Data-Centric Security in Cloud Computing

[1]Meena Kumari, [2]Neelam Yadav
[1]Dept. of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana, India
[2]Dept. of Computer Science and Engineering, DAV College, Kanina, Haryana, India

Abstract
With the inception of Cloud Computing in 2006, academia has paid much attention towards it. Apart from its marvelous advantages it provides, it also comes up with various concerns like loss of control, lack of trust, Governance and compliance and security attacks. One of the most critical concern regarding cloud computing is of security. To be more specific data security in cloud computing paradigm. Recently, much attention is paid on securing networks and hosts (i.e. Infrastructure) holding data, but these existing solutions are not sufficient. This paper emphasizes the need of adoption of a new approach towards data security namely data or information centric security. This paper discusses various enabling tools behind data centric security and also presents a security framework. This scheme ensures the security of data whether it is in transit or in motion. Security procedure or modules are embedded in data itself rather than the container of data.

Keywords
Data-Centric Security; Confidentiality; Integrity; Infrastructure,; Tokenization; Masking; Object Based Storage;

I. Introduction

Cloud Computing refers to the use of networked infrastructure software and capacity to provide resources to user in an on demand environment. With cloud computing, information is stored in centralized servers and cached temporarily on client that can include desktop computers, notebooks and handheld devices [8]. While convenience is increased, the reduction of ownership over the machines,which store the data, decreases the overall sense of control and trust, especially when it involves sensitive data such as healthcare, government, or financial data. This lack ofownership and transparency are some of the main obstacles for data-sensitive industries to trust the cloud [9].

One of the best way to maintain the cost and efficiency benefits of the cloud and virtualization while keeping sensitive information secure, is to protect the data using a security solution that is data-centric. That is, the security system is not supposed to reside within walls of an enterprise but moves along with the data.

Data-centric protection through encryption makes the data unusable to anyone that does not have the key to decrypt it. No matter whether the data is in motion or at rest, it remains protected. The owner of the decryption keys maintains the security of that data and can decide who and what to allow access to the data. Encryption procedures can be integrated into the existing workflow for cloud services [1]. But one main problem with encryption is that while processing data on intermediate nodes it has to be decrypted as processing cannot be done on encrypted data and file size may also increase after encryption.There is also an issue of sharing of secret keys. Although fully homomorphic encryption scheme was proposed by which data can be processed while it is in encrypted form but it takes many more years to implement this scheme practically as accepted by Craig Gentry [2].

This paper describes the existing technologies and tools behind data centric security concept. A data centric security framework is also proposed in this paper.
The rest of the paper is organized as follows. Section II provides an overview of related work. Section III discusses some enabling technologies for content centric security. Section IV provides a detailed description of the proposed framework and section V gives some concluding remarks.

II. Related Work

This section briefly discusses the existing work in the field of data centric security. As this field is new not much work has been done so far.
A data centric security reference architecture is proposed in [11]. At the center of the reference architecture is data. Components like labels for specifying authorization criteria, indexes for ease of retrieval, and cryptography for an extra layer of protection were embedded within it. Author further added that Data-Centric Security solution encompasses features like Fine-grained, cell-level security enforcement, Data labeling capability, Policy specification capability, Encryption, at-rest and in-motion and Secure search.

[1] advocates the need of a security solution that encrypts the data at the file-level before it leaves a trusted zone. IT administrators and end users can exercise control on data that travels to and fro in cloud premises to a certain extent without putting sensitive data at risk.

[7] has highlighted two main drawbacks of traditional security mechanisms. The first one is incompatibility among existing security measures and second one is focus on the infrastructure security rather than the data itself. Therefore, the authors emphasizes on the need for integrating information-centric security approaches with existing traditional approaches, and hence presented a corresponding novel security framework.

In [11], authors presented Cloudsweeper, an implementation of the heterogeneous documents strategy as a cloud-based email protection system. Cloudsweeper helps in securing data at rest at cloud provider site. Author further highlighted that this cannot be incorporated for securing data in motion.

In [12], Securosis inc. illustrated the key enabling tools, technologies and deployment models behind working of data centric security. To name a few Masking, Tokenization, Format preserving encryption and Redaction. Data centric security in context of big data was also elaborated on.

In [13], author highlighted the main advantages of data centric security as protection of data wherever it goes, cost and complexity reduction, and increased end user transparency etc. Author further provides a detailed overview of the basic requirements for a true information-centric approach to data protection namely Smart data with embedded policies, universal policy language and user friendliness.

In [14], Susan Morrow introduced the concept of content centric security by incorporating the concept of Information cards. Entities possessing information cards and based on some claims can access data.

III. Enabling Technologies and Tools

This section deals with the enabling technologies available in literature for content centric security. These technologies were explored briefly in the following sections.

A. Object Based Storage

An Object Storage platform provides an infrastructure to store files along with metadata added to them – referred as objects. With Object Storage, there is no file system hierarchy. Instead of file system hierarchy all storage nodes are treated as a single pool where all objects have similar level or priority. Users access object storage through applications that typically use a REST API (Representational State Transfer). It is a software architecture that is used for distributed application environments, such as the internet. An API is used for an application (client) to talk to its environment (backend servers, storage, databases etc.)). An identifier is created while storing objects. This identifier is used to locate these objects from the pool. Structure of a typical object is depicted in Fig.1. Applications which needs to access these objects can quickly retrieve the data for the users through the object identifier or by querying the associated metadata (information about the objects, like its name, when it was created and by whom etc.). The object based approach enables a faster access and minimum overhead as compared to locating a file through a traditional file system [3]. Security procedures like labeling and indexing can be made an integral part of metadata to implement data centric security.

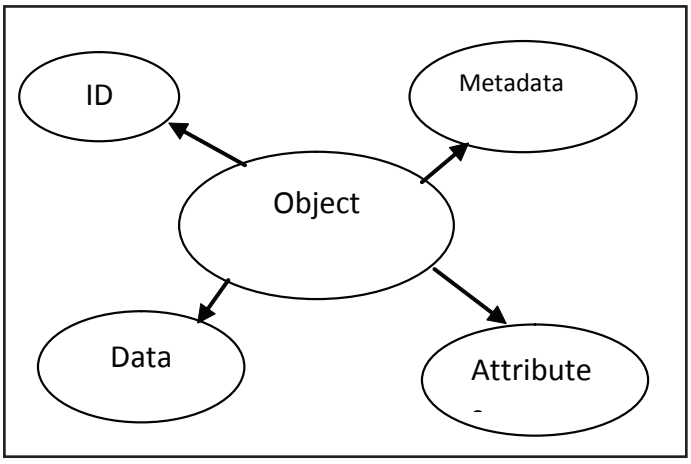

Fig. 1: Structure of an Object

B. Content Centric Networking

The NetInf [4-5] work by the 4WARD project is based on Content centric networking. The NetInf architecture is comprised of three major components: a naming scheme for Information Objects (IOs), a name resolution and routing system, and in-network storage for caching. The purpose of the name resolution and routing system is to link the IO IDs to the actual information objects, so that they can be queried upon. The naming scheme helps in making IOs independent from the devices storing them.

C. Tokenization

Tokenization is a process where critical data such as PAN number is replaced by any other irrelevant information called token [15]. In data centric security a token is provided in lieu of sensitive data. The token is an identifier that maps to the critical data through a tokenization system. The mapping from original data to a token uses methods which makes tokens infeasible to reverse in the absence of the tokenization system. Tokens replacement of data results is minimized exposure of sensitive data and reduction in risk of compromise to sensitive data [12].

D. Data Masking

Data Masking is the replacement of existing sensitive information with information that looks real but is of no use to anyone who might wish to misuse it [16]. Data masking is not same as limiting the visibility of data but data is visible to unauthorized entity but is of no use to them. Shuffling, substitution, number and date variance are some of the well-known techniques of data masking.

E. Format Preserving Encryption

Most of the encryption algorithms nowadays designed for files and disk volumes are not suitable for replacing data in databases. Once encrypted the encrypted data cannot be processed until unencrypted. Hence new advancements in this field of encryption made a data centrical solution viable. Format-preserving encryption (FPE) encrypts a plaintext of some specified format into a ciphertext of the same format—for example, encrypting a social-security number into another social-security number. Now encryption can be applied to any type of data — such as first and last names stored in a file — and still be processes by the application without error. And encryption keys can be provided to select users, keeping data secret from those not entrusted with keys [12].

This section discussed some of the tools and techniques which enable data centric security. Based on these tools a security framework is proposed in the following section.

IV. Proposed Framework

The proposed framework is shown in fig. 2. In the security framework proposed in [7] cloud operating environment and object storage and services are presented as separate modules but both modules can be merged because object storage is a storage (infrastructure) module in real sense. The proposed model incorporates the feature of Oracle labelling security for labelling of data items and for security policy enforcement. Following paragraphs discusses the various modules in detail.

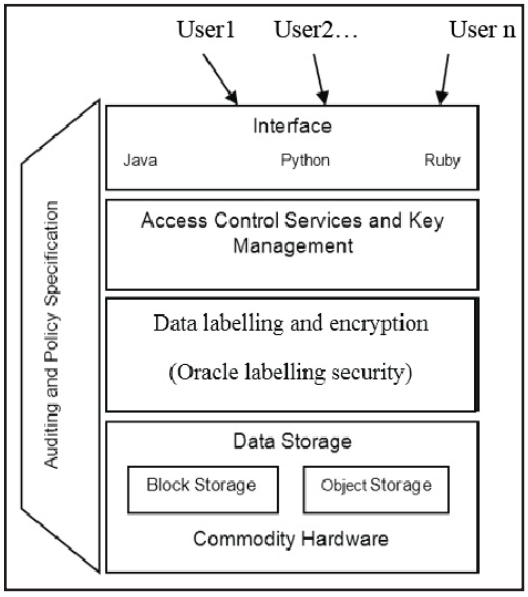

Fig. 2: Security Framework For Cloud Computing

A. Commodity Hardware

This layer deals with the IaaS (Infrastructure as a Service) layer of cloud computing. All the hardware required for networking and cores comes under this layer.

1. Data Storage

In cloud environment broadly two types of storage mechanisms can be adopted namely Storage Area Network (SAN) and Object based storage. In SAN dedicated storage nodes, organized in clusters are used to store data.SAN uses hierarchical storage structure and they have scalability issue. On the other hand Object based storage share a pool of objects without any hierarchical relationship. Objects are identifiable entities with a unique ID and associated metadata. In this framework Critical data is stored under object based scheme. Less critical data could be stored under block storage mechanism. Hence this model takes advantages of both the mechanisms.

B. Data Management Layer

As sensitivity of all information stored at cloud provider site is not same; hence uniform level of security is not required. For this purpose data should be labeled based on its criticality. Oracle labelling security enables to control the display of individual data items using labels that are assigned to data item and application users. Sensitive information visibility can be easily restricted to authorized users only. Oracle label security enables classification of data at row level and provides out of the box access mediation based on the data classification and the user label authorization or security clearance. It also helps in assigning label authorizations or security clearances to both database users and application users[17]. It works by comparing the row label with user's label authorizations and if it matches with the user authorization he/she may get access to that particular row. Visibility labels provide an assurance that a particular set of data item will be allowed to access by authorized user only.

Data whether in rest or in motion is also encrypted to add an additional level of protection.

C. Access Control Services and Key Management

As the name suggests this layer is dedicated to securely authenticate and authorize users. This layer also deals with the management of user credentials along with cryptographic keys.

D. Interfaces

Through this layer user will be able to access any cloud service (application or platform as a service).This layer provides a GUI or API, through which user can interact with the application.

E. Auditing and Policy Specification

This module addresses issues like:

1. Keep audit record of each and every request for data access. It not only maintains record for successful requests but also for failed requests. These audit reports helps in future analysis.
2. This module also specifies the privileges to individuals or groups to view data that has particular set of visibility labels after successful authentication of the entity involved.

This layer works throughout all the layers of the proposed model.

The model discussed above can fit in any cloud environment. Along with the specified layers some actions must also be undertaken to address the issues of governance and compliance.

IV. Conclusion

Despite of the fact that there are numerous advancements in the field of security in cloud computing but still there is no assurance of about being secure. The proposed scheme attempts to secure data in a cloud environment by adopting a data centric scheme. This paper investigates various enabling technologies for content or data centric security. A data centric security framework is also proposed which can fit in cloud environment effectively. Content centric security is a paradigm shift from traditional network or host based security. This does not mean that traditional container based security schemes were totally overlooked. Cloud computing security can be benefited by utilizing both the mechanisms. Through data centric security it is possible that data reach only intended audience.

References

[1] Kelley Diana,"How Data-Centric Protection Increases Security in Cloud Computing and Virtualization", Security curve white paper, 2011.

[2] Gentry Craig,"A Fully Homomorphic Encryption Scheme", Dessertation of Ph.D, Sept. 2009.

[3] Leyden Tom,"A Beginner's Guide To Next Generation Object Storage", Data Directs Network White paper, 2013.

[4] B. Tarnauca et al., Netinf evaluation, EC FP7-ICT-4WARD Project, Deliverable D-6.3, June 2010 [Online] Available: http://www.4ward-project.eu).

[5] Jacobson V. et al.,"Networking named content", In Proc. of CoNEXT'09 - 5th International Conference on Emerging Networking Experiments and Technologies, Rome, Italy, Dec. 2009.

[6] Ahlgren Bengt et al.,"Content, Connectivity, and Cloud: Ingredients for the Network of the Future", IEEE Communications Magazine, 2011.

[7] Aiash Mahdi et al.,"Introducing a Hybrid Infrastructure and Information-Centric Approach for Secure Cloud Computing", In: Proc. of 29th IEEE International Conference on Advanced Information Networking and Applications, Korea, Mar 2015.

[8] Tribhuvan MR et al.,"Ensuring data storage security in cloud computing through two way handshake based on Token management", In : Proc. of International conference on advances in recent technologiesin communication and computing,2010, pp. 386-389.

[9] Ko R.K.L. et al.,"Towards Achieving Accountability, Auditability and Trust in Cloud Computing ", In: Proc. of International workshop on Cloud Computing: Architecture, Algorithms and Applications, Springer, 2011, pp. 5-18.

[10] Sqrrl data inc,"Big Data and Datacentric Security", White Paper, 2014.

[11] Snyder Peter et al.,"Cloudsweeper: Enabling Data-Centric Document Management for Secure Cloud Archives", In: Proc. of ACM cloud computing workshop, 2013.

[12] Lane Adrian,"Securosis-Trends in Data Centric Security", Secrosis White paper, Version 1.0, September 14, 2014, pp. 8-16.

[13] Khosla Pradeep,"Information security for the next century", Carnegie Mellon CyLab.

[14] Morrow Susan,"Data Security in the Cloud", Cloud Computing Principles and Paradigms, Edited by Rajkumar Buyya et al., John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 573-586, 2011.

[15] SIG Scoping,"Information Supplement: PCI DSS Tokenization Guidelines", PCI Data Security Standard (PCI DSS), Ver2.0, August 2011.
[16] NET2000 ltd.,"Data Masking: What You Need to Know", [Online] Available: http://www.DataMasker.com, 2010.
[17] Oracle label security administrator's guide,12 C release 1(12.1), June 2014. [Online] Available: https://docs.oracle.com/database/121/OLSAG/E48437-04.pdf

Meena Kumari has received her MCA degree from Kurukshetra University, Kurukshetra, Haryana, India in 2012. Currently she is pursuing her PhD degree from Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana, India. She is working as an Assistant Professor in the Kurukshetra University. Her research interests includes Cloud Computing, security, cryptography, hashing and computer networks.

Neelam Yadav received her B.Tech. Degree in Computer Science and Engg. from Maharishi Dayanand University, Rohtak, Haryana, India in 2008 . Currently she is teaching as an assistant professor in Computer Science Department in DAVCET, Kanina, Haryana, India. Her research interests include networking, image processing and programming language and cloud computing.