cloud computing, data loss/Leakage, Security, confidentiality, authentication, integrity

1. INTRODUCTION

Cloud computing is a flexible, cost-effective, and proven delivery platform for providing business or consumer IT services over the Internet. Cloud resources can be rapidly deployed and easily scaled, with all processes, applications, and services provisioned *on demand*, regardless of the user location or device. In effect, cloud computing shifts much of the control over data and operations from the client organization to their cloud providers, much in the same way organizations entrust part of their IT operations to outsourcing companies. Even basic tasks, such as applying patches and configuring firewalls, can become the responsibility of the cloud service provider, not the user Unfortunately like any other IT system, the cloud faces higher challenges than those involved in developing of secure IT systems, cloud computing presents an added level of risk because essential services are often outsourced to a third party. The *externalized* aspect of outsourcing makes it harder to maintain data integrity and privacy, support data and service availability, and demonstrate compliance.

In [1] Security is always afforded to an automated information system in order to attain the applicable objectives of preserving the following features:

- **Strong Authentication**: The control of authenticity, the process of identification of parts involved in electronic transactions or exchange of information with electronic means.
- **Authorization**: The authenticated access to resources, database and informative systems, according to the user's permission rights and the roles.
- **Data Confidentiality**: The protection of information either locally stored or in transmission from unauthorized access.
- **Data Integrity**: The protection of information either locally stored or in transmission from unauthorized modification.
- **Non-Repudiation**: Ensuring that no part of an electronic transaction can deny its attendance in it.

These features embody the fundamental security objectives of both data and for Information and computing services. In cloud computing concepts a customer usually relies on the provider to ensure that all the features are implemented to protect and maintain his data.

Dealing with security issue in cloud have never been an easy task for the provider, to aid cloud providers, CSA developed "Security Guidance for Critical Areas in Cloud Computing", initially released in April 2009, and revised in December 2009 [2]. This guidance has quickly become the industry standard catalogue of best practices to secure Cloud Computing, consistently lauded for its comprehensive approach to the problems faced by the cloud providers and incorporating informational guide on how to manage their cloud when faced with such problems. Problems or threats listed in the 2010 document of the CSA include:

- Abuse and Nefarious Use of Cloud Computing
- Insecure Application Programming Interfaces
- Malicious Insiders
- Shared Technology Vulnerabilities
- Data Loss/Leakage
- Account, Service & Traffic Hijacking
- Unknown Risk Profile.

In most cases, these problems seem to be intertwined and one cannot just treat a single problem without bothering to carter of the others. (ie pay much regard to one while ignoring the other).

This paper focuses on "Data Loss /Leakage" in public cloud computing. It was selected as a key threat here because most organizations cite data protection as their most important security issue. Typical concerns include the way in which data is stored and accessed, compliance and audit requirements, and business issues involving the cost of data breaches, notification requirements, and damage to brand value.

In [2] Data Storage/ leakage are described as the various ways to compromise data. Either by Deletion or alteration of records without a backup of the original content is an obvious example includes the unlinking a record from a larger context may render it unrecoverable, as can storage on unreliable media. Loss of an encoding key may result in effective destruction. Finally, unauthorized parties must be prevented from gaining access to sensitive data. The threat of data compromise increases in the cloud, due to the number of and interactions between risks and challenges which are either unique to cloud, or more dangerous because of the architectural or operational characteristics of the cloud environment.

The Causes of this vulnerability leading to Data loss/leakage in cloud includes;

- Insufficient authentication, authorization, and audit (aaa) controls,
- Inconsistent use of encryption and software keys
- Operational failures
- Persistence and remanence challenges: which is associated to disposal challenges
- Risk of association (Multitenancy)
- Data center reliability and disaster recovery

Data loss or leakage can have a devastating impact on a business. Beyond the damage to provider's brand and reputation, a loss could significantly impact employee, partner (associate cloud provider), and customer morale and trust. Loss of core intellectual

property could have competitive and financial implications. Worse still, depending upon the data that is lost or leaked, there might be compliance violations and legal ramifications.
Thus data loss or leakage is a very sensitive issue which must be well defined, categorized (depending on the deployment model) and managed. Data protection must be an element of the secure architecture design which will produce zero degree of data protection vulnerability to the architecture.

In this paper, we have proposed a secure framework for solution to data loss/leakage issues in cloud computing. The paper is divided in to 5 sections. Section 2 describes the Factors making up data loss/leakage using a separate module. In section3, our proposed framework is discussed while section 4 describe. Section 5 concludes the paper and gives an overview of our future work.

2.0 PROBLEMS STATEMENT

In [3] the rapid transition towards the clouds, has fuelled concerns on a critical issue for the success of information systems, communication and information security. From a security perspective, a number of unchartered risks and challenges have been introduced from this relocation to the clouds, deteriorating much of the effectiveness of traditional data protection mechanisms, thereby introducing vulnerabilities which pose serious threat to both the customers and cloud providers. Data loss and leakage are very sensitive issues that transgress all features of security (confidentiality, integrity and Availability). Factors making up data loss/leakage are addressed using a separate module. The modules contribute a substantial part of the security measure that was taken and used in designing the framework. The proposed modules for the framework are as follows:

- Data Transit confidentiality module.
- Process – Data Separation module
- Server and Client Authentication module
- Cryptographic Separation of Data module.

The above modules are centered on but not limited to cryptography, specifically Public Key Infrastructure operating in concert with single-sign on (SSO), security assertion markup language(SAML) and lightweight directory access protocol (LDAP), to ensure the authentication, integrity and confidentiality of the involved data and communications systems.

3.0 METHODOLOGY OF SOLUTION

Causes of Data loss and leakage such as *Insufficient authentication, authorization, and audit (AAA) controls, inconsistent use of encryption and software keys, operational failures, persistence and remanence challenges (disposal challenges), risk of association,* can all be mapped to the poor construct of the security architecture built around the cloud, the architecture is based on

- Centralized security
- Data and process Segregation
- Redundancy and high availability

These are in other words the sensitive features of the architecture as they carry along with them some of the key security features (confidentiality, Integrity and Availability) which are the building blocks used in designing secure system.

3.1 CORRELATION OF SECURITY FEATURES AND DATA LOSS /LEAKAGE

Confidentiality: This refers to only authorized parties or systems having the ability to access protected data. The threat of data compromise increases in the cloud, due to the following;

☑ Increased number of parties, devices and applications involved, that leads to an increase in the number of points of access.

☑ Delegating data control to the cloud, inversely leads to an increase in the risk of data compromise, as the data becomes accessible to an augmented number of parties.

☑ A number of concerns emerge regarding the issues of multitenancy, data remanence, application security and privacy.

☑ Multitenancy refers to the cloud characteristic of resource sharing. Several aspects of the information system are shared including, memory, programs, networks and data. Cloud computing is based on a business model in which resources are shared (i.e., multiple users use the same resource) at the network level, host level, and application level. Although users are isolated at a virtual level, hardware is not separated. Multitenancy, as multitasking, presents a number of privacy and confidentiality threats. Object reusability is an important characteristic of cloud infrastructures, but reusable objects must be carefully controlled lest they create a serious vulnerability.

○ Data confidentiality could be breached unintentionally, due to data remanence. Data remanence is the residual representation of data that have been in some way nominally erased or removed. Due to virtual separation of logical drives and lack of hardware separation between multiple users on a single infrastructure, data remanence may lead to the unwilling disclosure of private data. But also maliciously, a user may claim a large amount of disk space and then scavenge for sensitive data.

D*ata confidentiality in the cloud is correlated to user authentication*. Lack of strong authentication, inconsistent use of encryption and software keys can lead to unauthorized access to users account on a cloud, leading to a breach in privacy. In a cloud environment the user is required to delegate **trust** to applications provided by the organization owning the infrastructure. Software applications interacting with the user's data must be certified not to introduce additional confidentiality and privacy risks. Unauthorized access can become possible through the exploitation of an application vulnerability or lack of strong identification, bringing up issues of data confidentiality and privacy.

Integrity: Data Integrity refers to protecting data from unauthorized deletion, modification or fabrication. Managing an entity's admittance and rights to specific enterprise resources ensures that valuable data and services are not abused, misappropriated or stolen. By preventing unauthorized access, organizations can achieve greater confidence in data and system integrity. Additionally, such mechanisms offer the greater visibility into determining who or what may have altered data or system information, potentially affecting their integrity (accountability).

Integrity is correlated to authorization. Authorization is the mechanism by which a system determines what level of access a particular authenticated user should have to secured resources controlled by the system. Due to the increased number of entities and access points in a cloud environment, authorization is crucial in assuring that only authorized entities can interact with data to increase Data center reliability and prevent data leakage. A cloud computing provider is again **trusted** to maintain data integrity and accuracy.

Availability: This refers to the property of a system being accessible and usable upon demand by an authorized entity. System availability includes a systems ability to carry on operations even when some authorities misbehave. Thus a provider Audit his system from time to time, *Auditing is a mechanism which is correlated to availability*, it is referred to as efficiency or effectiveness check carried out as an independent assessor to determine the reliability of a system.
The system must have the ability to continue operations even in the possibility of a security breach. Availability refers to data, software but also hardware being available to authorized users upon demand. Leveraging users from hardware infrastructure demands generates a heavy reliance on the ubiquitous network's availability. A customer who is subscribed to such a cloud provider can reliably **trust** the cloud service since it is immune to operational failures and thus customer does not have to resort to drastic actions such as disaster recovery.

Having laid this correlation between security features, data loss/leakage and their causes, we observed that for a system to be designed we must establish a common base which the solution to the problem of data loss and leakage can be built on. This base is no other than trust.

TRUST: The notion of trust in an organization could be defined as the customer's certainty that the organization is capable of providing the required services accurately and infallibly. A certainty which also expresses the customer's faith in its moral integrity, in the soundness of its operation, in the effectiveness of its security mechanisms, in its expertise and in its abidance by all regulations and laws, while at the same time, it also contains the acknowledgement of a minimum risk factor, by the relying party [3]. The notion of security refers to a given situation where all possible risks are either eliminated or brought to an absolute minimum.

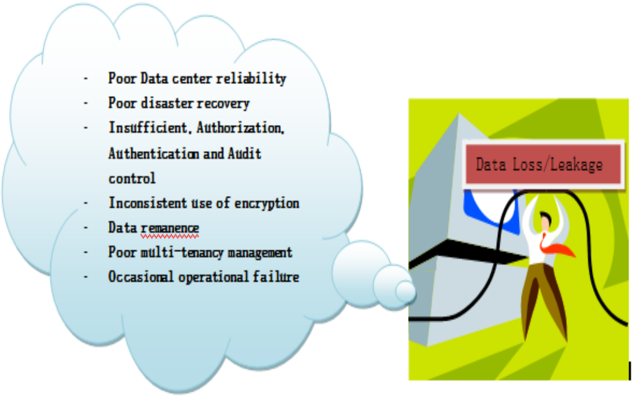
- Poor Data center reliability
- Poor disaster recovery
- Insufficient, Authorization, Authentication and Audit control
- Inconsistent use of encryption
- Data remanence
- Poor multi-tenancy management
- Occasional operational failure
Data Loss/Leakage

Fig.1 Factors that make up Data Loss and Leakage

3.2 PROCEDURE

We begin our framework by deploying a Trusted Third Party services within the cloud, this leads to the establishment of the necessary Trust level and provides ideal solutions to preserve the confidentiality, integrity and authenticity of data and communications. In cryptography, a Trusted Third Party (TTP) is an entity which facilitates secure interactions between two parties who both trust this third party. This layer is masked away from the user for security reason, as the user will engage in all interactions with the crowd service provider, not knowing that there is a separate third party handling the user's certification and access to the cloud services. The framework using the features provided by the TTP can be relied upon for the implementation of our modules:

- Data Transit confidentiality module.
- Process – Data Separation module
- Server and Client Authentication module.
- Cryptographic Separation of Data module.

☑ **Data Transit confidentiality**: Preventing data loss or leakage while data is in transit (data travelling) over the network is a hard and highly complex issue, while the threat of data modification and data interruption is continuously rising. A cloud environment increases this complexity as it does not only require protection of traffic towards the cloud but additionally between cloud hosts, as they lack a traditional physical connection. PKI enables implementing SSL for secure communications. SSL protocol generates end-to-end encryption by interfacing between applications and the TCP/IP protocols to provide client–server authentication and an encrypted communications channel between client–server. Due to the cloud environments unique characteristics, communications are required to be protected between users and hosts but also from host-to-host.

☑ **Process Data Separation:** In order to prevent operational failure and disaster recovery, we proposed the concept of separation of duty for cloud computing provider by targeting a most basic case where data need to be processed and stored. The main idea is to have two independent services responsible for data processing and data storage. Data are presented to users and are processed by the trusted Data Processing Service. When the data need to be stored, they are handed over to the Cloud Storage Service, which will make the data persistent and ready for retrieval in the future [3, 4].

The system (process – data separation) involves at least two independent service providers. Each service should be responsible for only one of the critical processes involved in a transaction. To maximize interoperability between communicating parties, it is a necessity to adopt widely used standards such as Security Assertion Markup Language (SAML), is an XML-based standard for exchanging authentication and authorization of data between the domains

Compartmentalization of data is also a key role performed by the cloud storage service so as to prevent malicious user from scavenging for sensitive data. Data confidentiality could be breached unintentionally, as a result of data remanence. Due to virtual separation of logical drives and lack of hardware separation between multiple users on a single storage infrastructure, data remanence may lead to the unwilling disclosure of private data, since a user may claim a large amount of disk space and then scavenge for sensitive data.

☑ **Server Client Authentication:** Digital signatures in combination with Single sign on (SSO) and Lightweight Directory access protocol (LDAP -is the internet standard way of accessing directory service that conforms to the X.500 data Model), implement the strongest available authentication process in distributed environments while guaranteeing user mobility and flexibility. The signing private key can be used to authenticate the user automatically and transparently to other servers and devices around the network whenever he/she wants to establish a connection with them. During the authentication process a user is required to navigate to his home organization and authenticate himself. During this phase information is exchanged between the user and his home organization only. After the successful authentication of a user, according to the user attributes/credentials, permission to access resources is either granted or rejected. The process in which the user exchanges his attributes with the resource server is the authorization process during which no personal information is leaked and can only be performed after successful authentication. The PKI certification authority is responsible for generating these required certificates while registering these within the web of trust. In other words, a Certification Authority builds the necessary strong credentials for all the physical or virtual entities involved in a cloud and it therefore builds a security domain with specific boundaries.

In the module, the service provider or the home organization may decide to implement a logger which monitors the activity of the user. This helps prevent malicious insiders from conducting any malicious activity that may hinder the performance of the services rendered by the server.

☑ **Cryptographic Separation of Data**: To remedy data remanence, multi-tenancy and improve privacy, we introduce cryptographic separation. Cryptographic Separation is a process by which processes, computations and data are concealed in such a way that they appear intangible to an outsiders. The protection of personal information or/and sensitive data, within our designed framework of a cloud environment not only ensures confidentiality and integrity, but also improves privacy of data which can be protected through encryption. Using a combination of asymmetric and symmetric cryptography (often referred to as hybrid cryptography) which can offer the efficiency of symmetric cryptography (host to host) while maintaining the security of asymmetric cryptography between the host and the service provider.

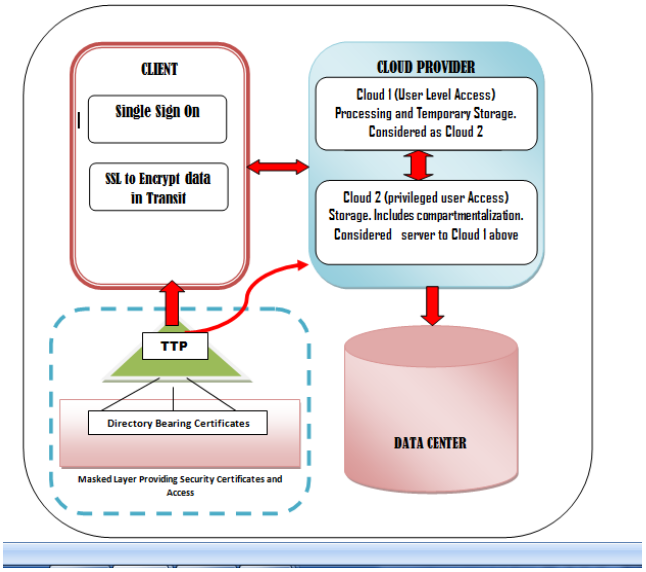
CLIENT
Single Sign On
SSL to Encrypt data in Transit
CLOUD PROVIDER
Cloud 1 (User Level Access) Processing and Temporary Storage. Considered as Cloud 2
Cloud 2 (privileged user Access) Storage. Includes compartmentalization. Considered server to Cloud 1 above
TTP
Directory Bearing Certificates
Masked Layer Providing Security Certificates and Access
DATA CENTER

Fig. 2 Proposed Framework for Treating Data Loss/Leakage

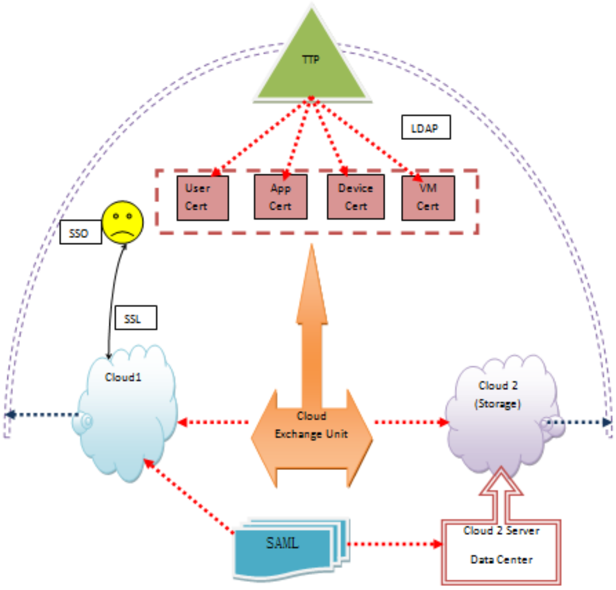

Fig. 3 Interaction Diagram for Proposed Framework

3.3 HOW IT WORKS

We have broken down the threat "Data loss/Leakage" by looking at its composition, and we have designed a framework which is built on trust which is leveraged by the trusted third party (TTP). To explain our model we begin by describing how a user access the services provided by the cloud provider.

Data in Transit confidentiality module operation introduces Digital signatures in combination with Single sign on (SSO) and Lightweight Directory access protocol; implement the strongest available authentication process in distributed environments while guaranteeing user mobility and flexibility. This operation is initiated when the user, using his smart card (SSO) authenticates his information which is transported in a secure connection to the TTP. Using a LDAP with a bind (authenticate) operation performs another authentication of user credentials before a certificate in the form of a session key is granted for the user to access the cloud. Access is provided to all subscribed entities and communication to the cloud entities is via an encrypted channel.

Session can last as long as the user intends to last using the cloud services but all activities conducted within the session are in a Bounded perimeter. Which means that all the interaction a user performs a within a dedicated boundary or domain. Separating his activities virtually from other user bound. This boundary implemented within the cloud provided to prevent malicious users from scavenging for sensitive data since they know they are sharing resources with other subscribed user (multi-tenancy).

Process – Data Separation module operation immediately set in as it also bears the same notion of boundary security but at this level the processing of data by application subscribed to is completely separated from the storage activity. For the Storage, The provider leverages another trusted cloud which will co-ordinate the storage processing. Before storage commences our fourth module comes into play. **Cryptographic Separation of Data module** data are concealed in such a way that they appear intangible to an outsiders this action is achieved via encryption which is also communicated back to the trusted third party also based on the application subscribed by the user each data is compartmentalized and stored separately in the data center. The cryptographic operations are transparent to the Data Processing Service and the Cloud Storage Service. The Data Processing Service and the Cloud Storage Service will not have access to the data without the cryptography key. Data access, such as reading and modifying the data, could be well protected by the cryptography key. In the case of encrypted data, the decryption key will be required. While in the case of digital signed data, all modification will need to be validated by producing new signatures with the needed keys. Of course the trusted third party also plays a vital role in this activity.

Communication between the cloud provider and the data center when retrieving data on-demand is done using a Security Assertion Markup Language (SAML) is an XML -based open standard for exchanging authentication and authorization data between security domains, that is, between an identity provider (Trusted Third Party) and a service provider. An SAML protocol is a simple request-response protocol which can ensure authenticity and provide authorization (at various levels)[5].

The Client Server Authentication module operates between the cloud providers, here the cloud that a storage cloud is leveraged to is the server while, the one providing the storage is in other words termed its server. For a user who has already being granted access to cloud 1 and all its device, a one-time private key is automatically generated (on-demand) should he choose to perform another operation involving cloud 2. The one-time private key authenticates the user to the cloud 2 server and transparently to devices around the cloud 2. During the authentication process a user is required to navigate to his home organization and authenticate himself. During this phase information is exchanged between the user and his home organization only. After the successful authentication of a user, according to the user attributes/credentials, permission to access the service is either granted or rejected. The process in which the user exchanges his attributes with the resource server is the authorization process during which no personal information is leaked and can only be performed after successful authentication. The PKI certification authority is responsible for generating these required keys/certificates while registering these within the web of trust.

In the module, the service provider or the home organization may decide to implement a logger which monitors the activity of the user should he decide to bypass authorization from his home organization. This helps prevent malicious insiders from conducting any malicious activity that may hinder the performance of the services rendered by the server.

4.0 DISCUSSION AND RESULTS

The Expected outcome of every security framework is to balance security and performance, our proposed framework employs the use of lightweight protocols/mechanisms in its implementation. Protocols such as LDAP, SAML, SSL, and SSO are flexible and elastic (capable of adjusting to increase in infrastructures deployed and easy to provision).

An end user is required to strongly authenticate himself with a cloud service and validate his access rights to a required resource. A secure SSL connection between user and provider is used for communication thus encrypting exchanged data and guaranteeing their security through the cloud infrastructure. The user is able to encrypt all personal data stored on the cloud to counter previously identified confidentiality risks. The cloud provider when communicating with the data center, can use his own certificate to authenticate himself in upcoming communications using a standard SAML protocol which provides security as well as interoperability between the two communicating entities. Also the use of certificate to encrypt and decrypt data can be enhanced to carry role information about a user or process (extended X.509 certificates). At the lowest level the hardware infrastructure are made available for use through the use Single sign on (SSO) authentication and authorization mechanism to enable secured communication between devices

The constant encryption and decryption of data could have a heavy toll on speed, inducing additional processing consumption. Using the cloud infrastructures flexibility within the context of demand on CPU, could leverage the system from this overhead and accelerate encryption/decryption. Also Key management is a critical issue in cloud infrastructures, thus we suggest that keys can be stored and protected in an environment deploying tamperproof devices.

Cloud security implementation such as the framework proposed above do take a centralized architecture, even though it is not a suitable security measure because of the “single point of failure” however, it easy to monitor, cheaper to implement, guarantees high performance due to the light nature of the mechanisms that were used to implement it and offers disaster recovery services.

5.0 CONCLUSION AND FUTURE WORK

However, based on the proposed framework a lot of ongoing research is has proved the idea to be viable: www.shibboleth.net Shibboleth is among the world's most widely deployed federated identity solutions, connecting users to applications both within and between organizations. Every software component of the Shibboleth system is free and open source. Shibboleth open source middleware software provides Web Single Sign On (SSO) across or within organizational boundaries. It allows sites to make informed authorization decisions for individual access of protected online resources in a privacy-preserving manner. Shibboleth technology relies on a third party to provide the information about a user, named attributes.

www. sandbox.safelayer.com Research & Development shows that, the aim of the PKI Webtop developed by sandbox.safelayer.com application is to provide a familiar environment, similar to the desktop of an operating system, to the user that provides security and trust services based on PKI technology for guaranteeing the authenticity, integrity, non-repudiation and confidentiality of application and Web-scenario data in a user-friendly and intuitive way.

Despite the security concerns in cloud computing, its benefits outnumbered its shortcomings. We have developed a framework aimed at minimizing data loss/ leakage security concern in cloud environment. Security in a cloud environment requires a systemic point of view, from which security will be constructed on trust, mitigating protection to a trusted third party. A combination of PKI, LDAP and SSO can address most of the identified threats regarding data loss/leakage in cloud computing and can help in addressing the issues of integrity, confidentiality, authenticity and availability of data and communications.

REFERENCES

[1] Wayne Jansen,Timothy Grance,"Guidelines on Security and Privacy in Public Cloud Computing" NIST ,2011.

[2]www.cloudsecurityalliance.org/guidance/csa guide-dom10-v2.10.pdf .

[3]Dimitrios Zissis, Dimitrios Lekkas, "Addressing cloud computing security issues", Future Generation Computer Systems Vol 28, Pp583-592, 2012.

[4] Gansen Z. et al (2010):Deployment Models: Towards Eliminating Security Concerns From Cloud Computing.

[5] Jason Goode, Ping Identity "The importance of identity security" Computer Fraud & Security, January 2012.

[6] Xiaojun Yu and Qiaoyan Wen "Design of Security Solution to Mobile Cloud Storage" Knowledge Discovery and Data Mining, AISC 135, pp. 255–263, Springer, 2012

[7] Asha Mathew "Security and Privacy Issues Of Cloud Computing; Solutions And Secure Framework" International Journal of Multidisciplinary Research Vol.2 Issue 4, April 2012, ISSN 2231 5780, Pp182-193

[8] Ayesha Malik, Muhammad and Mohsin Nazir, Security Framework for Cloud Computing Environment: A Review: Security Framework for Cloud Computing Environment: A Review VOL. 3, NO. 3, March 2012 ISSN 2079-8407, Pp390-394

[9] Prof Asha Mathew "Security and Privacy issues of cloud computing: solutions and secured framework international journal of Multidisciplinary Research, Vol.2 Issue4, (2012).