A Hybrid Cloud Security: Model for Preventing Account Hijacking

SHIVA PRASAD VANGARA[1], RAGHAVENDRA VIJAY BHASKER VANGARA[2], KAILASHNATHAN THIRUPATHUR[3], BHARATH TAVIDABOINA[4]

[1]*Department of Engineering Management, Indiana Tech University, Indianapolis*

[2]*Department of Computer Science, University of Missouri- St. Louis, Missouri*

[3]*Department of Computer Science, University of Bridgeport, Connecticut*

Abstract

In the current era, the dominance of electronic gadgets is rapidly increasing and making them into the smart devices is been trending in the market. These smart devices can connect over a network, exchange data to help someone with a particular task or learn from a process, this is internet of things. Now a days, it's all about network, devices and data. To store this huge data, people are opting cloud technology which in turn is very cheap, flexible, which makes them closely affiliated future technologies with one providing the other a platform for success. The main problem with this integration would be with the provision of security. In this paper, we present a two-step process, where we create a secret key for encrypted data through CP-ABE (Cipher text Policy – Attribute Based Encryption). This encrypted data can only be decrypted when the correct key is provided. The data can only be decrypted when the secret key is correctly produced. The second step is for ensuring more security leveraging threshold cryptography where secret key is further encrypted by RSA and then generated key is shared internally to a group of users. Shared key can be produced only if all the authorized users come together. In this way, we can minimize the usage of multiple number of keys and prevent malicious attacks to access the data.

Keywords: *Cloud Security; IOT Security; CP-ABE (Cipher text Policy – Attribute Based Encryption), Threshold cryptography*.

1. Introduction

The Internet of Things (IoT) paradigm is based on smart and auto-configuring nodes (things) interconnected over a dynamic and global network infrastructure. It is one of the most innovative, ubiquitous, and pervasive computing technologies. Internet of things is all about little things in the real world, their limited storage, their processing ability, and their problems with reliability, performance, security and privacy. To make it simple, the internet of things can be as simple as a smart mirror or a smart door lock or a smart security system which can caution its owners about a potential danger. The main objective of IoT devices is to transfer data over a network with designated IP address and stay connected to other devices as long as the network is open and active.

On the other hand, cloud computing is all about unlimited capacity of storage and its processing power. It is a revolution in IT economy and with its emergence, the market's perspective has been changed creating new opportunities, new business models and a new business era. In simple words, cloud computing can be explained as providing different types of services over internet. It can be a software, a networking resource everything can be delivered through cloud. The day to day

example of cloud services would be email server, electronic drives, mobile apps, games etc. The main features of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.

When these two technologies come together, they can solve many complex problems and surprisingly that's been the case so far. IoT cloud enables organizations to leverage the significant potential of IoT without having to build the underlying infrastructure and services from scratch. IoT cloud also helps promote and ensure standardization in key areas, including how devices communicate with each other, device management, and security. Since these two are new to the market, the primary concern would be providing security. Since both these etechologies dealt with private and confidential information of the customers, there is huge risk of malicious attacks and account hijacking from hackers. To prevent this attacks, traditional security provisions are not enough. There is always a need for new security mechanisms to handle such kind of risks in these innovative technologies. So, it is important to research such threats early and come up with new solutions to protect such technologies which are boon to the human world. We formulated a two- step process which would ensure security to this Cloud IoT paradigm.

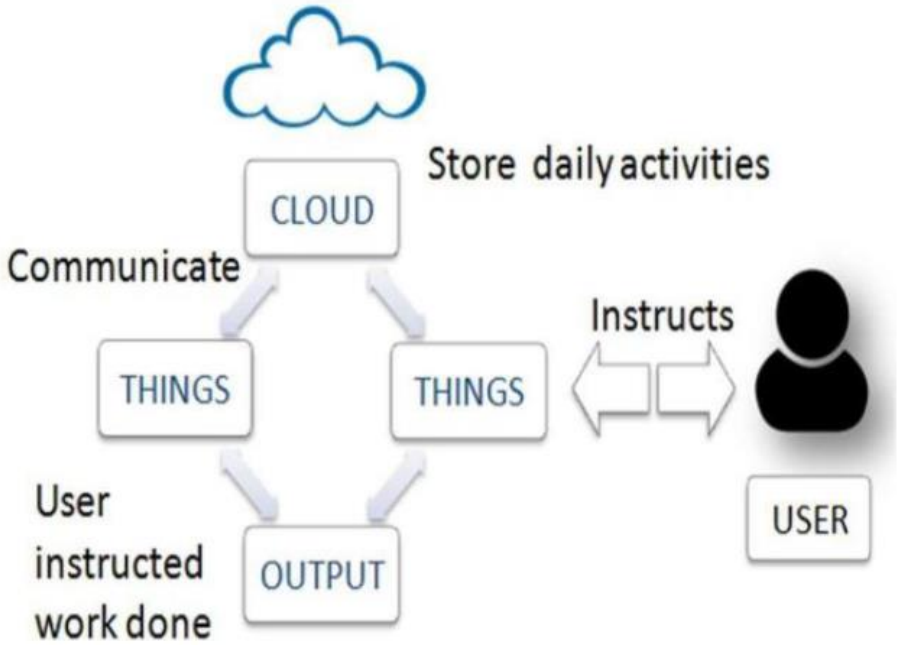


Fig:1 Cloud IoT Paradigm

The rest of the paper is organized as follows: Section II discusses the related work. In Section III, proposed work has been discussed in context of the log files dataset. Section IV we discussed the results of different data mining methods and compare the results. Conclusion of the paper and future work plans is presented in Section V.

2. LITERATURE SURVEY

Yadav P. et al. [1] proposed ecdsa algorithm to proof that elliptic curve based public key cryptography is far better than RSA based schemes. They compared its performance with RSA based algorithm in cloud.

Sunita R. et al. [2] proposed an encryption technique called hybrid algorithm in order to provide privacy in the cloud. It enables greater coverage and a far higher degree of accuracy, including first generation hybrid technology or dynamic and static testing conducted in isolation. Organizations can take advantage of this method for their transactions. Additionally, this technique can makes it possible to speed the remediation of critical issues, gain better insight into vulnerability root causes, and simplify software security assurance processes with unified reporting that organizes results by severity, correlation, common causes, and precise code location.

Arora, P. et al. [3] describes about the performance of different security algorithm on a cloud network and also on a single processor for different input sizes and advanced Encryption Standard security algorithm implemented for ensuring security framework.

Zissis, D. et al. [5] proposes introducing a Trusted Third Party, tasked with assuring specific security characteristics within a cloud environment. The proposed solution calls upon cryptography, specifically Public Key Infrastructure operating in concert with SSO and LDAP, to ensure the authentication, integrity and confidentiality of involved data and communications. The solution presents a horizontal level of service, available to all implicated entities, that realizes a security mesh, within which essential trust is maintained.

Sachin, G. et al. [7] proposed a methodology regarding the security in the cloud, in which they proposed a access control system so the unauthorized user cannot access the data easily.

Sachin, G. et al. [8] provides an advance Cross Site Scripting Attack, which could be harmful for any web site vulnerable to Cross Site Scripting Attack. After successfully implementing this attack over a vulnerable website, the attacker would be able to gain access over the content and source code of the website. The attacker will also be able to get user information from the database directory. Attackers don't have to send vulnerable link to the victim in order to steal user's data. a call for the developers, the Cross-Site Scripting Attack can be more harmful as it was in the past.

Wenge, O. et.al [10] explains concepts, mechanism and available tools that can support establishing of security-aware cloud markets. Furthermore, they introduce their information security governance driven cloud brokerage model with security labelling of tradable cloud products that can be the next step in the standardization process of tradable cloud products and optimize the selection of collaborative cloud partners.

3. PROPOSED WORK

We Built a pre-processing model of dataset which we download from cloud. File is in the form of log entries. We have also plan to use compression algorithm for the optimization purpose.

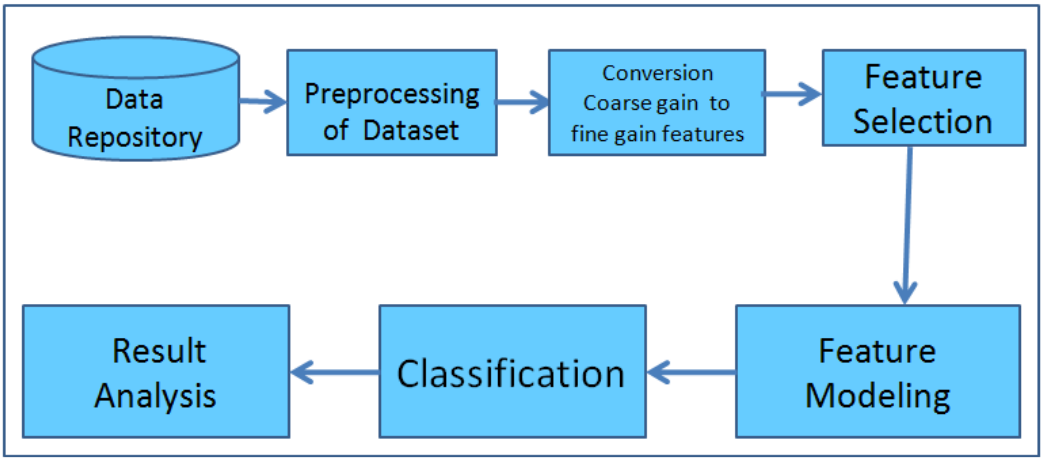


Figure 1: Block Diagram Proposed Method

Figure. 1 shown block Diagram of proposed model Our model had been passes through seven different steps. In first step we download the data from data Repository then we pre-process the dataset convert it and apply feature selection and feature modelling techniques. After these processes we apply classification model and done the result analysis of these models on different parameters.

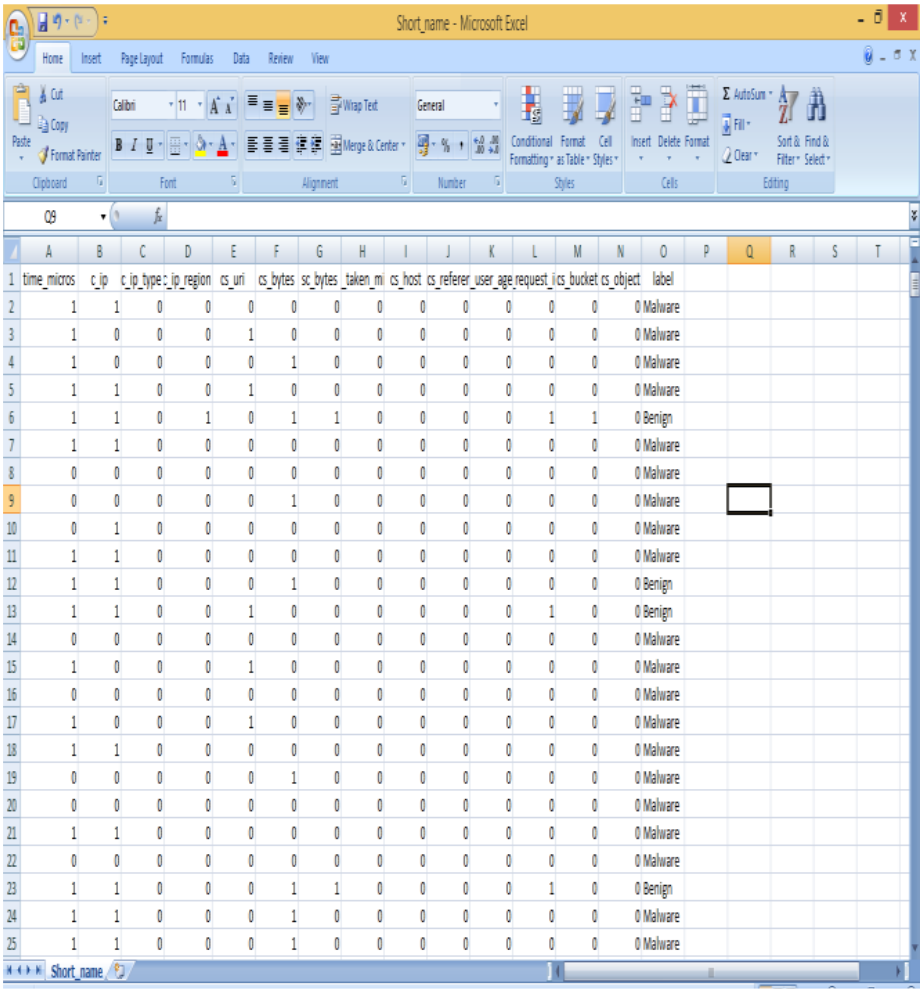
3.1. Proposed Algorithm:

Algorithm-

Step 1: Read the one line from the access log file.
Step 2: Calculate the length of current line
Step 3: If space Count is not as predefined count then got to step 1
Step 4: Else write the current line into pre-processed log file
Step 5: Extract the one log entry from the pre-processed access log file and store it into string
Step 6: Check parsed string is in the loopback address range then assign string variable user with "Admin"
Step 7: If parsed string is in range of private address range then assign string variable user with Insider.
Step 8: Else If assign string variable user with "Outsider"
Step 9: Apply Feature selection and Feature Appearance vector on pre-processed dataset.
Step 10: Apply Data mining classification techniques (Logistic Regression, Random Forest, Decision Tree) after feature selection and feature modelling on pre-processed log file dataset
Step 11: Evaluate the model's performance on Confusion Matrix, ROC and Lift curve and compare the performance. The data of log features downloaded from the cloud sources. The data contains the log features of the cloud file. We pre-process the data set and apply data mining techniques to dataset. To apply the classification models on the dataset we pre-process our target variable into two classes and apply a decision tree, random forest and logistic regression model on our dataset with R language.

3.1. Data Repository:

The standard dataset has been used for analysis; Dataset file is of approx 20 GB size that contains the access log file of cloud environment. sample dataset is represented in figure 2.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	time_micros	c_ip	c_ip_type	c_ip_region	cs_uri	cs_bytes	sc_bytes	time_taken_mi	cs_host	cs_referer	user_age	request_i	cs_bucket	cs_object	label					
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
3	1	0	0	0	1	0	0	0	0	0	0	0	0	0	Malware					
4	1	0	0	0	0	1	0	0	0	0	0	0	0	0	Malware					
5	1	1	0	0	1	0	0	0	0	0	0	0	0	0	Malware					
6	1	1	0	1	0	1	1	0	0	0	0	1	1	0	Benign					
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Malware					
10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
11	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
12	1	1	0	0	0	1	0	0	0	0	0	0	0	0	Benign					
13	1	1	0	0	1	0	0	0	0	0	0	0	1	0	Benign					
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
15	1	0	0	0	1	0	0	0	0	0	0	0	0	0	Malware					
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
17	1	0	0	0	1	0	0	0	0	0	0	0	0	0	Malware					
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
19	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Malware					
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
21	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Malware					
23	1	1	0	0	0	1	1	0	0	0	0	1	0	0	Benign					
24	1	1	0	0	0	1	0	0	0	0	0	0	0	0	Malware					
25	1	1	0	0	0	1	0	0	0	0	0	0	0	0	Malware					

Figure 2: Classified Database

Each log entry of access file contains „n" dimensional attribute vector (Fine grained features of log entries F: (f1, f2,…..,fn), Where fi is the value of attribute Fi. There are 2 classes: Malevolent (Cm) and Benevolent (Cb).

3.3. Pre-Processing:

The raw permissions do not arrive in a format conducive to fruitful detection. Pre-processing helps to improve the quality of data, efficiency and effectiveness. Therefore, substantial pre-processing must be applied. The pre-processing tasks applied in this work are –

- Data Cleaning and Filtering.
- Log files contains number of raw and irrelevant entries.
- Identify the duplicity of the record in access Apps.

3.4. Feature Selection:

Feature selection is dissimilar from dimensionality reduction. Both methods seek to reduce the number of attributes in the dataset. [9]

Many databases offer applications that may download and view freely and paid. Log entries features provide the information related to the request made by the users to

access the resources of end device. It may represent the behaviour of resources used. To simulate the behaviour of cloud has been extracted. Common features extracted from different cloud database applications.

3.5. Feature Modelling:

Selected features of the log entries have been parsed one by one and create the Feature Appearance Vector (FAV) [10]. The FAV in binary form only captures the types of feature appeared in the application or not. The ith element value in the vector is set to be 1 if the ith permission is present in the extracted permission. Otherwise, it is set to be 0 as shown in Table 1.

Log Entry	Cp_ip 1 (Le_1)	Cp_ip 2 (Le_2)	Cp_ip 3 (Le_3)	time_ micros (Le 4)	Cs _byts (Le5)	Sc_byts (Le_6)	….	Cs_h ost (F_n)	Class Label
L_1	1	1	0	0	0	0	…	0	C_M
L_2	1	1	0	0	1	0	…	0	C_b
L_3	1	1	0	0	0	1	…	0	C_M
L4	1	1	0	0	1	0	…	0	C_M
L_5	1	0	0	1	0	1	…	0	C_b
L_6	1	1	0	0	0	0	…	0	C_M
L_7	1	1	0	0	0	0	…	0	C_M

Figure 3: Database After Feature Selection

3.6. Classification Methods:

There are many classification data mining techniques, but we choose logistic regression method over other methods because we have imbalanced and noisy dataset. Other data mining methods are not work well on this type of dataset. It gives very good performance on our dataset and better accuracy rate rather than models. The constant (a0) in the logistic regression moves the curve left and right and the slope (a1) defines the steepness of the curve. the logistic regression equation can be written in terms of an odds ratio.

$$\frac{p}{1-p} = \exp(a_0 + ab_1 x)$$

Taking the natural log of both sides, we can write the equation in terms of log-odds (logit) which is a linear function of the predictors [11]. The coefficient (a1) is the amount the logit (log-odds) changes with a one-unit change in x.

$$ln(\frac{p}{1-p}) = a_0 + a_1 x$$

4. RESULTS AND DISCUSSION

To evaluate the performance of models we used confusion Matrix, ROC curve and Lift Curves of the models. To find the best accuracy is randomly divided into 10 smaller subsets where dataset contains 612 entries, 9 subsets used for training and 1 subset is used for testing. The process is repeated 10 times for every combination. To calculate the results of different classifications models we used R language ,machine learning methods with rattle that is written in R language , the package consist inbuilt different data mining techniques we run the model and find the different accuracy parameters. To find the results on our preprocessed log entries data set. we used three different data mining techniques random forest, logistic regression and decision tree, we run the these data mining techniques in R language with rattle data mining package and find the results, random forest gives the accuracy of 87.25% ,decision tree 87.09% and logistic regression 88.39%4.1. Data

Table No. 1. Comparison of Accuracies

Algorithm	Accuracy	Sensitivity	Specifity
Decision Tree	**87.09%**	**82.60%**	**87.45%**
Random Forest	**87.25%**	**89.74%**	**87.08%**
Logistic Regression	**88.39%**	**77.63%**	**88.16%**

As we shown table no.1 that in the lift curve of three predicted model decision tree random forest and logistic regression in the chart we show glm() function has highest lift value (which defined logistic regression) in the curve when we rank the probabilities of all three model and find the curve so in our experiments LM get the highest lift rate.

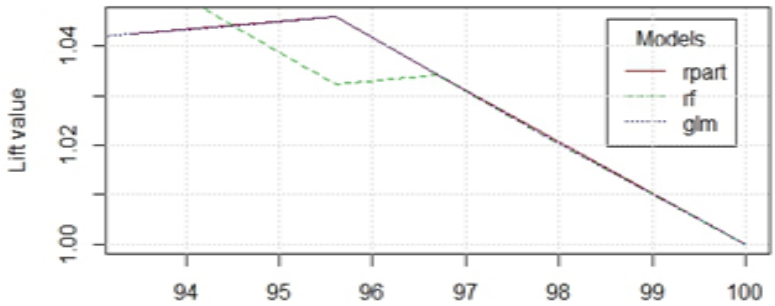


Fig .4: Comparison of Models with Lift Curve

In this work ROC curve are used because predicted probabilities of log entries are not properly classified or even when classes are highly unbalanced [12]. Roc Curve obtained by the decision tree model roc curve visualize plot between the true positive rate and the false positive rate [13]. The curve(fig no.5) shows that it has 0.62 values has under the curve. So it is a good model on our data set. In a below diagram we can see, that sensitivity at this threshold is 38% and the (1-specificity) is ~60%. To bring this curve down to a single number, we find the area under this curve (AUC)

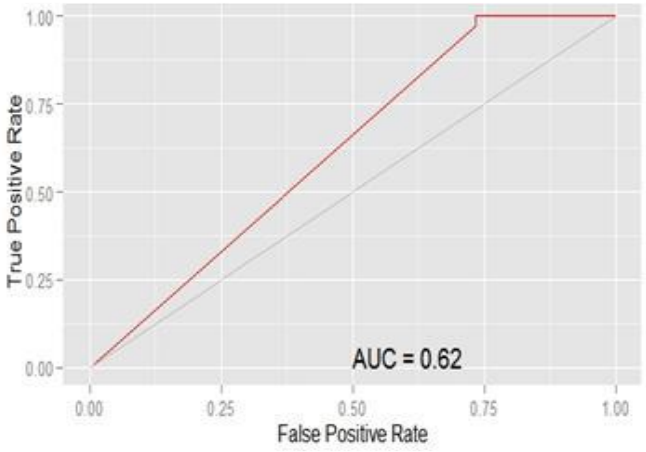


Fig .5 ROC Curve of Decision Tree Model

The curve (fig no.6) shows that it has 0.72 values has under the curve. So it is a good model on over data set. As a diagram we can see, that sensitivity at this threshold is 38.2% and the (1-specificity) is ~72%. To bring this curve down to a single number, we find the area under this curve (AUC)

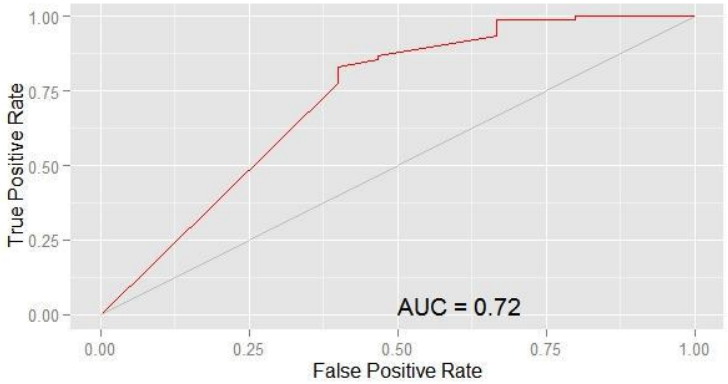


Fig. 6- ROC Curve of Random Forest Model

The curve (fig no.7) shows that it has 0.84 values has under the curve. So it is a good model on over data set. As a diagram we can see, that sensitivity at this threshold and the (1-specificity) is ~84%.To bring this curve down to a single number, we find the area under this curve (AUC)

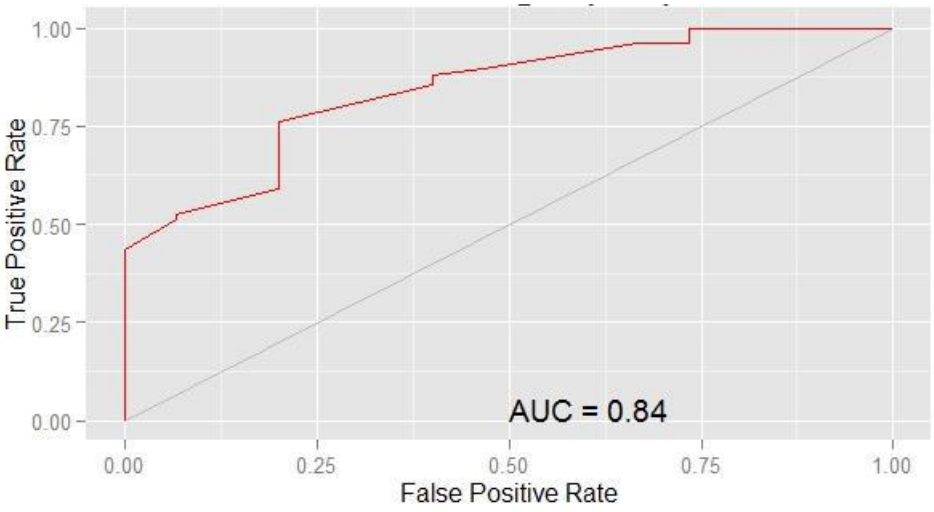


Fig 7: ROC Curve of Logistic Regression

5. CONCLUSION

In this study, we addressed the analysis of cloud log files to prevent account hijacking attacks in hybrid cloud. We discussed the security vulnerability threat and how attacks could happen and the consequences of it can be deadly. We proposed a concept to prevent attacks and implemented it on the dataset and produced results. We produced results with three different data mining techniques random forest, logistic regression and decision tree. We ran these data mining techniques in R language. After applying these three data mining methods to the dataset logistic regression gives the highest accuracy of 88.39%. For the future scope we can use different non-linear machine learning techniques to improve our result.

References

6.1. Journal Article

[1] Tripathi, Yadav, P. (2012). Enhancing Security of Cloud Computing using Elliptic Curve Cryptography, International Journal of Computer Applications (0975-8887) Volume 57-No. 1.

[2] Sunita, R., & Ambrish, G. (2012). Cloud Security with Encryption using Hybrid Algorithm and Secured Endpoints, International Journal of Computer Science and Information Technologies, Vol. 3, pg. 4302-4304

[3] Arora, P. & Tyagi, A. S. (2013). Evaluation and Comparison of Security Issues on Cloud Computing Environment. IOSR Journal of Computer Engineering (IOSR-JCE), Volume 11, Issue 1, May. - Jun. 2013, PP 39-45.

[4] Brunette, G., & Mogull, R. (2009). Security guidance for critical areas of focus in cloud computing v2. 1. *Cloud Security Alliance*, 1-76.

[5] Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. Future Gener. Comput. Syst., 28, 583-592.

[6] Bodkhe, A.P., & Dhote, C.A. (2015). Cloud Computing Security: An Issue of Concern.

[7] J. Ong, "Domain registrar Melbourne IT at centre of sea meddling with New York times, twitter," August 2013. Available: urlhttp://thenextweb.com/twitter/2013/08/28/domain-registrarmelbourne-it-at-center-of-sea-meddling-with-new-york-times-twitterwhois-info.

[8] Ratish Agrawal and Sachin Goyal Ankit Valdaya, "A Methodology for Development and verification of Access Control system in Cloud Computing," Advanced Research in Computer and Communication Engineering, vol. 4, no. 3, March 2015.

[9] Vangara, S. P., Vangara, R. V. B., Thirupathur, K., & Tavidaboina, B. (2017). "Security for Cloud storage through Deniable Attribute-Based Encryption". Journal of Engineering, Computing and Architecture (JECA), 7(12), 27-33.

[10] S. Alliance, "The notorious nine: Cloud computing top threats 2013," 2013. Available: urlhttps://downloads.cloudsecurityalliance.org/initiatives/top threats/The Notorious Nine Cloud Computing Top Threats in 2013.pdf.

[11] Wenge, O., Schuller, D., & Steinmetz, R. (2014). Towards Establishing Security-Aware Cloud Markets. 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, 1027-1032.

[12] Li W, Ping L. (2009). Trust model to enhance Security and interoperability of Cloud environment. In Proceedings of the 1st International conference on Cloud Computing. Beijing, China: Springer Berlin Heidelberg; 69–79.

[13] Siebenhaar, M., Wenge, O., Hans, R., Tercan, H., & Steinmetz, R. (2013). Verifying the Availability of Cloud Applications. CLOSER.

[14] Nabi, M., Toeroe, M., & Khendek, F. (2016). Availability in the cloud: State of the art. J. Netw. Comput. Appl., 60, 54-67.

[15] Naldi, M. (2013). The availability of cloud-based services: Is it living up to its promise? 2013 9th International Conference on the Design of Reliable Communication Networks (DRCN), 282-289.

[16] Karajeh, H., Maqableh, M., & Masa'deh, R. (2016). Privacy and Security Issues of Cloud Computing Environment.