International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.1, 2015

RESEARCH ARTICLE

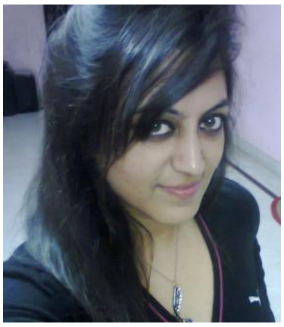
ISSN: 2321-7758

ENHANCEMENT IN CLOUD DATA SECURITY USING FOG COMPUTING

ASHADEEP[1], SACHIN MAJITHIA[2]

[1]STUDENT, [2]ASSISTANT PROFESSOR
[12]Chandigarh Engineering College, Landran, 140307

Article Received: 31/12/2014 Article Revised on: 05/01/2015 Article Accepted on:07/01/2015



ASHADEEP

ABSTRACT

Cloud Computing enables multiple users to share common computing resources, and to access and store their personal and business information A major amount of professional and personal data is stored on cloud.Cloud storage is being used enormously in various industrial sectors. In spite of the abundant advantages of storing data on cloud, Security still remains a major hurdle which needs to be conquered. Data on Cloud is being accessed with the new communication and computing paradigms which further arise new data security challenges. The subsisting methods of protecting data on cloud have failed in preventing data theft attacks. An altered approach is used known as fog computing which uses the two techniques viz. User Behaviour Profiling and Decoy Technology. In this paper, we propose a technique to solve the above mentioned problem we propose a new technique i.e. Cusum algorithm which detects changes in user access patterns by calculation of average fluctuation in the user behavior and in order to secure the real user information, we have implemented an enhanced HMAC technique by addition of pseudo-random generator in it This technique efficiently enhances the accuracy to detect the insider data theft attacks and securing useful information.

Keywords: Decoy information, cloud computing, fog computing, insider theft attacks.

INTRODUCTION

The Computing in which the resources like data, storage, various softwares are allotted over the network and are managed through the Internet by a service provider (one who provides cloud resources like software and storage space, etc.) is termed as cloud computing. It is also popularly called an Internet based computing because the users interact with the service provides through the Internet and also the customers are given the services via Internet Cloud computing is achieving popularity and gaining attention in business organizations. It offers a variety of services to the users. It is a widespread computing field which is easy to use, service is provided according to user need or request. Due this ease, software companies and other agencies are shifting more towards cloud computing environment. To achieve better operational efficiency in many organizations and small or medium agencies is using Cloud environment for managing their data. Cloud Computing is a combination of a number of computing strategies and concepts such as Service Oriented Architecture (SOA), virtualization and other which rely on the Internet.

The pictorial representation of cloud computing is shown below [1]:

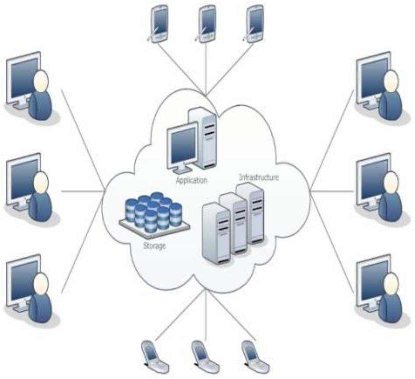


Figure 1: Cloud Computing Architecture

DEPLOYMENT MODELS:

Clouds are categorized into four deployment models based on their accessibility, organizational structure and the provisioning location. They are:

a) Public Cloud
b) Private Cloud
c) Community Cloud
d) Hybrid Cloud

Public Cloud: A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Public cloud services may be free on a pay per-per-usage model.

Private Cloud: Private cloud is infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally. Private cloud computing systems, use of the concept of visualization and emphasis on consolidating distributed IT services often within data centres enters belonging to the organization or enterprise. A private Cloud's usage is restricted to members, employees, and trusted partners of the organization.

Community Cloud: Community cloud shares infrastructure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally.

Hybrid cloud: Hybrid cloud is a composition of two or more clouds (private, community, or public) that remain unique entities, but are bound together, offering the benefits of multiple deployment models. Hybrid Cloud enables the use of private and public Cloud in a seamless manner. It can also be defined as multiple cloud systems that are connected in a way that allows programs and data to be moved easily from one deployment system to another.

1. ATTACKS ON CLOUD COMPUTING

While shifting from traditional computing to cloud computing many new security and privacy concerns has aroused. On the basis of attacks on cloud computing security can be divided into two categories one is the physical security and other is cyber security. Physical security means when there are security issues due to hardware or software failure. Security breached due to some natural calamity such as earthquake, flood, tsunami, etc. Cyber security concerns are attacked through the internet or with the use of devices like computer. Due to huge number of data on cloud the risk of such attacks increase. Consumers face problems of availability of resources as the attackers utilize these resources for their own criminal purposes. The following are the types of attacks which come under the category of cyber security [2], [3]:

- **Insider attacks:** The employees who can or were able to access the entire information about the organization are termed as insiders. Insider attacks are knowledge about consumers or providers and include every kind of attacks which can be executed from inside.
- **Flooding attacks:** These attacks include packets containing huge amount of information and are sent from an exploited resource. These attacks are mainly done through the Internet as cloud computing is also sharing of resources through internet. Moreover the connections used for such attacks are unauthorized network connections. These attacks block services to be used by authorized users by accessing a particular service for a long period. Such attacks are known as DoS attacks which are Denial of service attacks.
- **Backdoor channel attacks**: These attacks are linked to DoS attacks, the intruder or the attackers use a node of the cloud through which

he can inject zombies. Once the node is compromised the attacker can use it as a path for introducing attacks like DoS and DDoS. Viruses, Trojans, malware, etc. can be easily injected to the confidential data through the compromised node.

- **Data Modification:** After an attacker has read the data, the next logical step is to alter it. An attacker can modify the data in the packet without the knowledge of the sender or receiver. Even if no confidentiality for all communications is required, a user does not want any messages to be modified in transit.
- **Eavesdropping :** In general, the majority of network communications occur in an unsecured or "clear text" format, which allows an attacker who has gained access to data paths in network to "listen in" or interpret (read) the traffic. When an attacker is eavesdropping on communications, it is referred to as sniffing or snooping. The ability of an eavesdropper to monitor the network is generally the biggest security problem that administrators face in an enterprise. Without strong encryption services that are based on cryptography, data can be read by others as it traverses the network.
- **Identity Spoofing (IP Address Spoofing)** Most networks and operating systems use the IP address of a computer to identify a valid entity. In certain cases, it is possible for an IP address to be falsely assumed— identity spoofing. An attacker might also use special programs to construct IP packets that appear to originate from valid addresses inside the corporate intranet. After gaining access to the network with a valid IP address, the attacker can modify, reroute, or delete data.

Ways of performing insider theft attacks: Insiders have an advantage than external attackers that is they are familiar with the network architecture of the organization. Therefore, attacks can be performed through network. Organizations focus on mechanisms which can provide protection from external attacks therefore the chances of detecting insider theft decrease.

2. SECURING CLOUD WITH FOG COMPUTING

Fog Computing is an extension of Cloud Computing. As in a Cloud, Fog computing also provides data, compute, storage, and application services to end-users. The difference is Fog provides proximity to its end users through dense geographical distribution and it also supports mobility. Access points or set-up boxes are used as end devices to host services at the network. These end devices are also termed as edge network. 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. .Fog computing provides-Low latency and location awareness, it has Wide-spread geographical distribution, supports Mobility [3]. The main task of fog is to deliver data and place it closer to the user who is positioned at a location which at the edge of the network. Many methods are proposed to secure cloud data by encryption and standard access control but it is found that the methods are not full proof due to variety of reasons. Customer not only requires reliable cloud environment but also a healthy security for data and applications. Recovering the stolen or lost data is not possible. So we must have knowledge to deal with such incidences. If we decrease the value of stolen data by providing decoy documents then we can limit the harm of the system. The architecture of fog computing is shown in Fig. 2 below:

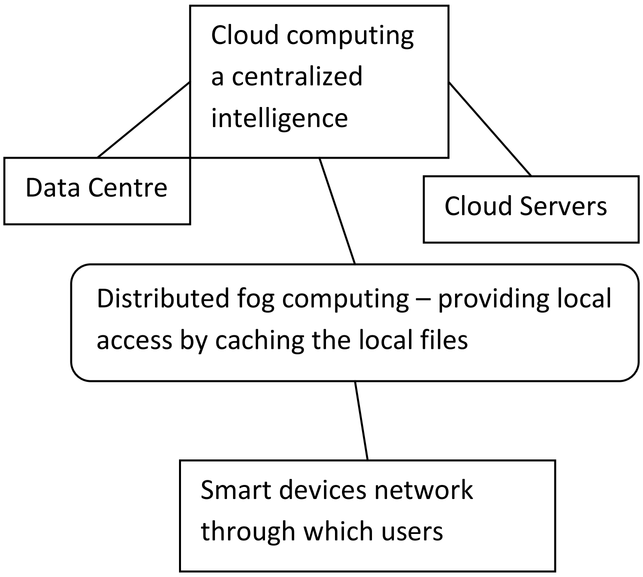

Cloud computing a centralized intelligence
Data Centre
Cloud Servers
Distributed fog computing – providing local access by caching the local files
Smart devices network through which users

Figure 2: Fog Computing Architecture

Salvatore J. Stolfo and Malek Ben Salem propose extra security features are as follows [4]:

- User Behaviour Profile
- Decoys

User Behaviour Profile: Search for specific files is likely to be targeted and limited the reason being that valid users of a computer system are familiar with the files on that system and where they are located. Search by a masquerade is liable to be extensive and untargeted because of his unfamiliarity with the structure and contents of the file system. Based on this key assumption, user search behavior is profiled and user models are developed trained with a one class modeling technique, namely one-class support vector machines. In a one-class modeling technique a classifier can be built without having to share data from different users. The data and privacy of the user is thus preserved. Abnormal search behaviors that exhibit deviation from the user baseline are monitored. A potential subterfuge attack is signaled by such detection

Decoy Technology: Traps are placed within the file system. The traps are decoy files downloaded from a Fog computing site, an automated service that offers several types of decoy documents such as tax return forms, medical records, credit card statements, e-bay receipts, etc. The decoy files are downloaded by the legitimate user and placed in highly-conspicuous locations that are not likely to cause any interference with the normal user activities on the system. A masquerade, which is not familiar with the file system and its contents, is likely to access these decoy files, if he or she is in search for sensitive information, such as the bait information embedded in these decoy files. Therefore, monitoring access to the decoy files should signal masquerade activity on the system.

3. LITERATURE REVIEW

Salvatore J. Stoflio et al. proposed a new technique and named it as Fog computing. They implemented security by using decoy information technology. They mentioned two methods, User behavior profiling and Decoy. In User behavior profiling they checked which information does a user usually checks. They monitored their user's activity to check for any change in the usual data access behavior of the user. Another technology is decoy in which bogus information such as honey pots, fake documents is provided to the attacker to protect the real one.

Govinda et al. discussed that leakage of sensitive data from the service provider is an alarming situation. In Cloud Computing resources are offered as a service which leverages virtualization and other Internet technologies. Further, they proposed an agent based model that would secure the users' data over the cloud and they also implemented various algorithms to secure cloud.

Sabahi, F. (2011) mentioned threats and response of cloud computing. They presented a comparison of the benefits and risks of compro mised security and privacy.

Mowbray M. et al.(2009) described a privacy manager for cloud computing, which reduces the risk to the cloud computing user of their private data being stolen or misused, and also assists the cloud computing provider to conform to privacy law. Also give an algebraic description of obfuscation, one of the features of the privacy manager; and describe how the privacy manager might be used to protect private metadata of online photos.

Park, Y. Et al. (2012) developed a technique that was a software decoy for securing cloud data using software. They proposed a software-based decoy system that aims to deceive insiders, to detect the exfiltration of proprietary source code. The system builds a Java code which appears as valuable information to the attacker. Further obfuscation technique is used to generate and transform original software. This deception technique confuses the insider and also obfuscation helps the secure data by hiding it and making bogus information for insider. Beacons are also injected into the bogus software to detect the exfiltration and to make an alert if the decoy software is touched, compiled or executed

Godoy et al explained that there is a need of such profiling strategies or methods through which user profiling can be done. As there is a huge amount of information available on the web or Internet therefore from last few years personal information agents are helping the users to manage their information. In this paper the authors have discussed a learning technique for data acquisition for user profiling and so they mentioned some methods for adaption with the changes which happen time to time with the change in user's interest. They said earlier only supervised learning technique was used in general. But for moving the information agents to the next level authors are focusing on assessment of semantically useful user profiles. They said that account hijacking is a disadvantage for such user profiling.

From this literature survey, we find that with the help of two techniques of fog computing named as decoy technology and user behavior profiling, we can prevent the data theft attacks in the cloud in a better way in comparison to the traditional encryption techniques. In the literature survey, the implementation of decoy technology involves HMAC code calculation using a message digest algorithm, but in this paper, we have implemented an enhanced technique of HMAC code calculation by adding a pseudo-code generator that further assures the security of the real user information. Further, we have implemented Cusum algorithm that calculates the average fluctuation in the access pattern of the user which adds to the efficient detection of an attacker and thus the accuracy of the detection of insider data theft attacks is enhanced efficiently.

3. PROPOSED TECHNIQUE

Detecting abnormal search and decoy traps together may make a very effective masquerade detection system. Combining the two techniques improves detection accuracy False positive rate of detector is lowered by combining the two techniques, and having the decoy documents act as an oracle for the detector on detection of abnormal user behavior.

User Behaviour Profiling: Legitimate Users of the Cloud system are acquainted with the documents and information on the Cloud system they have stored. The search for documents is to the point and limited. A masquerade gets access to the victim's system illegitimately, is unlikely to be acquainted with the structure and contents of the file system. Their search is not to the point and widespread. The user search behaviour is profiled and developed based on this key assumption, Cusum algorithm is used which calculates the average fluctuation and thus the user behaviour is noted when changed.

Decoy Technology: Decoy means the bogus information about the related data documents. If any suspicious activity is sensed then to mislead the attacker, fake information is released after the user search modeling. In order to make sure that the attacker fails to differentiate between the decoy files and the actual files the same database is used for both decoy as well as original file. There is direct linking to fog computing sites in case the attack on user's data is continued by the attacker. Through this the safety of the important data is increased. The actual user will now identify if the bogus data is being sent by the cloud as he is the owner of the data [5]. Thus, through a large number of means the response by the cloud can be altered, such as challenge questions to inform the cloud security system about its unauthorized and incorrect access. The decoy documents carry a keyed-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, we verify whether the document is a decoy document by computing a HMAC based on all the contents of that document. It is compared with HMAC embedded within the document. If the two HMACs match, the document is deemed a decoy and an alert is issued.

HMAC code:

HMAC is keyed hashed message authentication code which is used for calculating a message authentication code. It involves a cryptographic hash function along with a secret key. We are calculating the HMAC code of the document by using the MD5 Algorithm.MD5 processes a document of variable length into a fixed length output of 128 bits [8].

- Variable length to fixed length output.
- Input n-bit blocks
- Input divided into 512 bit blocks
- Padding is done
- Buffer initialization
- Output 128 bit

In our decoy technique, we have enhanced the security of the information by inserting a pseudo-code generator which further jumbles the code generated by the hashing or cipher generator algorithm. It uses a key to generate a jumbling mask. A mask is a pseudo-code which is passed to multiple multiplexers and it selects one from each multiplexer ensuring that no same block is selected by two or more multiplexers at any time.Once, it has selected the blocks, the input block pattern is jumbled depending on the key and hence the attacker can never they get the original information if ever he tries.

The advantages of placing decoys in a file system are threefold:

- The detection of masquerade activity.
- The confusion of the attacker and the additional costs incurred to identify the real information from bogus information.
- The combination of the two techniques: The combination of user behaviour profiling with decoy technology provides a strong evidence of illegal access and helps improve accuracy of detection. Only user one technique can produce false positive results.

By combining the two techniques the rate of detecting illegal access increases.

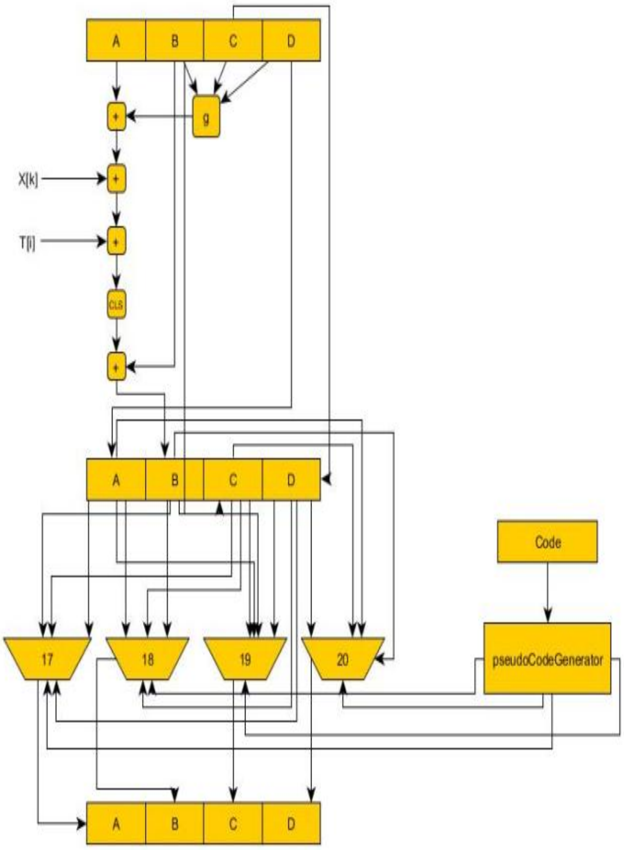


Figure 3: HMAC Technique

4. STEPS OF EXECUTION OF PROPOSED WORK

STEP 1: A file system in the cloud is made and some trap files are placed in it.

STEP 2: User files are kept secure by an authentication mechanism asked while accessing and editing. If someone accesses the trap file then with the help of HMAC technique, we get to know the access and even if the person gets the access to the useful documents, he won't be able to get the correct information because of the coded form of the information.

STEP 3: User behaviour is identified and monitored using CUSUM Algorithm which is explained below:

```
For applying cusum on N no of observations
Let, initial average av1 -> N =0;
Sump=Sum till previous observations =0;
For loop n=1 -> N
sump=sump (previous)+Current(n)
av(n)=sump/N
end for loop
```

Now **av** is the cumulative summation averages and difference in two **consecutive** averages gives the **fluctuation.**

STEP 4: If the average fluctuation varies more than 12 and too more than 10 counts, then the user is classified as an attacker else a legit imitate user.

STEP 5: After the user classification, two lists are prepared i.e. white and black list according to which the users are allowed and disallowed. According to the new incoming users, the list is updated.

STEP 6: If a user fulfils all the above conditions of an attacker, then, ask for email id and password as the last authentication step.

STEP 7: If the user enters the credentials correctly, then the attacker is the real one else the account of the user is temporarily locked. We take 1400 conditions in which 700 are attacker conditions and 700 non-attacker conditions. Total of 18 rounds are conducted to test the accuracy of our algorithm. The Flow chart of the execution is shown in Fig.5 below:

ANALYSIS AND RESULTS

In this section we evaluate the effectiveness of proposed technique. We have generated the graphs of our proposed method by using three parameters average fluctuation, load and time, and the final graph is of accuracy in terms of percentage.

Accuracy is being calculated by the formula given below:

$$\frac{\text{TRUE POSITIVES}}{\text{TOTAL NO. OF CASES}} \times 100 = \text{ACCURACY} \qquad (1)$$

Where the total number of true positives is the number of times our algorithm gave correct results i.e. recognised the attacker and non-attacker conditions correctly, the total number of cases are the total conditions to be tested i.e. 1400

Accuracy Graph

The current accuracy results are depicted in the table below:

FLOW CHART

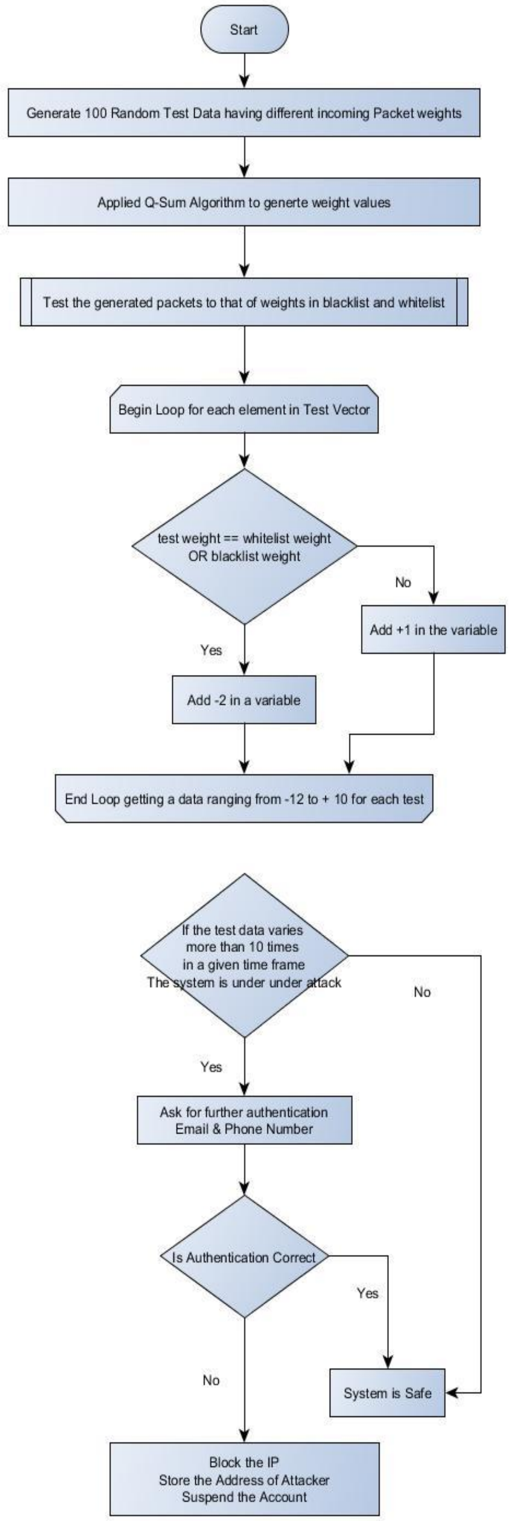

Start
Generate 100 Random Test Data having different incoming Packet weights
Applied Q-Sum Algorithm to generte weight values
Test the generated packets to that of weights in blacklist and whitelist
Begin Loop for each element in Test Vector
test weight == whitelist weight OR blacklist weight
No
Add +1 in the variable
Yes
Add -2 in a variable
End Loop getting a data ranging from -12 to + 10 for each test
If the test data varies more than 10 times in a given time frame The system is under under attack
No
Yes
Ask for further authentication Email & Phone Number
Is Authentication Correct
Yes
No
System is Safe
Block the IP
Store the Address of Attacker
Suspend the Account

Figure 5: Flow Chart of Execution

TABLE I: SIMULATION TABLE

USER NO.	TRUE POSITI-VES	ACCURACY
1	1344	(1344/1400)*100=96%
2	1316	(1316/1400)*100 94%
3	1386	(1386/1400)*100=99%
4	1232	(1232/1400)*100=88%
5	1162	(1162/1400)*100 83%
6	1190	(1190/1400)*100=85%
7	1232	(1232/1400)*100 =88%
8	1162	(1162/1400)*100= 83%
9	1204	(1204/1400)*100= 86%
10	1246	(1246/1400)*100 =79%
11	1246	(1246/1400)*100 =79%
12	1092	(1092/1400)*100 =78%
13	1260	(1260/1400)*100 =90%
14	1134	(1134/1400)*100 =92%
15	1162	(1162/1400)*100 =90%
16	1134	(1134/1400)*100= 81%
17	1162	(1162/1400)*100 =83%
18	1134	(1134/1400)*100 =81%

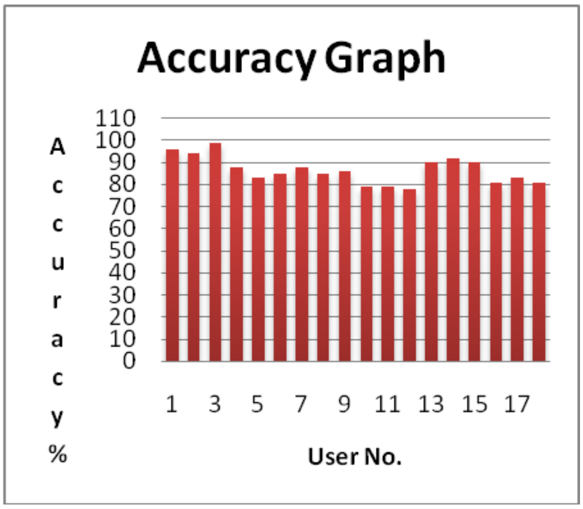


Figure 6: Accuracy Graph

Comparison graph of current and previous results

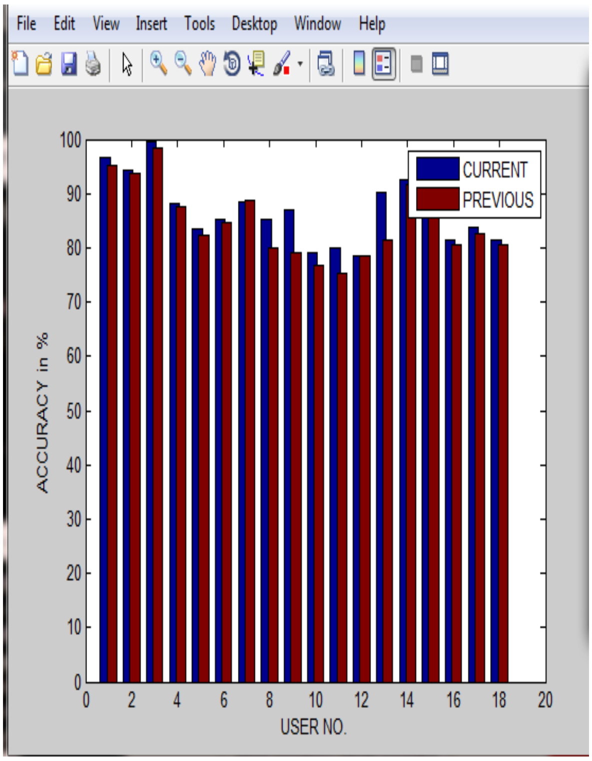


Figure 7: Comparison Graph

The table comparing the accuracy values of previous and current results corresponding to the Fig.7 is shown in the below table:

TABLE II: COMPARISON TABLE

Present Value	Previous Value
96	95.2
94	93.6
99	98.3
88	87.5
83	82.3
85	84.7
88	88.6
85	79.9
86	79.1
79	76.9
79	75.3
78	78.5
90	81.5
81	80.7
83	82.5
81	80.5

CONCLUSIONS AND FUTURE WORK

In this paper, we have presented the fog computing architecture and discussed the two techniques i.e. decoy technology and user behaviour profiling. With the help of these two techniques, we can efficiently prevent the data theft attacks in the cloud. We have implemented a CUSUM change point detection algorithm for detecting the abnormalities in user behaviour profile. Using CUSUM, average fluctuation in user profile or access behaviour is evaluated.And we have implemented an enhanced HMAC code calculation technique with insertion of pseudo-code generator in MD5; this enhances the security of real information of the user. On the basis of this technique the accuracy of the system is more enhanced.In future we can extend the working of algorithm, by calculating the accuracy with other attributes such as performance evaluation of the security mechanism.

REFERENCES

[1]. Marten van Dijk, Ari Juels, “On the Impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing” RSA Security Brief, March 2010

[2]. http ://cnc.ucr.edu/security/glossary.

[3]. http://technet.microsoft.com/en-us/library/cc959354.aspx

[4]. Salvatore J. Stolfo, Malek Ben Salem and Angelos D. Keromytis, “Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud”,IEEE 2012

[5]. Ben-Salem M., and Stolfo, “Decoy Document Deployment for Effective Masquerade Attack Detection,” Computer Science Department, Columbia University, New York

[6]. Cisco Cloud Computing -Data Center Strategy, Architecture,and Solutions http://www.cisco.com/web/strategy/docs/gov/CiscoCloudComputing_WP.pdf

[7]. Jay Heiser, Mark Nicolett, Assessing the Security Risks of Cloud Computing, 03 June, 2008

[8]. Sayali Raje, Namarta Patil, Shital Mundhe and Ritika Mahajan, “Cloud Security Using Fog Computing” Proceedings of IRF International Conference, 30th March-2014, Pune,

[9]. John Harauz,,Lori M. Kaufman, Bruce Potte, Data Security in the world of Cloud Computing

[10]. Ki-Woong Park, Sung Kyu Park, Jaesun Han, Kyu Ho Park, THEMIS: “Towards Mutually Verifiable Billing Transactions in the Cloud Computing Environment,” 2010 IEEE 3rd

International Conference on Cloud Computing (CLOUD), Page(s) 139-147, July 2010

[11]. J. Montelibano, A. Moore, Insider Threat Security Reference Architecture, 2012 45th Hawaii International Conference on System Science (HICSS), Page(s) 2412 - 2421, 4-7 January 2012

[12]. Nahla Shatnawi, Q.A., Wail Mardini (2011). "Detection of Insiders Misuse in database Systems" proceedings of the international Multi Conference of Engineers and computer Science 2011, Hong Kong Vol. I, IMECS 2011, March 16 - 18, 2011.

[13]. D. Godoy, A. Amandi, "User Profiling for Web Page Filtering," IEEE Internet Computing, vol. 9, no. 4, pp. 56-64, July – Aug.2005

[14]. S. Mathew, S. Upadhyaya, D. Ha, H.Q. Ngo, “Insider abuse comprehension through capability acquisition graphs”, IEEE, 2008 11th International Conference on Information Fusion, pp. 1-8, June – July 2008

[15]. J. Montelibano, A. Moore, “Insider Threat Security Reference Architecture”, 2012 45th Hawaii International Conference on System Science (HICSS), Page(s) 2412 - 2421, 4-7 January 2012

[16]. G. Briscoe, A. Marinos, “Digital ecosystems in the clouds: Towards community cloud computing”, 3rd IEEE International Conference on Digital Ecosystems and Technologies,” Istanbul, DEST '09, pp. 103-108, June 2009

[17]. Bonomi F., Milito R., Zhu J. &Addepalli S. , “Fog Computing and its role in the Internet Of Things”,IEEE 2012