Use of Hidden Markov Model as Internet Banking Fraud Detection

Sunil S Mhamane
M.E(C.S.E),Dept. Of Computer Science and Engg.
Walchand Institute of Technology
Solapur, India

L.M.R.J Lobo
Associate Professor, Dept of Computer Science and Engg.
Walchand Institute of Technology
Solapur, India

ABSTRACT
Now a day's Many Peoples are using internet banking for online Transaction we call it as E-commerce. As online transaction interest is increased associated with there are many frauds increasing such as using key logger, virus and worms to reveal internet banking account information such as password and ID. In this paper we explained about how Fraud is detected using Hidden Markov Model also care has been taken to prevent genuine Transaction should not be rejected by making use of one time password which is generated by server and sent to Personal Mobile of Customer. Hidden Markov Model is the statistical tools for engineer and scientists to solve various problems.

General Terms
Hidden Markov Model, Man in Browser Attack.

Keywords
Internet Banking, Hidden Markov Model, Probability, fraud detection, Transaction.

1. INTRODUCTION
Online banking (or Internet banking) allows customers to conduct financial transactions on a secure website operated by their retail or virtual bank, credit union or building society.

1.1 Online Banking Feature
Below Some Online banking feature are mentioned those features are application specific. The Common features fall broadly into several categories.

- Transactional:
  It includes following features
  1. Funds transfer between a two customer's account.
  2. Paying Third Parties.
  3. Investment purchase or sale.
  4. Loan applications
- Non-transactional
  1. Viewing Account Balance.
  2. Viewing Recent Transaction.
  3. Ordering cheque books.
- Support of multiple users having varying levels of authority
- Transaction approval process
- Wire transfer

To access internet banking, the customer would go to the financial institution's website, and enter the internet banking facility using the customer number and password. Some Financial institutions have set up additional security steps for access, but there is no consistency to the approach adopted.

1.2 Hidden Markov Model
A hidden Markov model (HMM) is a statistical model in which the system being modeled is assumed to be a Markov process with unobserved state. In our case type of purchase is modeled to different state. A HMM can be considered as the simplest dynamic Bayesian network. Difference between regular Markov model and Hidden Markov Model is the state is directly visible to the observer in case of Regular Markov Model while it is absent in case of HMM. Therefore the state transition probabilities are the only parameters in Regular Markov Model. Each state has a probability distribution over the possible output tokens. In our case possible output tokens are Low, Medium, High. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. Even if the model parameters are known exactly, the model is still 'hidden'. Hidden Markov models are used for their application pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics[1].

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other[2].

2. LITERATURE REVIEW
In "Credit Card Fraud Detection Using HMM" paper, They have proposed an use of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. They have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. They have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the spending profile of the cardholders. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions [3]

In "credit card fraud detection with a neural network" paper, Using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of

labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards**,** stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20***)*** over rule based fraud detection procedures. They discuss the performance of the network on this data set in terms of detection accuracy and earliness of fraud detection.[4].

In "Offline Internet Banking Fraud Detection" paper .Object of this paper is to demonstrate one successful fraud detection model which is established in Greece. Apart from the offline internet banking fraud detection system itself, which is described briefly, there scope is to present its contribution in fast and reliable detection of any "strange" transaction including fraudulent ones[5].

In "Security Analysis for Internet Banking Models" paper They stated that Internet banking fraud can be performed internally by genuine staff or externally by customers or suppliers. This paper presents a security analysis of the proposed Internet banking model compared with that of the current existing models used in fraudulent Internet payments detection and prevention. Several modern models in preventing and detecting fraud are evolving and being applied to many banking systems. However, they have no effective detection mechanism to identify legitimate users and trace their unlawful activities. Also they are not secure enough to prevent fraudulent users from performing fraudulent transactions over the Internet. The proposed model facilitates Internet banking Fraud Detection and Prevention (FDP) by applying two new secure mechanisms, Dynamic Key Generation (DKG) and Group Key (GK) [6].

In "Study on Fraud Risk Prevention of Online Banks" paper .The paper is aimed, in the first hand, at giving a discussion on the fraud risks of online banking, introducing the current application situation of information sharing mechanism in respect of internet fraud outside China as well as the development of such concept in China. Then, a system is designed for sharing internet fraud information. The paper finally proposing that all the online banks should put more joint efforts in perfecting this mechanism for sake of international co operation [7].

In "Fraudulent Internet Banking Payments Prevention using Dynamic Key" In this paper, They have proposed an efficient new scheme which can prevent fraud by applying different security algorithms, generating and updating limited-use secret keys. It uses advanced authentication technologies and is well adapted to any possible future technology. Moreover, it does not rely on fixed values where hacking one secret will not compromise the whole system's security. The generation of each set of keys is based on dynamically generated preference keys. The higher number the transactions performed, the less chance the system has of being compromised. The practical usefulness of the technique has been demonstrated by applying it to Internet banking payment systems. The results show that our technique enhances their security considerably. It has been shown that the proposed technique is secure against key compromise. For future work, we aim to analyze the security of the system that applies the proposed technique. Moreover, we aim to apply the proposed technique to other kinds of internet applications, especially mobile commerce [8].

In the paper "Parallel Granular Neural Networks for Fast Credit Card Fraud Detection" .A parallel granular neural network **(**GNN**)** is developed to speed up data mining and knowledge discovery process for credit card fraud detection. The entire system was working on the Silicon Graphics Origin 2000, which is a shared memory multiprocessor system consisting of 24-CPU, 4G main memory, and 200GB hard-drive. In simulations, the parallel fuzzy neural network running on a 24-processor system is trained in parallel using training data sets, and then the trained parallel fuzzy neural network discovers fuzzy rules for future prediction. The higher the fraud detection error is, the greater the possibility of that transaction being actually fraudulent [9].

3. ARCHITECTURE OF PROPOSED SYSTEM

Architecture of Proposed system consists of following component.

A. Actual User:

Basically he is the authorized client who is having internet banking account in particular bank .He can do online Transaction with the help of Internet Banking account legally

B. Fraudulent User:

He is the unauthorized user who is not having legal Internet banking account in bank. who makes use of authorized users Internet banking account to do Transaction. Hence He is Fraudulent User. He obtains the password of Particular Customer By doing attacks that are mentioned in 1.4.

C. Bank Server:

Bank Sever is managed by Bank Manager who is responsible to add customers for internet banking account. All the processing of internet banking done here where Manger can change account status(Lock to Unlock and vice versa). It also records the Customer Transaction Pattern Using HMM algorithm. This is used to Detect pattern and to find Fraudulent Transaction. In case of Violation Bank server sends the one time password to the Mobile Number which is registered in the Bank Database for Particular Customer.

D. Bank Database:

It stores the Information about customers such as (Name, Contact Number, Email id, Account Number).It also Stores the previous Transaction information of Customer made also it records Sequence of Transaction and it is modeled through Hidden Markov Model. For about 10^{th} Transaction are recorded sequence and from 11^{th} Transaction HMM algorithm is running to find Fraudulent Transaction.

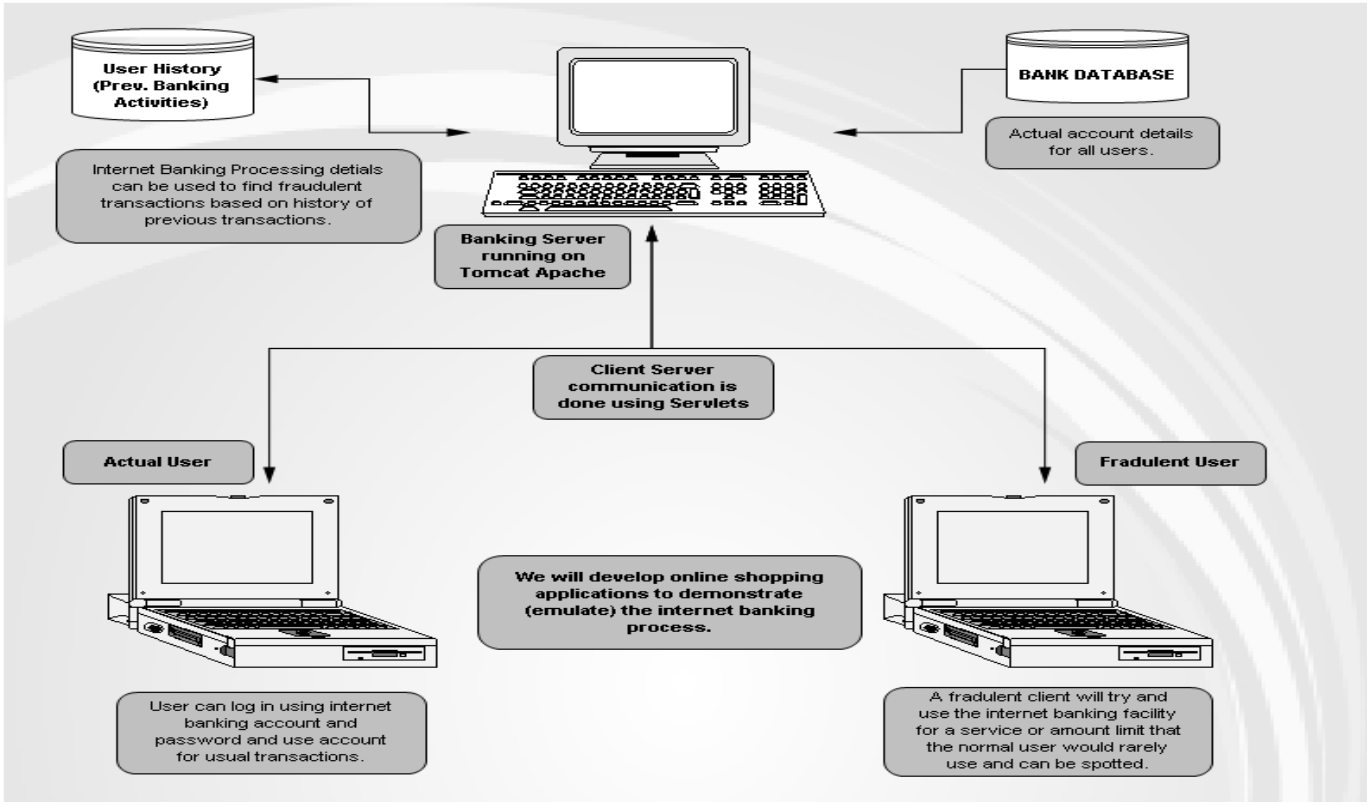

Fig 1: Architecture of Proposed System.

4. USE OF HMM TO DETECT FRAUD

4.1 Process Flow Diagram

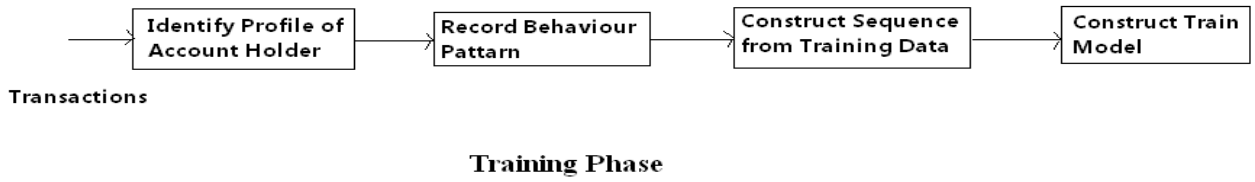

Fig 2: Training Phase of Process flow diagram.

The figure below illustrates about the two phases of the detection system used by HMM. In the training phase Train Model is created and based on the initial set of transactions Behavior profile of Internet Bank Holder is identified. This directs for expected transaction sequence for each account holder and the system is trained accordingly.

In the detection And Prevention phase(Fig 3) the system looks for the deviation in expected and actual outcome and fraud is recognized. When fraud is Recognized Bank server sends One time passwords to Mobile Number.

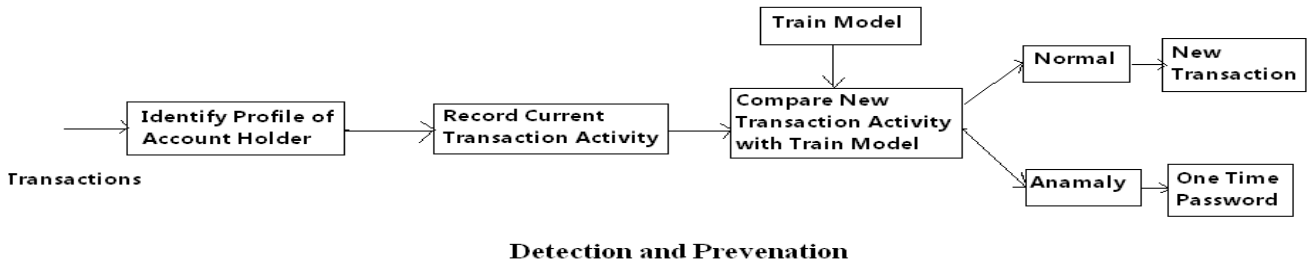

Fig 3: Detection and Prevention Phase of Process flow diagram.

4.2 HMM Model and Algorithm

4.2.1 HMM Model

Here in above Figure HMM Model for fraud detection is presented. Three Different kind of Purchases are shown they are represented as states of HMM TT(Travel Ticket),MT(Movie Ticket),BP(Book Purchase). V and NV are two Observation Symbols either of one will active for particular state they are shown on each state. V indicates Violation means if incoming transaction violates the Behavior sequence then V will be observed symbol to that state and OTP is sent to the Customers Mobile Number. NV indicates Non-Violation means there is no anomaly and incoming Transaction is Normal. No action is performed in this Observation. All others lines and curves indicates Transition from one state to another state. States will increase if sequence of number of Purchase made is increased.

Here Purchased Transaction amount is categorized into low, medium and high. if purchased amount is below 1000 then it is low. if in between 1000 and 2500 then it is medium and if more that 2500 it is high.

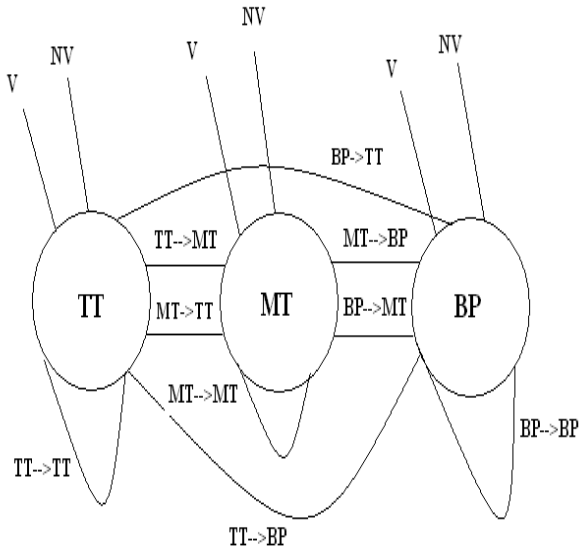

Fig 4: HMM Model for Fraud Detection.

4.2.2 HMM Algorithm

To use a HMM we need to calculate the HMM Parameters such as state and transition probabilities. Those parameters are calculated using Baum-Welch algorithm [10].Baum and Welch algorithm is given below.

Baum-Welch Algorithm::

1. Initialize the parameters (state, transition) to some values
2. Calculate "forward-backward" probabilities based on the current parameters
   - Forward probability is calculated as below.
     a. At time t, the probability that we're in state i.
     b. the previous observation thus far has been o1 … ot.
     c. $\alpha(i) = P(i,o1....ot/ \lambda)$
   - Backward probability
     a. At time t and we're in state i, the probability that
     b. the observation that follows will be ot+1 … oT.
     c. $\beta(i)= P(Ot+1....OT/i, \lambda)$.
3. Use the forward-backward probabilities to estimate the expected frequencies
   - Expected number of transitions from state i
   - Expected number of being in state j.
4. Use the expected frequencies to estimate the parameters.
5. Repeat 2 to 4 until the parameters converge.

5. RESULTS AND COMPARISION WITH PREVIOUS METHODS

5.1 Results

In Table 1 information about purchased item of customer is recorded. These details are sent from client to server for processing. In sever these details are stored on user history behaviour.HMM Algorithm is applied on these transaction. Here for first 8 transaction training phase is applied as shown in fig 2.from these training data sequence is constructed .in table 1 after M(Medium) there is L(low) this sequence is repeated for 3 times.HMM algorithm starts working from 9^{th} transaction. After L(low) there is M(Medium) 2 times and H(High) one time only. My 8^{th} transaction is L(Low) and it expects 9^{th} transaction to be M (Medium).if 9th transaction is medium then it is case of Non-Violation(NV) it will not ask for One time password(OTP) because user is doing in sequence. if 9^{th} transaction is either H(High) or M(Medium) then it is case of violation(V). It will ask for one time password which is sent to Customer Mobile number.

Table 1: Customer Purchase Information

Transaction Number	Type of Purchase	Amount	Category	OTP
1	Apparels	1000	M	YES
2	Education	700	L	YES
3	Entertainment	1000	M	YES
4	Food	100	L	YES
5	Travel	2500	H	YES
6	Utility	100	L	YES
7	Jewelry	1500	M	YES
8	Monthly Payments	100	L	YES
9	Entertainment	1000	M	NO
10	Travel	2500	H	YES

5.2 Comparison with Previous Methods

5.2.1 "Credit Card Fraud Detection Using HMM" Paper.

1. In this paper HMM is applied on credit card Transaction but In our paper which is applied on Internet Banking Transaction which is different from Credit card Transaction.
2. In this paper fraud detection was made but there was no arrangements made if genuine customer did transaction in different behavior. Alarm will rang at server side but it is no use of genuine transaction but in our paper if there is case of violation One time password is sent to Customers Mobile number so that genuine transaction will be successful. Mobile is personal so really if he is genuine user he will enter OTP. This is very important improvement in our paper.

5.3 Efficiency of Method

Testing internet banking fraud detection system using real data set is a difficult task. Banks do not, in general, agree to share their data with researchers.

We created the 16 customers account database and given to 16 persons. for first 8 transaction they did the transactions according to normal behavior. After all ones 8 transaction all the 16 persons had given other persons login information. Now from 8th transaction all persons are doing Fraud up to 5 transactions. We saw how many of them are getting OTP that we counted. If they are not getting OTP then transaction will be successful however person is doing fraud. Following table shows the efficiency of system.

Table 2: OTP Measurement.

Persons	OTP ASKESD (out of 5)
Person1	3
Person2	4
Person3	2
Person4	4
Person5	4
Person6	3
Person7	3
Person8	5
Person9	3
Person10	4
Person11	5
Person12	3
Person13	4
Person14	4
Person15	3
Person16	3
Total	**57**

$$\text{Efficiency of System} = \frac{\text{Total no of OTP ASKED}}{\text{Total No of Fraud Transaction.}} * 100$$

$$\text{Efficiency of System} = \frac{57}{80.} * 100$$

Efficiency of System = 71.25%.

Above Efficiency shows that our system is working 71.25% correctly.

Fig 5 indicates the True Detection and False detection series. The line graph values are taken from observation made from Table 2. Where true detection indicates no of transaction for which OTP is asked and False detection indicates no of transactions where OTP is not asked where transaction is successful in case of fraud.

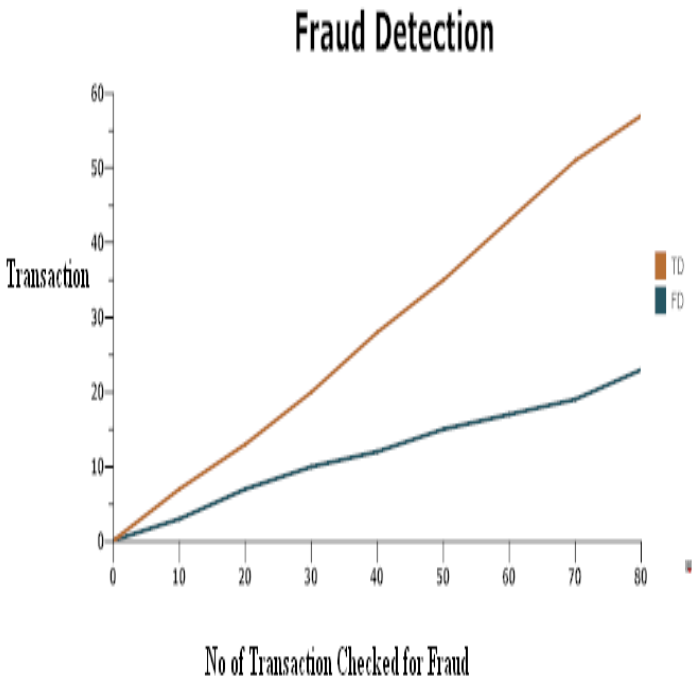

Fig 5: Line Graph for Fraud Detection.

6. CONCLUSION

The proposed methodology is aimed at detecting fraud in case of internet banking. In Internet Banking a Fraud detection system will run at the banks server and its Function to detect fraud in online transaction. This is a Prediction system. Fraud detection is carried out using Hidden Markov Model which uses baum-welch algorithm. Initially After detecting a fraud it sends a Onetime password to Mobile number. Here prices are divided into three ranges. Low, Medium, and High. We model the sequence of operations in online banking transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of an account holder. If an incoming online banking transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, we ensure that genuine transactions are not rejected. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of Hidden Markov Model. Comparative studies reveal that the Accuracy of the system is close to 72 percent over a wide variation in the input data and results are good as compared to previous papers. The system is also scalable for handling large volumes of transactions.

7. REFERENCES

[1] Hidden Markov Model by Jia Li. Department of Statistics "The Pennsylvania State University" http://www.stat.psu.edu/~jiali/course/stat597e/notes2/hmm.pdf.

[2] "A Revealing Introduction to Hidden Markov Models" by mark stamp.

[3] "Credit Card Fraud Detection Using Hidden Markov Model" By Abhinav Srivastava, Amlan Kundu, Shamik Sural. *IEEE Transaction, January-March 2008.*

[4] "credit card fraud detection with a neural network" by Ghosh and Reilly. *IEEE" Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994.*

[5] "Offline Internet Banking Fraud Detection" by Vasilis Aggelis.

[6] "Security Analysis for Internet Banking Models" By Osama Dandash, Phu Dung Le and Bala Srinivasan. *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing IEEE DOI 10.1109/SNPD.2007.5321142*

[7] "Study on Fraud Risk Prevention of Online Banks" By Qinghua Zhang. *2009 International Conference on Networks Security, Wireless Communications and Trusted Computing.*

[8] "Fraudulent Internet Banking Payments Prevention using Dynamic Key" By Osama Dandash Yiling Wang and Phu Dung Leand Bala Srinivasan. *"JOURNAL OF NETWORKS, VOL. 3, NO. 1, JANUARY 2008".*

[9] Parallel Granular Neural Networks for Fast Credit Card Fraud Detection Mubeena Syeda, Yan-Qing Zbang and Yi Pan. *IEEE Transaction.*

[10] "Hidden Markov Model and Baum-Welch algorithm" by Lloyd R. *Welch IEEE Information Theory Society Newsletter Vol. 53, No.4, December 2003*