International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 7 Issue 6 , June-2017, pp. 295~302
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196

SECURED SYSTEM TO ACCESS HDFS DATA USING MiLAMob MIDDLEWARE-LAYER FRAMEWORK

Surbhi Singh

Research Scholar,

Department of Computer Science & Engineering

Deenbandhu Chhotu Ram University of Science & Technology, Sonipat

Abstract- Rapid advancements in the computing technologies and constant reduction in the valuable computing resources led the people to depend upon the cloud computing for the reliable and remote data accessibility However cloud era has arrived with several issues including security, confidentiality, availability policies and standards. This research work proposed the secured public cloud computing model to improve user's mobility, secure data access dealing with the user's identity tokens and storing huge amount of data related to mobile consumers on cloud.

This paper simplifies the authentication, authorization and token storage problems by introducing the middleware-centric framework called MiLAMob being implemented with the public cloud service i.e. HDFS. In order to reduce storage space, processing and the communication latency generated due to submission of huge amount of identification credentials by the consumer, it initially registers the identity tokens without storing them and then matches it with the identity credentials stored on the MiLAMob framework. Secondly, it improves the authorization process by implementing the flexible user based policies. This model employs the OAuth 2.0 technique to provide secure access to users via social networking sites also.

Keywords- Cloud Computing, Authentication, Authorization, OAuth 2.0, middleware framework, HDFS.

1. Introduction

In the fast pacing world of technology revolution, huge amount of data generated in daily operations and eventually increasing demands for the secured remote data accessibility by the mobile consumers across organizations is being pushing them to outsource their private and public data on the cloud storage services called Storage Service Providers (SSP) such as Amazon S3, Yahoo briefcase, HDFS, Gmail etc. However, comparing the present scenario with the previous one where designing and deployment of an application was of high concern, developers now aim at making online cloud storage more secure. The reason being, keeping data on huge cloud has rendered users from having the full control over its security, accessibility and the privacy.

CIA (Confidentiality, Integrity, Availability) are the three prime important features to be fulfilled by any networking services. Integrating all the above properties to store the data securely over the cloud storage services has become the critical component for the enterprises as well as for the mobile consumers. To provide the security to the consumer's data, enterprises widely adopted private cloud deployment which employs internal data protection policies. But as the communication between different organizations and the user's increases, there is a great demand for implementing security to access private data stored on the public cloud services such as HDFS.

This paper addresses the following questions efficiently:

- How the data stored on cloud hosted platform: HDFS can be securely accessed by the mobile users and that too with minimum HTTP traffic?
- What authorization mechanism grants the users of different access privileges to fetch data from the private HDFS data objects without even providing the security credentials all the time assigned by the IAM service?
- Using the randomly generated and not stored identity tokens (biometric) for authentication, how the verification of the mobile users to access the cloud stored data is being done?
- How the social networking sites can be used by the mobile consumer to securely access the data?

In order to answer the aforementioned questions, this paper presents more intrinsic approach of providing security in terms of authorization and authentication by using middleware-centric framework called MiLAMob.

2. Literature Survey

Data Mining in Cloud computing, this paper deals with data mining and how it is used in cloud computing. As already explained mining is the process of extracting potentially useful links or information which can be used for forecasting from raw data. The paper also explains how SaaS is useful in cloud computing. There is regular confrontation with target advertising and many businesses have become more efficient and economical by using data mining techniques, which are becoming very common. Data mining applications can derive information about human population and concerning customers that was earlier not known or hidden in warehouses. Recently there has been an overwhelming raise in the use of data mining techniques which target applications such as identifying suspects in crime, fraud detection, and prediction of potential terrorists. It is assumed that mining systems which were developed for grids, clusters and distributed clusters have faced scarcity of processors and other resources and hence they are shared. When processors or other resources become available, then the data is moved to the processors.[8]

In paper Security and Privacy in Cloud Computing: A Survey, there is a complete analysis of security issues in cloud environment. Big companies like Google, Amazon, and Microsoft are stepping into cloud computing but obviously there are major security and privacy concerns. Availability, audit, control, confidentiality, data integrity are the major security aspects. Multiple locations of data and services also make privacy concerns more prominent. Cloud will flourish even more once these security and privacy issues have been resolved.[9]

In the paper titled Secure Computing in the Cloud, Cloud Computing offers a lot of benefits for end customers like high-end machines, large amounts of storage, high availability and everything available at the touch of a button. This paper concentrates on compute clouds, which are clouds, which do not only offer storage but also computations that can be outsourced in form of virtual machines. Outsourcing computations as well as data to a third party, in this case the cloud provider, are accompanied by the uneasiness of confiding data to the cloud provider based only on service level agreements. Some other issues discussed are the involved risks; create threat models as basic assumptions that describe the untrusted cloud entities and present solutions that augment trust in the cloud provider, the integrity and verifiability of computations and data processed in the cloud.[10]

In one of the paper, it discusses that how advantages like flexibility, cost efficiency and availability of cloud services comes with the risk of privacy and security and privacy users' data. So privacy concern has become a major barrier in the minds of users and a hindrance to the growth of cloud computing worldwide. One of the major concerns of cloud is data mining based privacy attacks that involve analyzing data over some time to extract information which can be very valuable. In current cloud architecture a client stores data with a single cloud provider. It gives the cloud provider and some outside attackers' unauthorized access to cloud and an opportunity to analyze client data over a long period to extract sensitive information that causes privacy violation of clients. So data security is a big concern for many clients of cloud. This paper identifies the data mining based privacy risks on cloud data and proposes a distributed architecture to eliminate certain risks.[11]

3. The MiLAMob Framework

In this research work, highly efficient middleware-centric authentication model called MiLAMob is being implemented with the HDFS (Hadoop Distributed File System) so as to authenticate the enterprise consumers to access their data files stored in the HDFS. This model has already been implemented with the public cloud service Amazon S3 and has proved it's efficiency by minimizing the communication latency generated due to the large header size; data protection has been achieved as it provides semi-transparency to the users and even to the administrator by hiding the security credentials from them, saves a lot of bandwidth of the communication channel. It employs the use of open standard authentication approach OAuth 2.0 protocol for login and accessing data via social networking services such as facebook, google+, yahoo, twitter etc.

3.1 Authentication

The growing use of Internet web applications gives rise to the problem of managing the necessary digital identities and preserving their privacy. In an open large-scale domain such as the Internet, preserving user privacy is not a straightforward task. Identity theft, which occurs when an impostor uses a legitimate user's identifying information without his/her consent, is becoming one of the biggest security concerns both for users and for organizations accessing data stored on cloud. In order to reduce the storage space and to provide secure authentication it implements multi-layer biometric authentication architecture in it and prevent relying on third party for claims. The multi-layer architecture comprises of the following components as shown in figure 1.

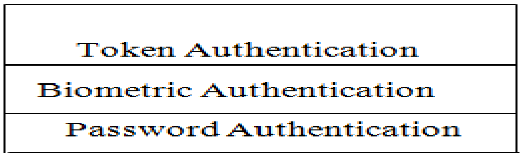

Fig. 1 Multi-Layer Architecture

The digital identities are first registered into the cloud directory and then every time these are matched with the stored credentials. The token is generated by taking the hash of the password entered by the user plus biometric template and the required file number that is to be accessed by the user as shown in figure 2.

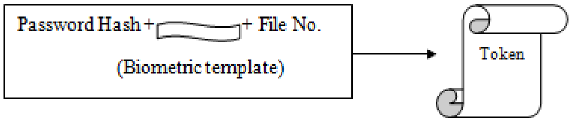

Fig. 2 Token Generation

This token is then either send to user's phone number or email id and is not stored on the cloud storage.The next time when user access the data he will first enter the password, then provide biometric pattern and finally the previously generated biometric token. If the whole template matches then he can access the data. The process is shown below in figure 3.

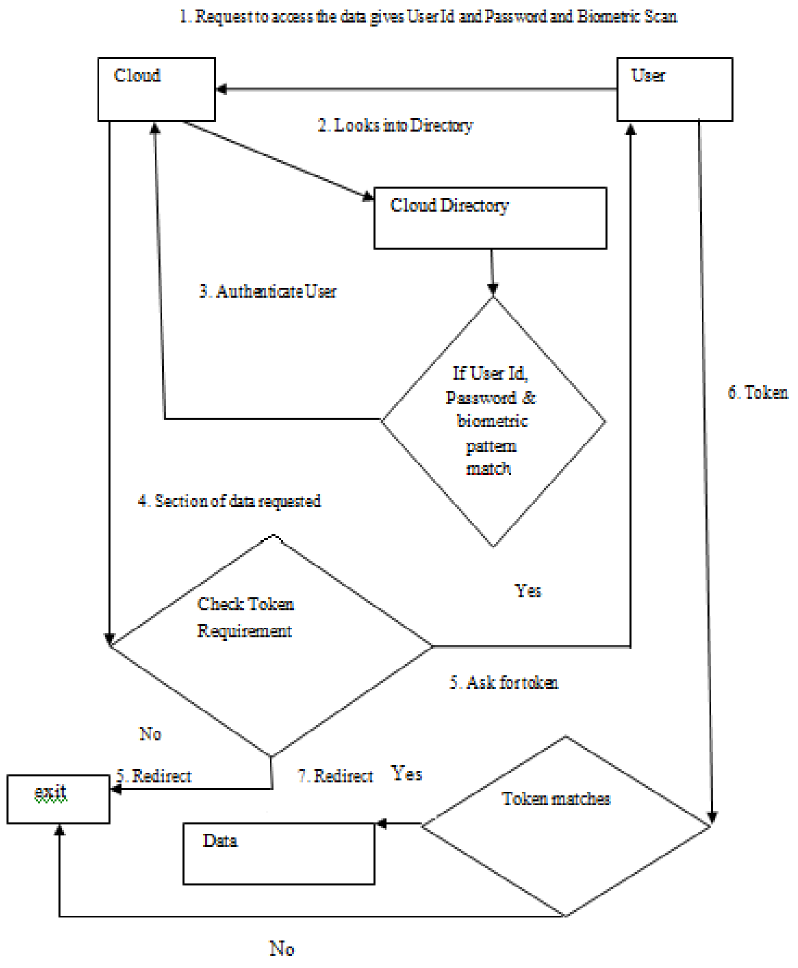

Fig. 3 User Authentication using biometric tokens

3.2 Authorization

This research paper aims at providing sophisticatedly implemented and secure data accessing architecture which keeps in mind both of its complexity and compatibility with the other systems. The authentication mechanism being discussed in the previous section uses three layered data access gateway which encompasses biometric technique along with the password and token authentication. This framework previously implemented on Amazon S3 cloud service uses simple token generation technique which in turn allows users to access their data stored on the cloud. Due to its loopholes, MiLAMob architecture is now implemented with HDFS storage service with improved security.

In order to make this architecture more computationally infeasible to password cracking, use of one-time password is made. When the user first accesses the system, the framework generates random password which can be used only for that session. Next time, the token which is being used for accessing the data stored on cloud, will be generated from previously used one-time password and then highly complex AES encryption technique convert that result into some hash value which will be going to store in the repository. Getting back to the original value from this hash value is not possible; hence it makes the whole framework complex to be cracked and easy, secure for the mobile consumers to access their private files.

Regarding the data access permissions, MiLAMob framework solves the great complexity of providing group access policies in a more refined way. The following code provides strict data access privileges to the whole group instead of giving individuals the separate set of permissions each and every time:

```
{    “Statement”:

  [
          [
              “Effect”: “Allow”,
              “Action”: “*”,
              “Resource”: “*”
          ]
  ]
}
```

The HDFS cloud service employs different user group policies to provide access to data. In this model, it follows the process of providing some specific individuals within a group with some special data access privileges instead of giving each one of them different access rights to reduce the complexity and processing effort. MiLAMob framework takes less request-response time as compared to the traditional TVM (token vending machine). The basic framework has been shown in the following fig 4.

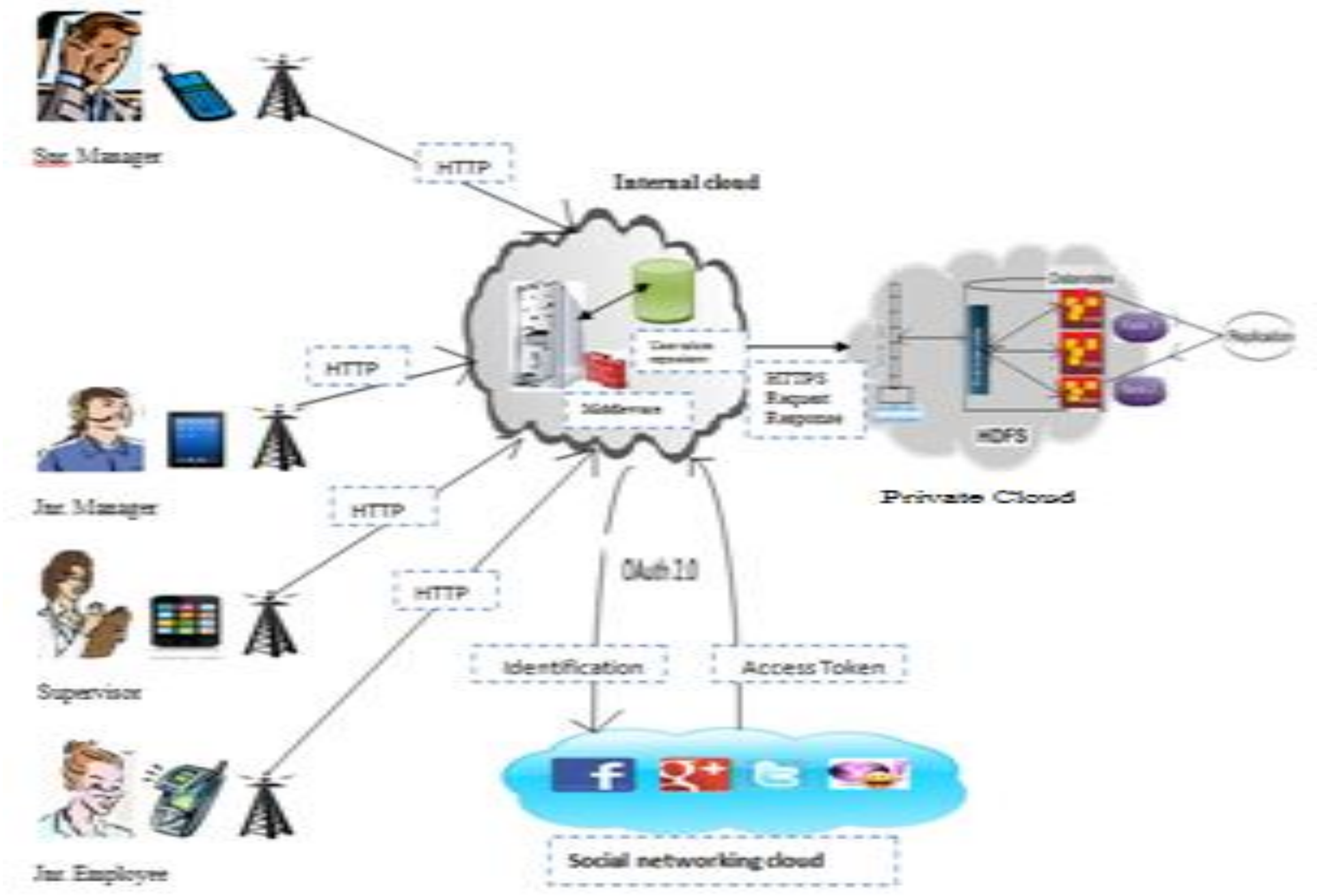

Fig. 4 Architecture design of MiLAMob framework interacting with HDFS [7]

The MiLAMob reduces the HTTP header length by approximately 70 bytes whereas TVM uses 520 bytes header, thus improves the performance.

4. COMPARISON WITH TVM

The traditional Token Vending Machine (TVM) are the proxy based applications that are used to generate the temporary security credentials used to authenticate the end users to access the cloud stored data. But as the keys are not persistent, this might generate a lot of overhead of sending the request over the HTTP to access the data.

The MiLAMob provides an efficient performance over traditional TVM as it reduces the HTTP header size to approximately 70 bytes only as compared to 520 bytes by TVM as the request process and credentials are now handled up by the middleware-oriented framework instead of the mobile users directly communicating with the cloud stored data. Thus proving itself a more secure, highly reliable and efficient framework over TVM as shown in Fig. 5.

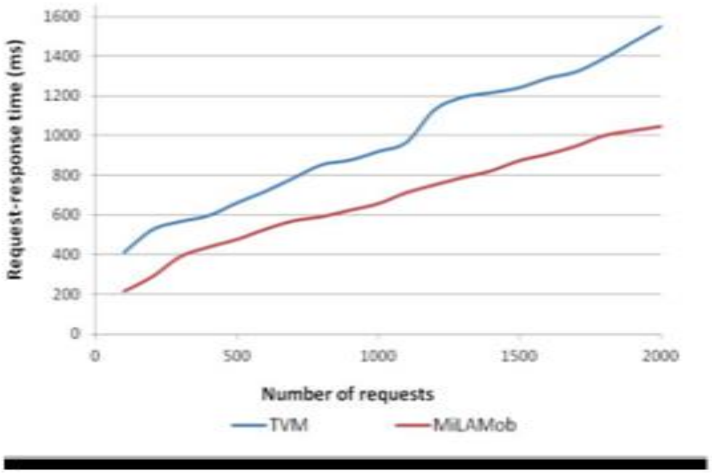

Fig. 5 Comparing the Request-Response Durations for 2000 Requests [7]

5. CONCLUSION

This paper presents the solution to the security problems occurred while accessing the public cloud storage services. Besides it also reduces the communication latency by reducing the size of the HTTPS header and thus improves the performance of the system. The use of OAuth 2.0 protocol to access data of HDFS via social sites has helped the mobile consumers to access the data via their mobiles anywhere and anytime. The multi-layer architecture employs the use of biometric tokens which helps in improving the security and privacy of accessing the data stored on cloud. The implementation of the middleware-centric MiLAMob framework with the highly fault-tolerant, quite expandable and economical HDFS system has provided such a great practical results as shown in the bar graph proving itself to be quite advantageous in performance as compared to the TVM.

REFERENCES

[1] Richard Lomotey, K., Ralph , Deters. : Middleware - Layer for Authenticating Mobile Consumers of Amazon S3 Data. IEEE International Conference on Cloud Engineering (2013)

[2] Qingni, Shen., Yahui, Yang., Zhonghai, Wu., Xin ,Yang., Lizhe, Zhang., Xi, Yu., Zhenming, Lao., Dandan, Wang., Min, Long.: SAPSC: Security Architecture of Private Storage Cloud Based on HDFS. 26th International Conference of Advanced Information Networking and Applications Workshops (2012)

[3] Qian, Quan., Wang, Tian-hong., Zhang, Rui., Xin, Ming- jun.: A Model of Cloud Data Secure Storage Based on HDFS. IEEE (2013)

[4] Richard Lomotey, K., Ralph, Deters. : SaaS Authentication Middleware for Mobile Consumers of IaaS Cloud Test Case for Amazon S3, Dropbox and MEGA. IEEE Ninth World Congress on Services (2013)

[5] Feng,Yang., Sathiamoorthy, Manoharan.: A security analysis of the OAuth protocol. IEEE (2013)

[6] Hrishikesh, Dewan., R C, Hansdah.: A Survey of Cloud Storage Facilities. IEEE World Congress on Services (2011)

[7] Singh, Surbhi., Sharma, Sangeeta.: Improving Security Mechanism to Access HDFS Data By Mobile Consumers using Middleware-Layer Framework.Computing, Communication and Networking Technologies (ICCCNT) (2014).

[8] Bhagyashree, Ambulkar., Vaishali, Borkar.: Data Mining in Cloud Computing. (2012)

[9] Zhou, Minqi., Zhang, Rong., Xie, Wei., Qian, Weining, Zhou, Aoying.: Security and Privacy in Cloud Computing: A Survey. Sixth International Conference on Semantics, Knowledge and Grids. (2010)

[10] Pinkas, Benny., Sadeghi, Ahmad-Reza., and Nigel, P.: Secure Computing in the Cloud. (2007)

[11] Himel, Dev., Tanmoy, Sen., Madhusudan, Basak , Mohammed Eunus ,Ali: An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks.Department of CSE, Bangladesh University of Engineering and Technology(2012).