iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** XII **Month of publication:** December 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

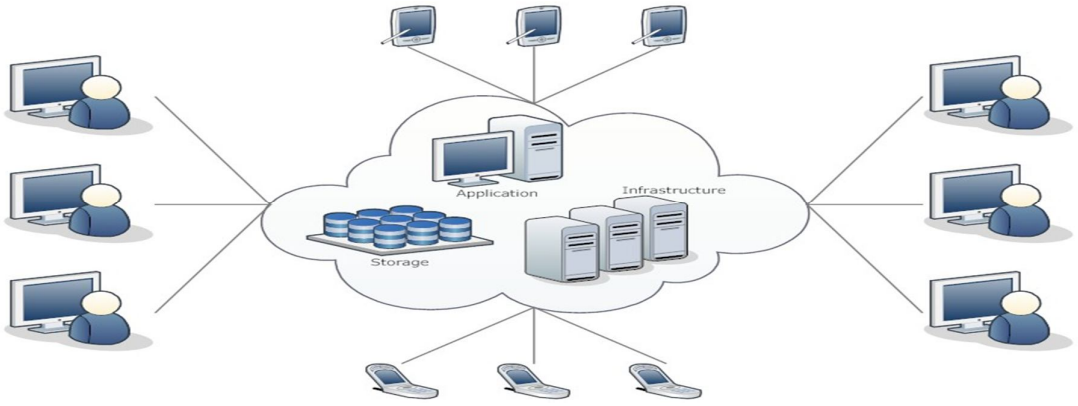
Hindering Data Theft Attacks in the Cloud Using Fog Computing

Ashadeep[#1], Sachin Majithia[*2]
[#*]Information Technology, CEC Landran, Punjab Technical University

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. 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 arises 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. Inspite of the these techniques, there occurs a problem of the wrong information appearing to the correct user and also the misidentification of user as an attacker then the situation goes up side down and user faces inconvenience and dissatisfaction. In this paper, we propose a technique to solve the mentioned problem i.e. implementation of Cusum algorithm which detects changes in user access patterns and calculates average fluctuation efficiently and thus enhances the accuracy to detect the insider data theft attacks.***
Keywords— Cloud computing, decoy information, fog computing, insider data thefts.

I. 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:



A. Deployment Models:

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

Public Cloud
Private Cloud
Community Cloud
Hybrid Cloud

1) *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.
2) *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.
3) *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.
4) *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

II. LITERATURE SURVEY

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 behaviour profiling and Decoy. In User behaviour 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 behaviour 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 compromised 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.

III.PROBLEM FORMULATION

Due to limitation of communication media and data privacy due to third party usage and ownership of data, it is a challenging task to maintain the confidentiality, integrity and authorizations for data.With cloud computing there come ease of access and a

convenience of way beyond the normal computing. It will significantly change the way we use computers. With the increase in convenience, there comes problems of the new level at which security is a prime concern. Your data in the cloud means your data at server, which can be an easy target if not secured properly. Proper breach detection and threat removal method must be there to prevent data from going into dark hands. Traditional methods have failed in providing up to the mark security to the clients, primarily those wreaked by an insider to the service provider (cloud).

A new technology of security had arrived called the fog computing this security mechanism uses a very simple method for threat detection and data misuse, but there have been some flaws which they were not able to count up-to. This method comprises a unique model for data protection in the cloud using offensive decoy technology for flooding the intruder with false data. But in some case this might become problematic to the user appearing as wrong information to the correct user and sluggish response because of so many unnecessary to the user security files also if the user reaches the trap then the situation goes upside down and user faces inconvenience and dissatisfaction.

IV. CLOUD COMPUTING TO 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. 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.

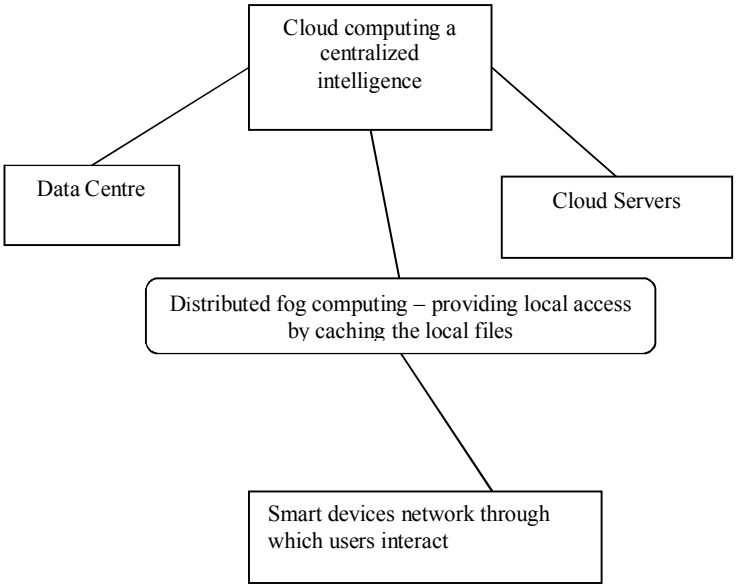


Fig. 1 Architecture of fog computing

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. Salvatore J. Stolfo and Malek Ben Salem propose extra security features are as follows [5]:

A. User Behaviour Profile

*S*earch 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 behaviour is profiled and user models are developed trained with a one class modelling technique, namely one-class support

vector machines. In a one-class modelling 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 behaviours that exhibit deviation from the user baseline are monitored. A potential subterfuge attack is signalled by such detection

B. 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.

V. 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 behaviour.

A. 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.

B. Decoy Technology

The file system is packed together with traps, these traps are uploaded on the system by the Cloud service provider. These traps can contain documents like credit card details, tax returns, bank statements. These documents are places in highly egregious places. A masquerader who is not acquainted with the system and who has an ill intent may is likely to click on these false documents. Thereby the system can be notified of masquerade activity.

The hash code of all the legitimate and decoy documents upload on the system is calculated. The hash code of every document downloaded is matched with the hash code of the decoy document. If a match is found then the document is deemed to be a decoy document and an alert is generated. An insider attacker would not be able to escape detection if they access a decoy document. The hash code is based on keyed-Hash Message Authentication Code (HMAC).

1) *HMAC code:* HMAC that 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.

a) Variable length to fixed length output.
b) Input n-bit blocks
c) Input divided into 512 bit blocks
d) Padding is done
e) Buffer initialization
f) Output 128 bit

In our decoy technique, we have enhanced the security of the by inserting a pseudo-generator which further jumbles the code and changes the position of each element in the code so that the attacker can never the get the original information if ever he tries.

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

1) The detection of masquerade activity
2) The confusion of the attacker and the additional costs incurred to identify the real information from bogus information.

3) 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.

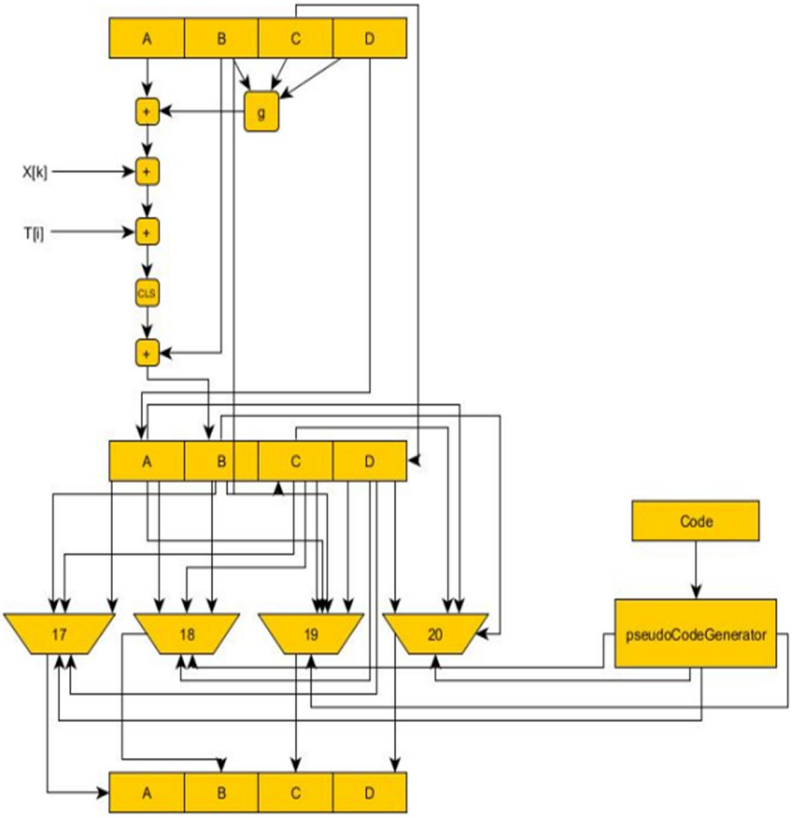


Fig.2 HMAC CODE WITH MD5

VI. CUSUM ALGORITHM

For applying cusum on **N** no of observations

Let, initial average av**1 -> N =0;**

Sump=Sum till previous observations =0;

For loop n=1 -> N

sump=sump (previous)+Current(n)

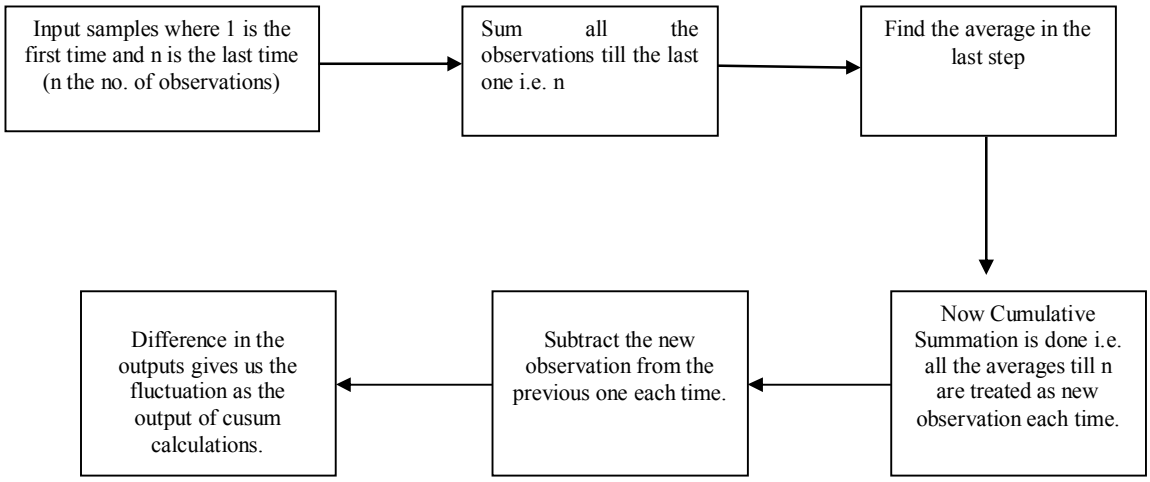
av(n)=sump/N

end for loop

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

TABLE I

CUSUM FLOW TABLE



VII. METHODOLOGY

A. Implementation Setup

This section describes the implementation environment and used system components. The implementation of CUSUM algorithm is done in MATLAB.

MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, we can analyse data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable us to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. MATLAB can be used for a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology.

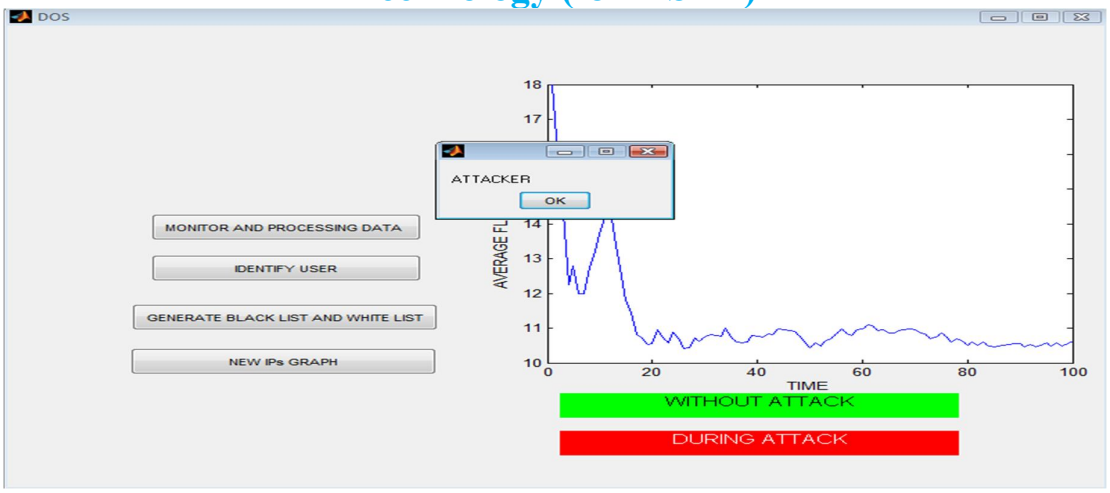
We chose MATLAB to implement our algorithm because of the following reasons:

1) High-level language for numerical computation, visualization, and application development
2) Interactive environment for iterative exploration, design, and problem solving
3) Built-in graphics for visualizing data and tools for creating custom plots
4) Development tools for improving code quality and maintainability and maximizing performance
5) Tools for building applications with custom graphical interfaces
6) Functions for integrating MATLAB based algorithms with external applications and languages such as C, Java,.NET, and Microsoft® Excel®

Various toolboxes used in MATLAB are as follows:

B. Communications System Toolbox

1) Data Acquisition Toolbox
2) Database Toolbox
3) Simulink

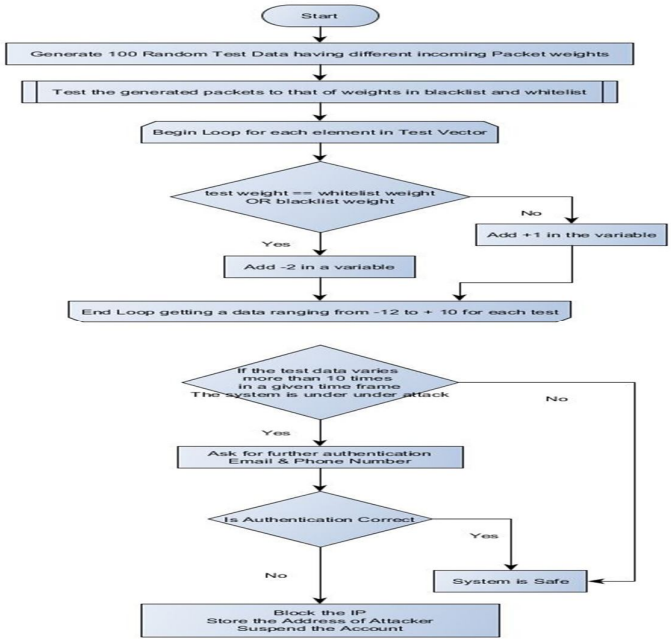


B. Methodology Used

This section will discuss a methodology and its related parameters, experiment factors.

1) *System Parameters*- The experiments are conducted using 1 GHz processor with RAM 256MB or higher and hard disk 10 GB or higher.
2) *Experiment Factors*-In order to evaluate the performance of CUSUM algorithm on the basis of time, average fluctuation and true positive.

C. Working Steps



VIII. SIMULATION RESULTS

This Section will show the result obtained from the simulated environment for CUSUM algorithm. Results of the simulation have been shown below in the form of graphs .Accuracy of the algorithm has been calculated in terms of percentage by dividing the total true positives i.e. number of times an algorithm detects the correct conditions by the total number of test cases.

$$\text{Accuracy} = \left(\frac{\text{True Positives}}{\text{Total Cases Taken}}\right) \times 100$$

TABLE 2
ACCURACY GRAPH

USER NO.	TRUE POSITIVES	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	1288	(1288/1400) * 100 = 92%
15	1260	(1260/1400)* 100 = 90%
16	1134	(1134/1400)* 100 = 81%
17	1162	(1162/1400)* 100 = 83%
18	1134	(1134/1400) * 100 =81%

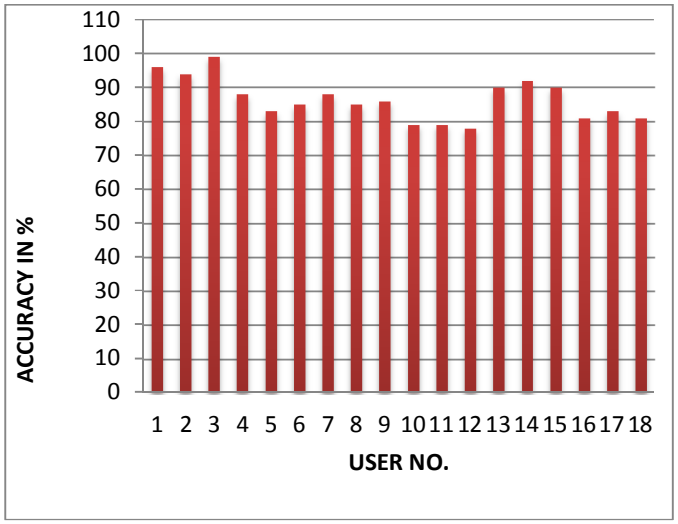


Fig.2 Accuracy graph

IX.CONCLUSION

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, time, load and average fluctuation in user profile or access behaviour is evaluated. And we have also proposed a new enhanced technique of HMAC calculation by addition of pseudo-random generator to further secure the useful information. On the basis of these above mentioned techniques the accuracy of the system is more enhanced and the insider data theft attacks are prevented and the user information is secured .Through this research we have concluded that decoy technology and fog computing together can provide security to real world problems like insider data theft attacks. In future we can extend the working of algorithm, by calculating the accuracy with other attributes such as performance evaluation of the security mechanism. The concept of fog computing is very vast other than security of data we can extend this research for network security through fog computing and also localising the user data a secure geographical locations.

ACKNOWLEDGMENT

This research paper is made possible with the help and support of my parents, teachers, family, friends, and all the people who guided me throughout my work. Especially, I am thankful and I express my gratitude to the following people who contributed and helped to make this work possible. I would like to thank Er. Sachin Majithia for his support and encouragement that motivated me to and write this paper. He kindly read my paper and suggested me advices on grammar, matter, and the title of the paper. I sincerely thank my parents, family, and friends, who gave me emotional and financial support. Without the support of these kind people the product of this research paper would not be possible.

REFERENCES

[1] Salvatore J. Stolfo, Malek Ben Salem and Angelos D. Keromytis, (2012) "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud", Security and Privacy Workshops IEEE Columbia Univ., New York, USA, pp. 125 – 128, May 2012

[2] Ben-Salem M., and Stolfo (2011) "Decoy Document Deployment for Effective Masquerade Attack Detection" in Proceedings of the 8th international conference on Detection of intrusions and malware, and vulnerability assessment, ser. DIMVA'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 35–54.

[3] F. Rocha and M. Correia(2011) , "Lucy in the sky without diamonds: Stealing confidential data in the cloud," in Proceedings of the 2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks
Workshops, ser. DSNW '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 129–134.

[4] B. Grobauer, T. Walloschek, E. Stocker (2011) " Understanding Cloud Computing Vulnerabilities, Security & Privacy", IEEE (Volume:9 , Issue: 2), March-April, 2011, Page(s) 50-57

[5] Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud Position Paper Salvatore J. Stolfo Computer Science Department Columbia University New York , NY, USA Email: sal@cs.columbia.edu

[6] Jiang Zhu, D.S. Chan, M.S. Prabhu, P. Natarajan, Hao Hu, F. Bonomi (2013) "Improving Web Sites Performance Using Edge Servers in Fog Computing Architecture," 2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE), pp. 320-323, March 2013.

[7] M. Ben-Salem and S. J. Stolfo (2011) "Modeling user search-behavior for masquerade detection," Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection. Heidelberg: Springer, September 2011, pp. 1–20.

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

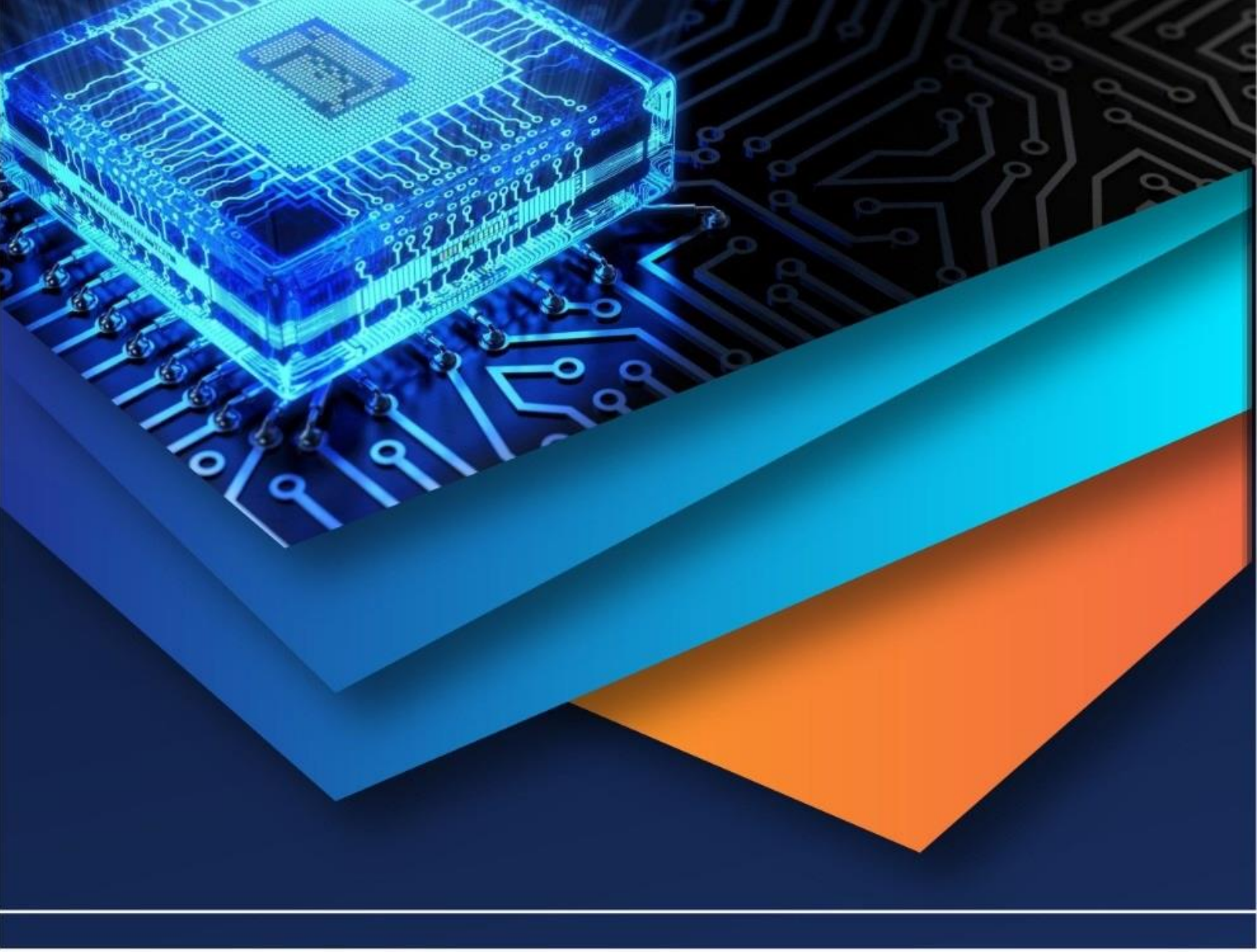
[9]

[10]Ki-Woong Park, Sung Kyu Park, Jaesun Han, Kyu Ho Park, (2010) "Towards Mutually Verifiable Billing Transactions in the Cloud Computing Environment," 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] Marten van Dijk, Ari Juels (2010) "On the Impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing" RSA Security Brief, March 2010



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY