



# **International Journal of Advanced Research in Computer Science and Software Engineering**

**Research Paper** 

Available online at: www.ijarcsse.com

# A Three Layered Security Model for Data Management in Hadoop Environment

# Monika Kumari

Scholar, M.tech

Department of Computer Science and Applications KurukshetraUniversity,Kurukshetra, India Dr.Sanjay Tyagi

Assistant Professor

Department of Computer Science and Applications KurukshetraUniversity,Kurukshetra, India

Abstract— Security is one of the basic aspects required in any network model. But when the access is on any shared system such as Hadoop, the security criticality is increased. This kind of systems is defined along with service and resource sharing services as well as to perform the data management effectively. These services are integrated with public as well as private environment. The Hadoop system increases the criticality because of available limited resources in mobile devices. In this present work, a three stage security model is presented for Hadoop Environment. The work is applied on secure file management and distribution over the secure Hadoop environment. The paper includes the exploration of the proposed security model.

Keywords — Authentication, Cloud, Data Management, Hadoop, Security.

### I. INTRODUCTION

Cloud computing is the evolutionary distributed platform to provide the services, resources and the hardware in an integrated environment to cloud users. It helps a user to use the storage system, hardware and the application software without performing any deployment or installation. Cloud computing is becoming one the most popular technology among the business enterprises because of infrastructure reduction and cost reduction. The users are also attracted to this environment because of the fast and integrate service access over the cloud system. The cloud system itself defines different platforms, services, applications to all public, private and limited users. Beyond the effective integration between the cloud servers and clients, it also suffers from security challenges because of its global virtual environment [1] [2].

Security is the key issue associated with cloud system that is required on the client side as well as on the vendor side. The security requirements in this public environment are shown in figure 1. The main consideration among these issues is the authentication and the authorization issue. This security concern shows the threats are again the hacking and the malware activity in the cloud system [3]. Once the authentication is proven, the next work is to perform the secure communication so that the reliable data will be transferred to cloud server and to client side in a secure way. Another security concern of cloud system is the authorization as well as access control. Authorization is about to avail the services, products or the resources based on the profile match as well as to keep safe the information from others. The profile match defines the user level identification to achieve the security. The trust level analysis also comes under the security specification.

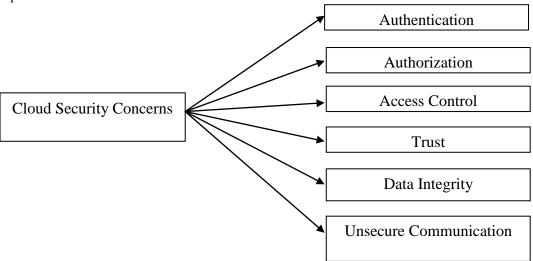


Figure 1 Security Concerns in Cloud System

The trust is analyzed for the customer as well as the vendor. The trust certificates are distributed to prove the trust. The data integrity is the security issue that deals with the data distortion or the error generated in the data communication or the availability. The most concerned issue with the security system is the communication level security [4] [5]. When the data is being transferred, the issue can be in the form of attacks or the incomplete transaction. The session level security is defined to handle this kind of problems in cloud system.

### A. Cloud Service Architecture

Cloud system is an organized architecture that is defined in several means. One of such effective representation is the service level based architecture. This architecture is defined with three main service layers or the model called IaaS, PaaS and SaaS. The IaaS (Infrastructure-as-a-Service) is described as the machine on demand service that avail the physical resources or the hardware in the form of remote service to the customer. PaaS (Platform-as-a-service) is defined as the complete application environment by using which the developers can interact with development software in a shared remote server system. SaaS (Software-as-a-Service) gives the concept of public cloud where an end user can interact with the system in an integrated environment and multiple vendors are available to provide the requested services [6]



Figure 2 Cloud Architecture.

(Source.http://www.csoonline.com/article/2126885/cloud-security/saas--paas--and-iaas--a-security-checklist-for-cloud-models.html)

In this paper, the security aspects related to the cloud service model are explained. These aspects include the issues and the relative solutions. In this section, the exploration of the cloud system and its security concerns is defined. This section also explained the cloud service model. In section II, the Hadoop Architecture is defined. In section III, the work cloud service security models are explained along with the issues and the solutions. In section IV, the proposed model is presented with section V showing results. In section VI conclusion of the paper work is described.

# II. EXISTING WORK

Security is always one of the most common and open research area, because of this lot of work is already done in the area of security system in a cloud environment. In this section, some of the work done by the earlier researchers is discussed. V. D. Cunsolo [3] performed a work to achieve the information security in distributed systems. To resolve the security problem in network based distributed system, the author suggested a light weighted cryptographic approach. The objective of work was to provide a secure asymmetric approach to provide secure communication of data as well as file system. Author proposed a secure distributed file system with asymmetric or symmetric structure. Author defined the secure interfacing with cloud and grid based systems.

Christian Schridde [4] provided a secure cloud infrastructure based work to provide the security over the cloud system. Author presented a secure infrastructure to provide service over the cloud environment. The work includes the identity based cryptographic model based on a public key system. Author provided the cloud based data transmission under the trust analysis. Author also provided the comparative analysis between the approaches.

Yingjie Xia [5] defined an ECC model over the cloud system to improve the security on cloud system. Author defined a hybrid ECC system for cloud data. It provided a platform to provide secure file communication, backup system and the resource sharing on distributed cloud. Author provided different security levels for different kind of cloud and avail different secure services with confidential protocol and privacy. Author combined the hash key based cryptography and enhanced it using ECC to provide secure user control system.

Neha Tirthani [7] has defined a work on data security and integrity in a cloud environment to perform reliable service distributed in a cloud network. Author defined a data or storage oriented secure service distribution mechanism so that the service distribution benefit will be taken by the cloud users. Author defined a work on key based authentication for cloud security analysis. Author used a combined secure approach for information sharing using ECC and Diffie Hellman approach. Author used the symmetric bivariate polynomial information sharing system for cloud environments. Author defined a trusted third party system where multi-server system is extended to get fit to the environment. Author defined a multi server system so that effective, secure service provider is established. Author proposed an effective, secure service mechanism in a cloud environment.

Yanping Xiao [8] presented a secure middleware in a cloud environment. Author defined a survey on this middleware technology under different platforms such as AppScale, Altocumulus and Cloudify etc. Author also presented the

analytical study over the cloud to achieve the secure integration and provide a security standard for the future technology research.

M.Venkatesh [9] defined a work over the secure data storage in cloud system with public audit ability. Author uses the internet feature and software support to improve the communication capability in the cloud system. Author defined the secure remote communication to utilize the cloud resources. Author used the RSA based secure storage system with public auditing to improve the cloud system. The public key cryptography is here implemented to improve the security support along with reduction of the computation time on cloud system. Obtained results show that the work has improved the security over the existing method.

Vasyl Ustimenko [10] presented a key based secure and scalable cloud environment for the application based security. Author provided a trustful cloud environment to provide the secure communication based on secure key management scheme. Author provide provided the secure application mechanism to achieve the coordination between the owner and multiple users.

Dexian Chang [11] defined a trust analysis in a cloud environment. Author defined the trusted relationship over the cloud environment under the flexibility and scalability parameters. Author defined the cloud virtualization under the different user domains. Author defined a trusted service domain for multiple user domains to achieve the cloud virtualization platform. Author also provided the inter domain communication and migration facility to provide the reliable communication over the system.

Chang-Ji WANG [12] provided the attributed oriented encryption analysis with constant size with the cipher text. Author provided a new cryptographic algorithm to provide finer grained data sharing with decentralized access control system. Author defined the secure key policy system with cipher text and to achieve the attribute and private key association over the system. Author provided the trustful cloud storage over the cloud system under the KP-ABE scheme. Author defined an application level security system to embed the security under the cloud storage environment. Author defined the monotonic structural access over the cloud system and also provided the secure key exchange mechanism using a Diffie Hellman algorithm.

### III. HADOOP ENVIRONMENT

Hadoop is a substitute to Cloud environment, but it also gives the extension to the traditional cloud architecture. This extension is in terms of service and new features included in cloud structure. These services and features are included in the cloud environment in terms of the API so that the new integrated cloud storage and synchronization application can be designed. Hadoop provides a free service for the storage and fee based architecture so that effective and secure storage of data, photos and other media information can be stored. The cloud environment is defined with big data centers of Hadoop servers. Apple provides such cloud architecture so that Application free environment will be generated [13] [14].

Hadoop is fully integrated with mobile devices including the iPhone, iPod, iPad etc. Different platform environment supports the Hadoop architecture. The Apple TV and computer based operating system so that the use of parts of Hadoop, photos and music. This cloud system is defined with SaaS model along with the integrated IaaS model [15]. The architecture of Hadoop system is shown in figure 3.

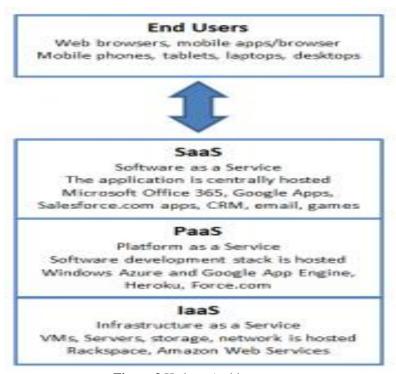


Figure 3 Hadoop Architecture

(Source http://tritoneco.com/cloud-iaas/cloud-word-of-the-day/)

Once the Hadoop is activated, the user can choose the settings respectively to the supported applications. These settings are data oriented to identify Hadoop is storing the data or not. There are a number of separate setting page so that the relative options will be selected and identified. Hadoop defined the work on internet connection. Author defined the connection based on different versions of the document. Hadoop is specially designed for apple applications. It also controlled by Microsoft Windows and the control panel so that the mails, nodes and photo features will be transmitted effectively [16].

## IV. PROPOSED MODEL

Today, instead of maintaining the data on individual systems, whole data and information are generally placed on some centralized system with distributed environment. Such distributed system can have multiple service providers as well as multiple users. This kind of environment is provided by Hadoop environment. A Hadoop is the distributed system for the new era that provides the shared and distributed infrastructure, services and the products. It provides a model based environment adapted by most of the web clients to avoid the individual installation of software, security etc. As the Hadoop system is open, publicly using the internet, it is having the main challenge in the form of security.

The presented work is about to provide the secure communication with data, Hadoop for the public and private access over the system.

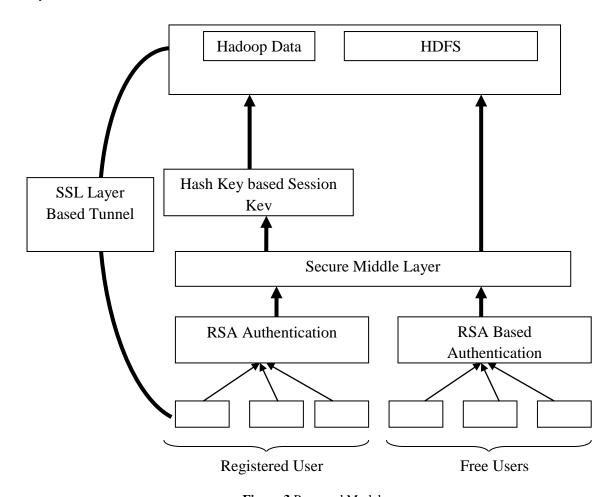


Figure 3 Proposed Model

In most of the existing approaches a generalized cryptographic approach is implemented to achieve the security over the Hadoop system. In this work a user's perspective security scheme is being presented. According to this approach, to a secure tunnel based transmission is provided for the frequent communicating authenticated users. For such users, one time authentication will be performed using RSA algorithm. Once the session is established, SSL layer is activated to provide the secure transmission over the Hadoop. The second level of security is provided for the authenticated Hadoop users that avail the Hadoop services rarely. For such users, each time authentication is performed using the RSA approach, but no tunnel will be defined. At the final stage, for the free uses, an RSA based authentication will be performed and allow the public area for the access.

As shown in the figure, the Hadoop server is having the raw data or the file as the available data resources. A system can have single or multiple Hadoop system. This Hadoop server is the top layer that will provide the resources to all users publicly. The users that will perform the data request can be registered user or the free visiting users. The security is here mainly incorporated for the registered users. To provide the security over the system, security is here implemented in the middle layer called the security layer. The work of this security layer is divided in three parts.

#### A. Authentication

The authentication is here provided at two levels. For the free users, the authentication is provided using an RSA cryptography approach where as for the registered users, the authentication will be achieved using hash key based RSA algorithm. As the user will enter into the system, the authentication check will be performed using the cryptographic approach. A free user is a visiting user that can visit the public pages of the Hadoop, but cannot perform any data oriented operation over the Hadoop. But the registered user is allowed to perform the data downloading on Hadoop.

#### B. Secure Session

If the authenticated registered user wants to download some data from the Hadoop server, the session key will be generated. This key will be activated for the specific period. As the session will be established, the next work is to perform the secure data transfer at the client end of the server. To perform this secure transmission SSL enabled secure tunnel will be generated between the client and the server with specific bandwidth. The communication will be performed using this tunnel. As the communication will end, the session key will be deactivated.

# C. Secure Data Management

Data over the cloud will be managed in the cryptographic form. To perform the data encryption over the cloud the RSA based cryptography approach will be implemented.

### V. RESULTS

The presented work is implemented in an integrated Hadoop environment. The work is applied on real time data or files. The work is tested on the text files. As the work is applied, the first stage is to perform the registration of a new user to the environment. After the generation of user, the security aspects are presented in the form of secure file management. That includes the generation of dynamic key and dynamic cryptographic operations using RSA approach. The work includes the secure authentication, data management and the secure transmission over the system. Here figure 4 is showing the authentication screen for cloud system.

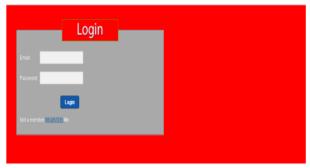


Figure 4 Authentication System

After entering into the system, the secure file management is being applied. Here figure 5 is showing the security management for the particular file and folder. In the same way, the authenticate communication is managed over the Hadoop environment.

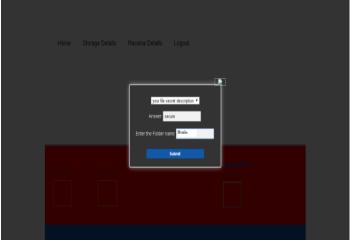


Figure 5 Secure Data Management

# VI. CONCLUSION

In this paper, an exploration of the Hadoop service model and the integrated security aspects is defined. In this work, a three stage security model is suggested that combines the authentication, secure data management and secure data transmission over the system.

### REFERENCES

- [1] Minqi Zhou, Rong Zhang, Wei Xie and Weining Qian, Aoying Zhou, "Security and Privacy in Cloud Computing: A Survey", Sixth International Conference on Semantics, Knowledge and Grids., 2010.
- [2] Jianfeng Yang and Zhibin Chen, Cloud Computing Research and Security Issues, IEEE, 2010.
- [3] V. D. Cunsolo, "Achieving Information Security in Network Computing Systems", Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.
- [4] Christian Schridde, "An Identity-Based Security Infrastructure for Cloud Environments", Proc. of IEEE International Conference on Wireless Communications, Networking and Information Security WCNIS2010, 2010.
- [5] Yingjie Xia, "Hierarchy-Aware ECC Model for Cloud", 2nd International Conference on Industrial and Information Systems, IEEE 2010.
- [6] M. O'Neil, "SaaS,Paas,Iaas," 31 january 2011. [Online]. Available: <a href="http://www.csoonline.com/article/2126885/cloud-security/saas--paas--and-iaas--a-security-checklist-for-cloud-models.html">http://www.csoonline.com/article/2126885/cloud-security/saas--paas--and-iaas--a-security-checklist-for-cloud-models.html</a>. [Accessed may 2014].
- [7] Neha Trithani and Ganeshan R, "Data Security in Cloud Architecture Based on Diffie Hellman and Elliptical Curve Cryptography", 2013.
- [8] Yanping Xiao, "An Efficient Privacy-Preserving Publish-Subscribe Service Scheme for Cloud Computing", IEEE 2010.
- [9] M.Venkatesh, "Improving Public Auditability, Data Possession in Data Storage Security for Cloud Computing", IEEE 2012.
- [10] Vasyl Ustimenko, "On some mathematical aspects of data protection in cloud computing", IEEE 2012.
- [11] Dexian Chang, "TSD: A Flexible Root of Trust for the Cloud", 11th International Conference on Trust, Security and Privacy in Computing and Communications, IEEE 2012.
- [12] Chang-Ji WANG, "A Key-policy Attribute-based Encryption Scheme with Constant Size Ciphertext", Eighth International Conference on Computational Intelligence and Security, IEEE 2012.
- [13] Apache hadoop nextgen mapreduce (yarn), <a href="http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn/hadoop-yarn/site/YARN.html">http://hadoop.apache.org/docs/current/hadoop-yarn/hado
- [14] An introduction to the hadoop distributed file system, <a href="http://www.ibm.com/developerworks/library/wa-introhdfs">http://www.ibm.com/developerworks/library/wa-introhdfs</a>.
- [15] Vishwas Churihar, "*Knowledge Management in Cloud Using Hadoop*", International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV,2013.
- [16] Jared Evans CSCI B534 Survey Paper. "Fault Tolerance in Hadoop for Work Migration" <a href="http://salsahpc.indiana.edu/b534projects/sites/default/files/public/0\_Fault%20Tolerance%20in%20Hadoop%2">http://salsahpc.indiana.edu/b534projects/sites/default/files/public/0\_Fault%20Tolerance%20in%20Hadoop%2</a> Ofor %20Work%20Migration Evans ,%20Jared %20Matthew.pdf