A Secured Approach to Credit Card Fraud Detection Using Hidden Markov Model

Twinkle Patel, Ms. Ompriya Kale

***Abstract*: - As the usage of credit card has increased the credit card fraud has also increased dramatically. Existing fraud detection techniques are not capable to detect fraud at the time when transaction is in progress. Improvement in existing fraud detection is necessary. In this paper Hidden Markov model is used to detect the fraud when transaction is in progress. Here is shown that hidden markov model is used to detect credit card fraud with reduced false positive transaction. HMM categorizes customers profile as low, medium and high and based on spending profile set of probability for amount of transaction is assigned to each cardholder. Amount of new transaction is checked against the profile of card holder, if it justifies a predefined threshold value then transaction is accepted else it is considered fraudulent. But still HMM is not secured for initial some transaction during training so HOTP is used as secured approach with HMM to reduce the fraud and to increase the security. HOTP is one time password which is used one once and it is send to the client mobile when HMM detects that amount is more than threshold value if the user enters valid HOTP then only transaction is allowed to progress else it is detected as fraud.**

***Keywords*: Hidden markov model; credit card; fraud detection, HMAC one time password (HOTP).**

I. INTRODUCTION

Credit card frauds are increasing day to day as the use of credit card is increasing. Instead of various fraud detection techniques fraudsters are so expert that they finding new ways for committing fraudulent transaction. Occurrence of credit card fraud has increased dramatically both online and offline. Credit card based purchase can be done in two ways: (i) physical card (ii) virtual card. In physical card purchase, the cardholder presents his card physically to the merchant for making payment [2]. For this type of fraud attacker has to steal the credit card. In virtual card purchase only the information about the card is stolen or gathered like card number, secure code etc. such purchases are done over an internet. For these type of fraud attacker needs only card details so only way to detect these type of fraud is to analyze the spending pattern of card holder.

II. FRAUD TECHNIQUES

Various types of fraud techniques are as follows:

A. Site Cloning

In site cloning the fraudster clone an entire site or just the payment page of the site where customer make a payment. Customer feels that they are viewing the real site. The customer handover a credit card detail to the fraudster and then fraudster sends the customer a transaction receipt via email as real site. Thus fraudsters have all detail of customer credit card so they can commit fraud without customer's awareness.

B. Stolen / Lost Credit Card

When customer card is lost or stolen by fraudster he gets all the information of the cardholder in the easiest way without investing any modern technology. It is difficult form of credit card fraud to detect.

C. Skimming

Skimming is one of the popular forms of credit card fraud. It is a process where the actual data on a card is electronically copied to another. It is very difficult for cardholder to identify this type of fraud.

D. Credit Card Generator

In credit card generator the computer program generates the valid credit card number and expiry gate. This generator creates a valid credit card highly reliable that it looks as the valid credit card number only and are also available for free download off the internet.

E. Phishing

In phishing the fraudster sends lots of false email to card holder. The e-mail looks like they came from the website where the customer trust for example customers bank. The e-mail asks the customer to provide personal information like credit card number. With the help of these details fraudster commits crime.

F. Internal Fraud

The employee or owner access customers detail. The steal the customer's personal information to commit crime or pass on the information about cardholder to fraudster for money.

III. LITERATURE REVIEW

S. Esakkiraj and S. Chidambaram et al describes "A predictive approach for fraud detection using hidden markov model" [3]. In this paper they have used hidden markov model for credit card fraud. In using this method they have explained

how the hmm can detect whether incoming transaction is fraudulent or not.

Abhinav srivastava et al describes "Credit card fraud detection using hidden markov model" [1]. In this paper they have used the transaction amount as the observation symbols and suggested a method for finding the spending profile of cardholder with the help of which HMM can detect the incoming transaction is fraudulent or not.

Raghavendra Patidar and Lokesh Sharma et al describes "Credit card fraud detection using neural network" [2]. In this paper they used neural network along with genetic algorithm to detect fraudulent transaction. They have explained that if neural network is trained properly it can work as a human brain. The working of neural network is based on the neurons and thus the output varies in which the topology varies.

I-Cheng Yeh and Che-hui lien[7] have examines and compares the performance six major classification techniques. They have shown the results of the classification and predictive accuracy of six data mining techniques. The result shows that there are little difference of error rates and big difference of area ratio among six different techniques. ANN performs better and more accurate among five methods.

A. Fraud Detection Techniques

Data Mining is a process of discovering patterns from large quantities of data so it is one of the powerful tools for decision support system and plays a key role in fraud detection. Various data mining techniques for fraud detection are mentioned below.

1) Logistic Regression

Logistic regression is a special case of linear regression analysis. Logistic Regression is used for predicting the outcome of a dependent variable based on one or more predictor variable. Here predictor variable can be either categorical or numerical.

Advantage: It produces a simple probability formula for classification. It works well with linear data for credit card fraud detection.

Disadvantage: It does not deal with non linear data for credit card fraud detection.

2) Decission Tree

Decision tree are commonly used in credit card fraud detection. Decision tree is a flow based structure in which internal node represent an outcome of the test on an attribute and branch represents an outcome of the test and leaf node represent classes. Root node is the top most node of the tree. Decision tree predicts the output of the target variable based on one or more input variable Decision tree algorithms are ID3, C4.5, CART, MARS.

Advantage: It can handle non linear and interactive effects of input variables.

Disadvantage: It has complex algorithm. Even a small change in observed data might change the structure of a tree. Choosing splitting criteria is also difficult.

3) K-Nearest Neighbour Algorithm

The concept of k-nearest neighbour can be used in many analogy detection techniques. Credit card fraud can be detected by using k-nearest neighbor algorithm. Here in KNN new transaction is classified based on the closeness i.e. distance. In KNN we classify any new incoming transaction by calculating closeness or distance to other transaction. If they are close then transaction is ok else the transaction is indicated as fraud.

Advantage: It does not require establishing any predictive model before classification.

Disadvantage: accuracy is highly dependent on the measure of distance.

4) Naïve Bayes Algorithm

Naïve bayes classifier makes a conditional independence assumption that the effect of an attribute value of a given class is independent of other attributes. It is based on bayes theorem.

Advantage: It only provides a theoretical justification to the fact but does not use bayes theorem.

Disadvantage: In real practice the dependences exists between the variable.

5) Artificial Neural Network

Although there are several fraud detection techniques based on knowledge detection, expert system, data mining etc but still they are not capable to detect the fraud at the time when fraudulent transaction is in progress but with the help of techniques like Neural Network, Hidden Markov Model the fraudulent transaction can be detected during the transaction is in progress.

When a customer use a credit card there is a particular pattern of credit card use made by a customer. By using this previous data neural network is train about particular customer. As shown in figure 1 neural network is trained based on his income, occupation, location, number of large purchases , location where large purchases are done etc. Based on these pattern the neural network classify whether a particular transaction is fraudulent or genuine

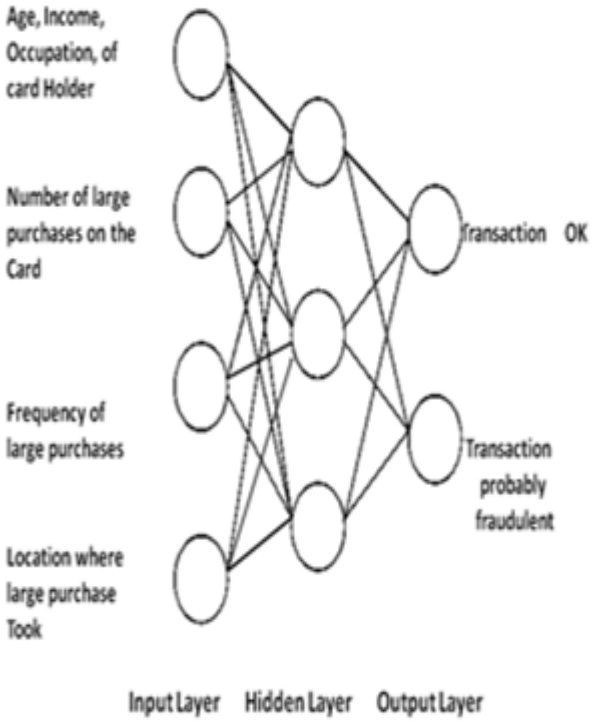

Figure 1: Layer of Neural Network in Credit Card [3]

The neural network produces output in real value between 0 to 1. If the neural network produce output below .7 then the transaction is ok and if it produce a output above .7 then the change of transaction being illegal increases.

Advantage: It detects the fraudulent transaction at the time when transaction is in progress.

Disadvantage: 1) Number of parameters to be set before training begins. There are no clear rules to set these parameters. 2) Network differ in the way their neurons are interconnected i.e. topology of a network has a large influence on the performance of network and so far there is no method that determine optimal topology for a given problem.

IV. COMPARISION OF VARIOUS TECHNIQUES

Fraud Detection Techniques	Advantage	Disadvantage
Logistic Regression	It produces a simple probability formula for classification. It works well with linear data for credit card fraud detection.	1) It does not deal with non linear data for credit card fraud detection. 2) **It is not capable to detect the fraud at a time when fraudulent transaction is in progress**.
Decision Tree	It can handle non linear and interactive effects of input variables.	1) It has complex algorithm. Even a small change in observed data might change the structure of a tree. Choosing splitting criteria is also difficult. 2) **It is not capable to detect the fraud at a time when fraudulent transaction is in progress**.
K-Nearest Neighbour Algorithm	It does not require establishing any predictive model before classification	1) Accuracy is highly dependent on the measure of distance. 2) **It is not capable to detect the fraud at a time when fraudulent transaction is in progress**.
Naïve Bayes Algorithm	It only provides a theoretical justification to the fact but does not use bayes theorem	1) In real practice the dependences exists between the variable. 2) **It is not capable to detect the fraud at a time when fraudulent transaction is in progress**.
Artificial Neural Network	**It detects the fraudulent transaction at the time when transaction is in progress.**	1) Number of parameters to be set before training begins. There are no clear rules to set these parameters. 2) Network differ in the way their neurons are interconnected and so far there is no method that determine optimal topology for a given problem.
Hidden Markov	1) **Hidden Markov Model**	1) It detects the fraud only after some

Model.	**(HMM) is capable to detect the fraudulent transaction is in progress**. 2) The main feature of the HMM-based model is **reducing in False Positive (FP) transactions predict as fraud** by a fraud detection system even though they are really genuine customer.	transactions so it is not secured for initial some transactions.

V. HIDDEN MARKOV MODEL

In HMM considered three price ranges such as low (l), medium (m) and high (h). First we need to find out transaction amount to a particular group either it will be in low, medium or high [1]. For example let l= (0, $150) m= ($150, $ 250) and h= ($250, credit card limit). If transaction of card holder is $70 then price range or observation symbol is low.

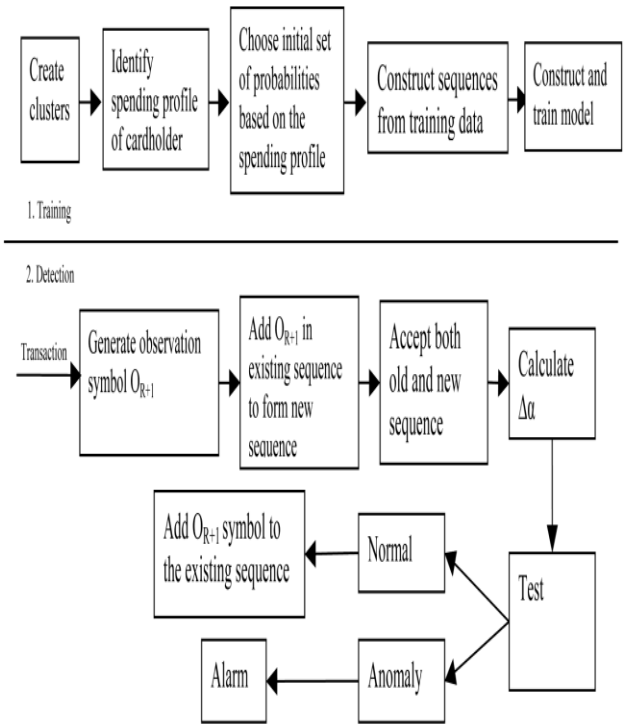

Figure 2: Process flow of credit card fraud detection system [1].

As shown in figure 2 there are two phases of HMM. In training phase card holder transaction amount is converted in observation symbols i.e. low medium or high and form sequences from them. After the sequence is formed threshold is calculated from the sequence of amount.

In the detection phase client enters amount and form an initial sequence of symbol. Let O_1, O_2, O_3.....O_R be such sequence of length R up to time t. This sequence is the input to HMM and from that we compute the threshold of acceptance α_1. Let O_{R+1} be a symbol of new transaction at time t+1.now with new transaction we generate a new sequence O_2, O_3,.....O_R ,O_{R+1}. We input these sequence in HMM and calculate the new threshold of acceptance if amount is less than threshold than amount is added in new sequence else the it is detected as anomaly

VI. HMAC ONE TIME PASSWORD (HOTP)

HMAC is hash message authentication code. HOTP is used for generating an 8 digit one time password for security. An important steps to obtain HOTP is describe in Figure3.

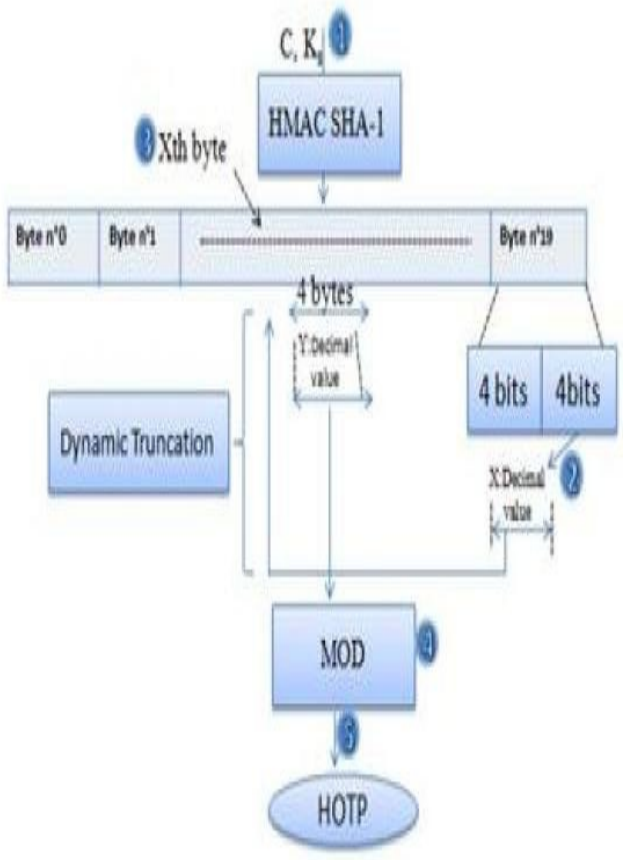

Figure 3: HMAC one time password [8]

As shown in figure step1 is generating HMAC-SHA1 value. Let HS= HMAC SHA1 (K, C) [9] where K is secret key and C is a counter. Step2 and step 3 is generating 4 bytes string i.e. dynamic truncation. Lets understand that by taking example [9]

Byte Number

|00|01|02|03|04|05|06|07|08|09|10|11|12|13|14|15|16|17|18|19

Byte Value
|1f|86|98|69|0e|02|ca|16|61|85|**50|ef|7f|19**|da|8e|94|5b|55|**5a**|

----------*********----------++

As shown in example the last byte has hex value 5a which means the value of lower four byte is 10 (a). Now the value of four byte starting at byte 10 is 50ef7f19. Now we will convert 50ef7f19 into number. $5\times16^7+0\times16^6+14\times16^5+15\times16^4+7\times16^3+15\times16^2+1\times16^1+9\times16^0$ = 1357872921

Step4 is to get 8 digit numbers by performing modulo. HOTP-Value = HOTP (K, C) mod 10^d [9] where d is desired numbers of digit. For example 1357872921 mod 10^8=**57872921** thus 57872921 is 8 digit HOTP. This is how HMAC one time password is generated and send to mobile as sms for security.

VII. PROPOSED WORK

In the proposed work HMM is used along with HOTP to make HMM more secured as we have seen above HMM needs training and during training some transactions are involved and fraud is not detected during training but it is detected after training so HOTP is used for secured approach in HMM so make initial transaction secure by sending one time password i.e. security code to clients mobile if the security code entered by client is correct then only transaction is done successfully else transaction is not allowed to progress. But once the HMM is trained and ready for detection client does not need to enter any security code unless HMM detects the transaction is above threshold value. If the transaction is above threshold value security code is send to mobile and client need to enter that security code then only transaction is done successfully else transaction is not allowed to progress.

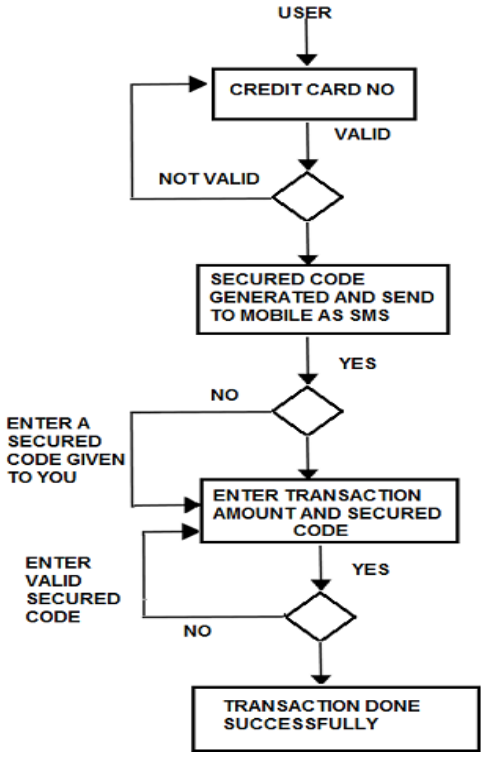

Figure 4: Proposed model during training.

As shown in figure 4 here client has to enter a valid credit card number then a security code i.e. HOTP will be generated and sent to mobile as sms client has to enter that security code and amount if it is valid then only transaction is done successfully. During training of HMM the client has to enter the security code every time then only transaction can be done.

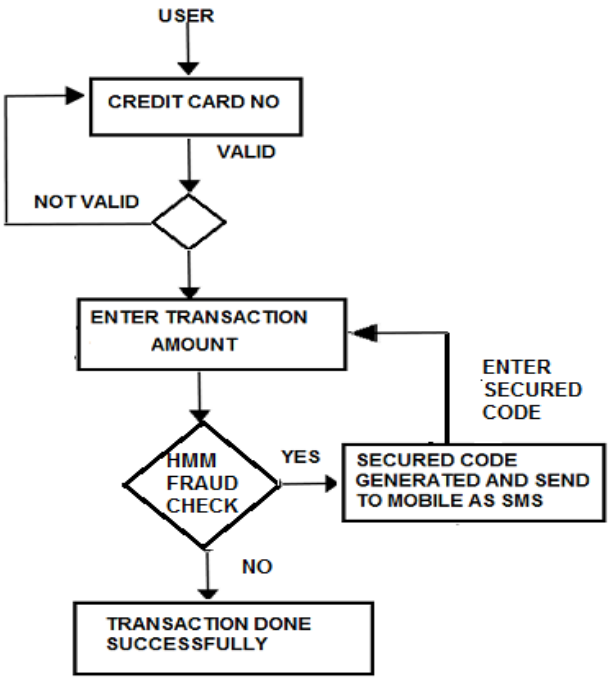

Figure 5: Proposed model of credit card fraud detection after training during detection

As shown in figure 5 after training the client need to enter credit card number and amount if HMM detects that amount is less than threshold value then transaction is done successfully else if amount is more than threshold value HOTP is generated and send to clients mobile. Client has to enter that security code if it is valid than only transaction is successfully else transaction is not able to progress and transaction cannot be done. After training client do not have to enter the security code every time only when amount is more than threshold value client need to enter security code generated for the security reason.

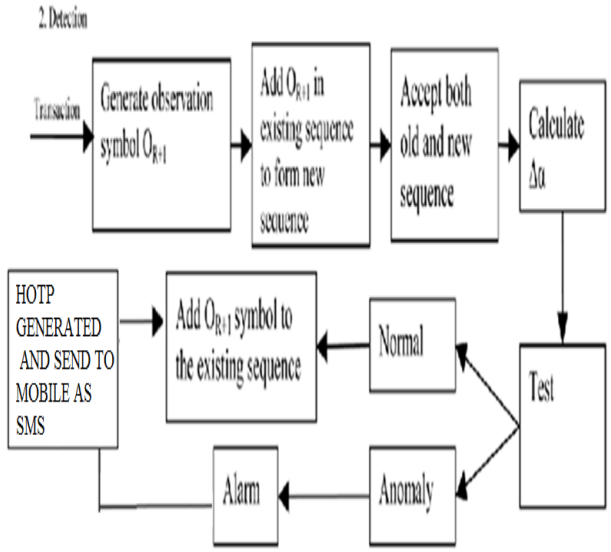

Figure 6: Hidden markov model along with HOTP during detection

During detection as shown in Figure 6the client enters the amount which is observation symbol whether it is low, medium or high then that amount is added in the sequence and both the sequence is accepted and new threshold value is calculated and amount is tested if it is above threshold value then the HOTP is generated and send to mobile as sms and then if client enters the valid HOTP amount is added in sequence else if amount is below threshold value then the transaction is done without need of HOTP to be entered by the client.

VIII. RESULT AND ANALYSIS

Result for HOTP generated and send to mobile as sms are shown in figure 7 below. HOTP is 8 digit one time password which is generated as shown in figure.

Figure 7: HOTP generated and send as SMS

Results during the training are as shown in Figure 8 during training clients need to enter a valid credit card number and secured code which is send to mobile as sms for successfully transaction.

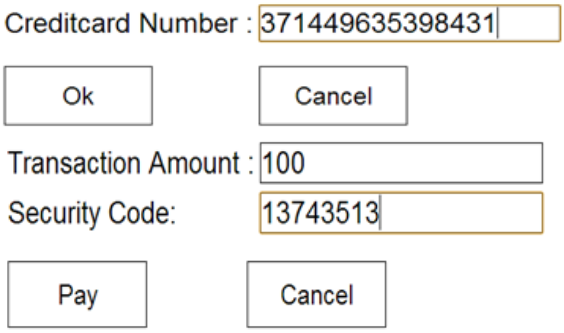

Figure 8: Result during training

Results after training are shown in figure below where client don't have to enter the secured code every time client can do transaction if HMM detects if amount is less than threshold else the HOTP is generated and send to mobile which client need to enter in secured code fields for security reason as shown above in figure.

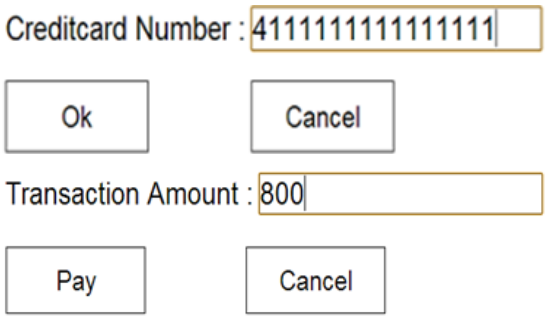

Figure 9: Result during detection

For analysis of clients amount the cluster is formed which shows the no of transaction and the amount that falls in low medium or high.

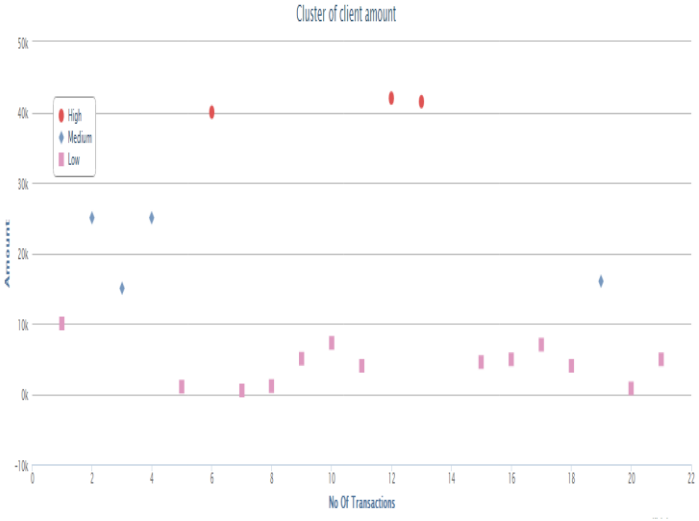

Figure 10: Cluster of client amount

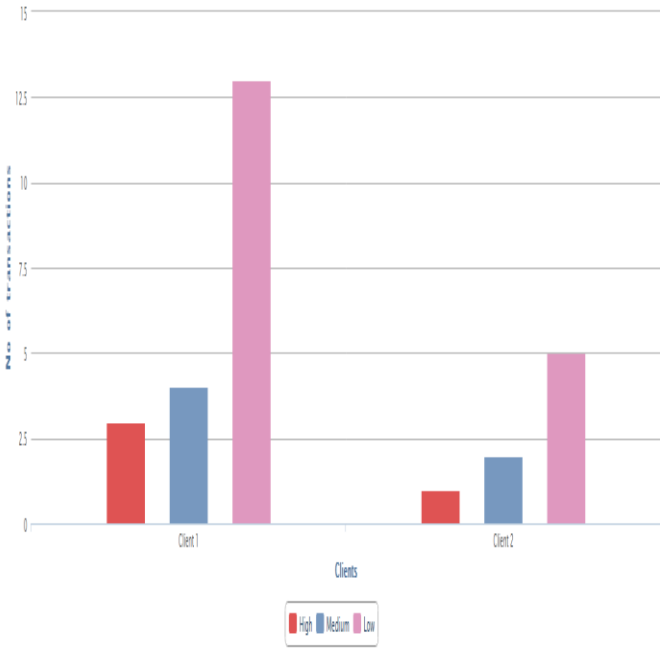

Figure 11: Different transactions of clients

As shown in figure11 shows the number of transaction of different clients red color is high transaction, blue is medium transactions and pink in low transaction based on these transaction the threshold is decided.

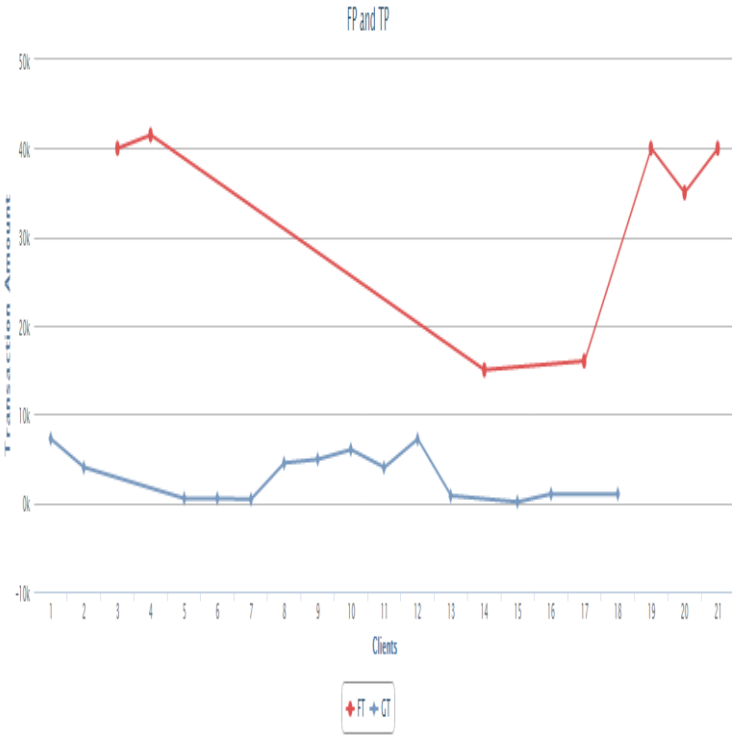

Figure 12: FT and TP transactions

As shown in figure 12 analyses show False positive and true positive transaction detected by HMM. Once HMM detects the false positive transaction HOTP is generated and send to clients mobile if client enter that correct

HOTP then only transaction is done else client cannot proceed. Thus by using HMM with HOTP we get more security.

IX. CONCLUSION

In this paper we have brief discussion on credit card fraud detection using Hidden Markov Model. Here we have shown how the HMM can detect whether an incoming transaction is fraud or genuine. HOTP is use to generate 8 digit unique security code. In proposed model we have use HMM with HOTP to provide more security and to reduce the fraud.

REFERENCES

[1] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, et al. "Credit Card Fraud Detection Using Hidden Markov Model", "IEEE transactions on dependable and secure computing", vol. 5, no. 1, January-march 2008 .

[2] S. Esakkiraj and S. Chidambaram, et al. " A Predictive Approach for Fraud Detection Using Hidden Markov Model ", "International Journal of Engineering Research & Technology (IJERT) ", Vol 2, January 2013

[3] Raghavendra Patidar and Lokesh Sharma, et al. "Credit Card Fraud Detection Using Neural Network" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011

[4] Anshul Singh and Devesh Narayan , et al. "A Survey on Hidden Markov Model for Credit Card Fraud Detection" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012

[5] V.Dheepa and Dr. R.Dhanapal, et al. "Analysis of Credit Card Fraud Detection Methods", "International Journal of Recent Trends in Engineering", Vol 2, No. 3, November 2009

[6] Siva Parvatni Nelluri, Shaik Nagul, Dr. M.Kishore Kumar, et al"Credit card fraud detection using Hidden Markov Model (HMM), " International Journal of Engineering Research and technology(IJERT)" ISSN:2278-0181,Vol. 1 Issue 5, July 2012

[7] I-Cheng Yeh and Che-hui Lien , et al. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients", "Expert Systems with Applications" 2009.

[8] Hamdane, Balkis, et al. "Using the HMAC-Based One-Time Password Algorithm for TLS Authentication." *Network and Information Systems Security (SAR-SSI), 2011 Conference on*. IEEE, 2011.

[9] D. M'Raihi, M. Bellare, F. Hoornaert, and D. Naccache, et al "Hotp: An hmacbased one-time password algorithm," RFC 4226, Dec. 2005

Twinkle Patel
Master in computer Engineering, L J Institute of Engineering and Technology Ahmedabad, Gujarat, India

Ms. Ompriya Kale
M.Tech (CSE), Assistant Professor, L J Institute of Engineering and Technology Ahmedabad, Gujarat, India