Shreya Waghmare et al. 2017, Volume 5 Issue 3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Securing Cloud using Fog Computing with Hadoop Framework

[1]Shreya Waghmare, [2]Shruti Ahire, [3]Himali Fegade, [4]Pratiksha Darekar

Abstract

The need to store data is increasing day by day may be as a record or as a memory. The conventional way of storing was with the hard disks of computers or in the smartphones. With the increase in number of profiles of individuals there was a parallel increase in the data store. This lead to the insufficiency of the storage space thus leading more and more people is today getting accustomed to storing their data onto the cloud. The cloud computing has flexibility, scalability, efficiency and multi-portability. Even though efforts are been taken to secure cloud there are still loopholes in it which is restricting the users from using cloud. Thus instead of securing the data by merely using authentication credentials like user name and password, an approach of using fog computing that is concoction of user profile mapping using various behavior parameters and decoy data technology that is having a fake file of every file format which will be used to launch a disinformation attack in case the user gets detected as an intruder came into being in order to maintain confidentiality of data. This paper proposes an approach of using fog computing for securing the data with efficient algorithms and large data processing framework for accurate results.

Keywords: Fog Computing, Decoy data Technology, User Behavior Mapping, and Hadoop, Cloud Security

Introduction

Today numerous people are moving toward digitization for having a carefree life. Digitization has helped people to store their data including confidential information and media files on their devices to avoid fail to recall scenario and of course for easy access, but due to the want to store more amount of such data, devices have become incapable which conceived the idea of cloud. Cloud computing spreads it's computing services over the World Wide Web.

Data for a person is as important as one's identity and so storing it safely and securely becomes the moto. Why would any one store the data at all? The simple answer is to avoid the scenario of "fail to recall", but this does not mean a compromise on the safety of data.

All medium and large scale industries use cloud to store their data. This obviously supports better operational efficiency, but comes with risks, perhaps the most serious of which are data theft attacks. This is considered as one of the top threats to cloud computing by the Cloud Security Alliance. Yahoo claimed that hackers stole personal information from 500 million accounts, including birth dates, hashed passwords and security answers used to verify an account holder's identity. The internet company attributed the breach as a "state sponsored" attack [10]. Another data theft attack was of a scan of first lady Michelle Obama's passport been posted online [11]. This clears that the data stored on cloud is not safe from intruders. Card data of 3.2 million customers was stolen between 25 May and 10 July from a network of Yes Bank Ltd ATMs managed by Hitachi Payment Services Pvt. Ltd, but it was only in September that banks and payments services providers became aware of the extent of the breach

[12]. To overcome these threats just a string (password) of particular specifications is insufficient, we need a more sophisticated approach of handling the authenticity of the user. The proposed system uses user behavior mapping in order to identify the intrusion and on detection of intrusion decoy data is provided to the attacker.

Related Work

The authors propose a system that monitors data and provides data security from malicious intruders and also helps in confusing the attacker about the real information [1]

Much research in Cloud computing security has focused on ways of preventing unauthorized and illegitimate access to data by developing sophisticated access control and encryption mechanisms. However these mechanisms have not been able to prevent data compromise. Also fully homomorphic encryption, often acclaimed as the solution to such threats, is not a sufficient data protection mechanism when used alone [1].

Authors propose a completely different approach to securing the cloud using decoy information technology, that they have come to call as Fog computing. They use this technology to launch disinformation attacks against malicious insiders, preventing them from distinguishing the real sensitive customer data from fake worthless data [1].

The decoy documents carry a Hash Message Authentication Code (HMAC), which is hidden in the header section of the document. The HMAC is computed over the file's contents using a key unique to each user. When a decoy document is loaded into memory, they verify whether the document is a decoy document by computing a HMAC based on all the contents of that document. They compare it with HMAC embedded within the document. If the two HMACs match, the document is deemed a decoy and an alert is issued [1].

They monitor for abnormal search behaviors that exhibit deviations from the user baseline.

According to their assumption, such deviations signal a potential masquerade attack. Their previous experiments validated assumptions and demonstrated that one could reliably detect all simulated masquerade attacks using this approach with a very low false positive rate of 1.12 [1]. SVM for user profile mapping is used by [1] but SVM is a margin based classifier which may lead to misclassification if the instance is around the boundary.

The proposed plan is worked upon a local system, and thus is tested-ok to be used for cloud as mentioned by [1].

According to [3] SVM and Naïve Bayes have same measures of AUC but Naïve Bayes is simpler to implement than SVM as it calculates probability using simple mathematical operations.

MD5 is used for having hash key to the decoy file to detect intrusion [5]

Cloud computing offers many advantages such as increased utilization of hardware resources, scalability, reduced costs, and easy deployment. As a result, all the major companies including Microsoft, Google and Amazon are using cloud computing. Moreover, the number of customers moving their data to cloud services such as iCloud, Google Drive, Dropbox, Facebook and LinkedIn are increasing every day [6]

This technique confuses the intruder by providing him/her with decoy data. By the time he/she realizes it to be worthless the original data is secured more [6]

Using SVM in multi clouds environment helps to reduce data theft. In multi clouds large data sets needs to be processed, we can use BIG DATA Analytics technique called HADOOP. HADOOP is a product Apache software foundation. It is an open source framework which gained much popularity in recent times. Map and Reduce functions are the two most common techniques in HADOOP that can help in processing large data sets. These functions can be used along with SVM to avoid unauthorized data access [8].

Shifting from cloud to fog: Fog computing improves the Quality of service and also reduces latency. According to Cisco, due to its wide geographical distribution the Fog computing is well suited for real time analytics and big data [9]

Discussed a technique that confuses the insider and also used obfuscation which helps to secure data by hiding it and making it bogus information for insider [9]

Fog for Secure Cloud

The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We posit that secure Cloud services can be implemented by following security features:

User Profile Mapping

It is method of detecting the intruder's pattern by comparing with the already existing patterns. User behavior profiling deals with the behavior of the user. This technique can be used applied to check that, when and how much a client access their data in the cloud. Such normal user behavior can be continuously checked to determine whether abnormal access to a user's data is experienced. If any person get access in the cloud, then the system start detecting behavior of that person on the basis of following characteristics [2]:

1. Login Time

2. Session Time

3. Upload Count

4. Download Count

5. How many files are read and how often.

System compares all above new dataset with already present datasets which we store in the database and identify that person is authorized person or not and according to that system will send the data. Comparison of data can be done using Data mining algorithms like SVM or Naive Bayes.

Decoy Data Technology

It is the garbage data which is provided to the user if he/she is detected as an intruder Decoy files are the files which not useful for the authorized users but act as trap for unauthorized user. The illegitimate users will believe that the files accessed by them are the original files. Whenever abnormal access to a cloud service is noticed, the decoy information or files may be returned by the Cloud and delivered in such a way that the files appear to be completely legitimate and normal files as that of original files.

Encryption – SHA-1 and AES algorithms:

SHA-1 is a one-way hash function that can be used to act as a 'signature' of a sequence of bytes. It is very unlikely that 2 different byte sequences would produce the same value (though not impossible) thus useful for using at the authentication end.

AES 128-bit encryption is a data/file encryption technique that uses a 128-bit key to encrypt and decrypt data or files. It is the most secure encryption method used in most modern encryption technologies. 128-bit encryption is considered to be logically unbreakable.

HADOOP

Hadoop is a free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is a framework used for data storing and processing. Two key components of Hadoop includes HDFS and MapReduce. MapReduce is the over- all software frameworks that allows applications to implement processing of data through a cluster of nodes through distributed processing algorithms. The logic of Naive Bayes algorithm will be implemented using the MapReduce [4].

Methodology

The proposed system involves Hadoop setup at server. The user's behavior is saved in SQL server and detection of abnormal access behavior is done using data mining algorithms and map reduce. When unauthorized access is detected, decoy data is provide and the owner is been informed. In case, of owner been detected as intruder the problem is been resolved using OTP, on second request for the same file. Here the user data is encrypted using a user defined key and then stored onto the server. This encrypted data can be used as the decoy data

- **Modules**
- **Cloud Server**

The server is responsible for handling the user's profile data and user's stored data. The server converts the user password into message digest and the file data to crypt form before storing. It is also responsible for detecting intrusive sessions based on user behavior and sending decoy file to intruder.

- **Client**

The client will be having a webapp on his desktop through which access to the cloud will be done. Various modules of the webapp will be registration, login and homepage through which the client will be able to see the existing files and an option to upload the file by encrypting it with a user defined key.

- **Admin**

Admin is responsible for adding decoy files and for loading the training dataset.

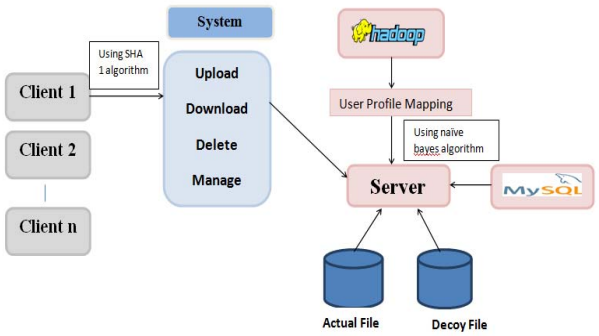


Fig 1: Implementation architecture for the system

The work flow for account registration and authentication in a hash-based account system is as follows:

1. The user creates an account.

2. Their password is hashed and stored in the database. At no point is the plain-text (unencrypted) password ever written to the hard drive.

3. When the user attempts to login, the hash of the password they entered is checked against the hash of their real password (retrieved from the database).

4. If the hashes match, the user is granted access. If not, the user is told they entered invalid login credentials.

5. Steps 3 and 4 repeat every time someone tries to login to their account

The workflow of user behavior mapping:

1. Parameters like duration, upload-rate, download-rate, slot, blacklisted count will be considered to predict requester as valid or invalid.

2. 1st the initial probability will be found in the training phase where system will find total count of yes i.e. valid out of total records and no i.e. invalid out of total records.

3. After this, conditional probability will be found i.e. for yes and no for every value of every parameter. for example say, duration has 3 values low, medium, high then, duration=low, yes=3/7 no=4/7 this means records having duration as low are 3 and out of those 3 are detected as valid and 4 are detected as invalid.

4. After step 3 all the yes values are combined and all no values are combined.

5. Finally, conditional (yes) * initial (yes) = x

conditional (no) *initial (no) = y

If x is greater than y then the user is detected as valid and if x is less than y then user is detected as invalid.

The workflow for file encryption is as follows:

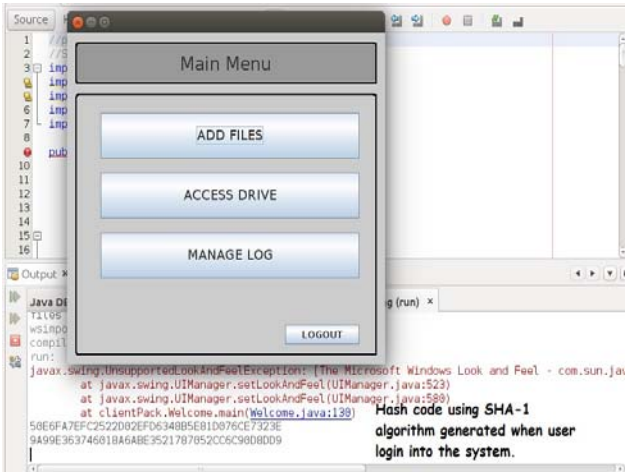
1. When client uploads the file, file is divided into chunks and is send to the server.
2. On Arrival at the server file is encrypted with static key using AES (Advanced Encryption Standard).
3. After encryption file is stored on MySql database server.
4. On request for the file, the file is fetch from the database, decrypted using the static key and send to the user.

Applications

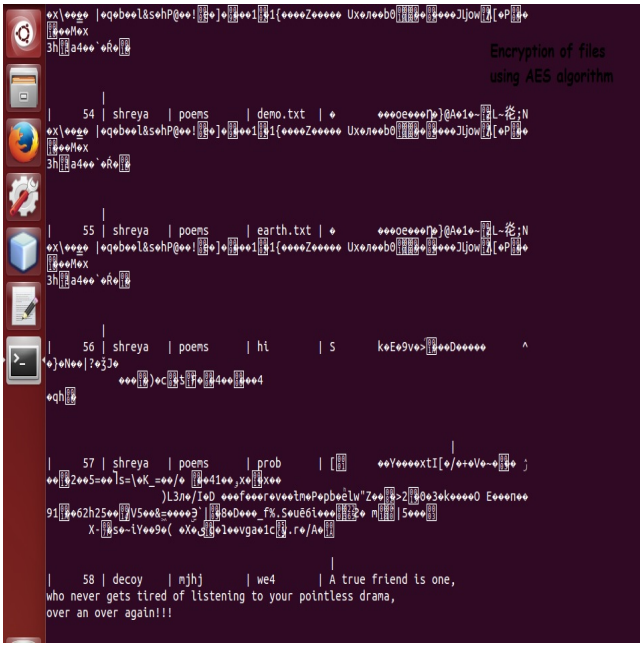
1. Securing the file or data stored by normal users on cloud from data theft just like DIGI-LOCKER.
2. Useful in government sector or organization to secure citizen or employee database from intruders intended to duplicates someone identity.
3. In Software Companies to secure software projects which are deployed on the cloud.
4. Banks can used to secure their account-holders details and transaction details.

Results:

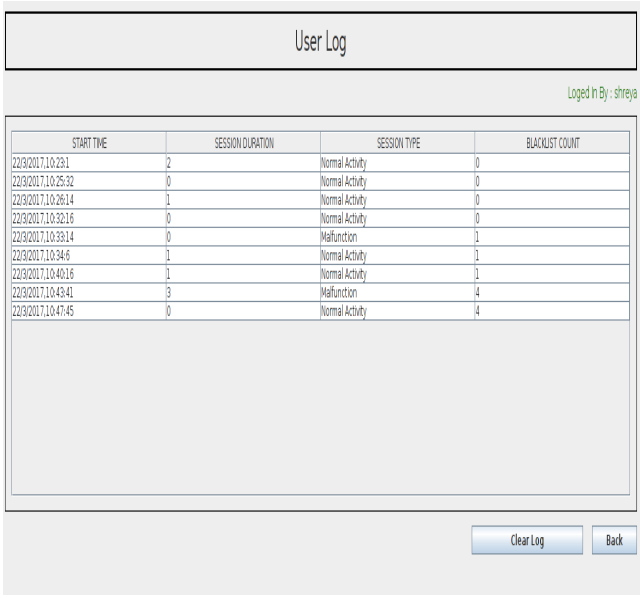
1. Generation of Hash code using SHA-1



2. Result of Encryption using AES



3. Result of Naïve Bayes



Conclusion

Thus we are able to secure the data on the cloud in our proposed system by using algorithms like SHA1 and Naive Bayes having high accuracy and precision rate spread across 2-level security structure compared to the one's proposed and used at same level thus detecting intruders more accurately. The system will be using Hadoop for User profile mapping and also for reducing the detection time.

Acknowledgements

We take this opportunity to thank our project guide Prof. Suruchi Nannaware and Head of the Department Prof. S. V. Todkari for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of JSPM's Jayawantrao Sawant College of Engineering, Hadapsar, Pune-28 for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the Institute for providing the required facilities, Internet access and important books.

References

[1] Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud by Malek Ben Salem, Salvatore J. Stolfo , EEE Symposium on Security and Privacy Workshops 2012

[2] Fog Computing: preventing Insider Data Theft Attacks in Cloud Using User Behavior Profiling and Decoy Information Technology

[3] Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy, Jin Huang, Jingjing Lu, Charles X. Ling

[4] Big Data and Hadoop,

[5] FogComputi¬¬¬¬ng: Securing the cloud and preventing insider attacks in the cloud. Aatish B. Shah1, Jai Kannan2, Deep Utkal Shah3 Prof. S.B.Ware4, Prof. R.S.Badodekar5 2016.

[6] Younghee Park, Salvatore J. Stolfo, Software Decoys for Insider Threat, ACM 2012.

[7] http://www.sha1-online.com/

[8] Secure Data Access control in Cloud Environment, 1 G. Praveen Babu, 2 B. Sushma Rao , / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1734-1737

[9] Fog Computing Providing Data Security: A Review, Manreet Kaur, Monika Bharati, International Journal of Advanced Research in Computer Science and Software Engineering

[10] http://www.telegraph.co.uk/business/2016/09/22/half-a-billion-yahoo-users-data-stolen-in-state-sponsored-hack/

[11]http://www.telegraph.co.uk/news/2016/09/22/mich-lle-obamas-passport-scan-posted-online-in-apparent-hack/

[12]http://www.livemint.com/Industry/Ope7B0jpjoLkemwz6QXirN/SBI-Yes-Bank-MasterCard-deny-data-breach-of-own-systems.html

Author's details

[1,2,3,4]UG Student, Dept. of Information Technology, JSPM's JSCOE, Pune, Maharashtra, India,

Email: [1]waghmare2shreya@gmail.com, [2]13ahireshruti@gmail.com, [3]himalifegade16@gmail.com, [4]pratikshadarekar1@gmail.com