Markova Scheme for Credit Card Fraud Detection

B.Sanjaya Gandhi , sanjaygandhi.b@gmail.com, R.Lalu Naik, rlalunaik@yahoo.com, S.Gopi Krishna, gks24@rediffmail.com, K.lakshminadh l_nadh@yahoo.co.in

Abstract*— as credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In this paper, we model the sequence of operations in credit card transaction processing using a Hidden Markov Model and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behavior of a cardholder. If an Incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected.

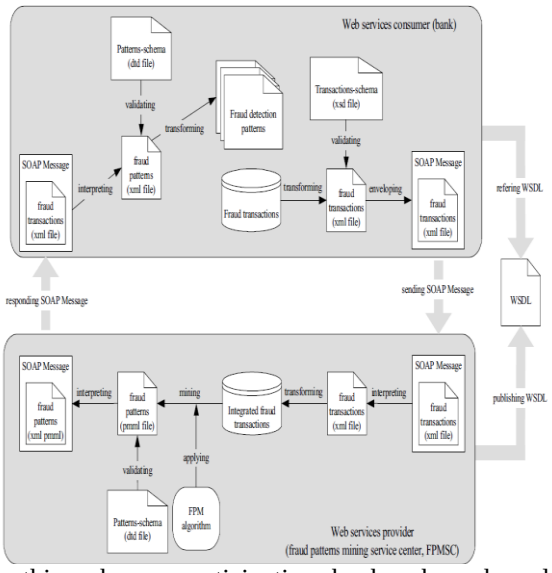
INDEX TERMS: ONLINE SHOPPING, CREDIT CARD, SECURITY, FRAUD FINDING, HMM.

I. INTRODUCTION

Credit-card-based purchases can be categorized into two types: 1) physical card and 2) virtual card. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In the second kind of purchase, only some important information about a card (card number, expiration date, secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behaviouristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc

II.RELATED WORKS ON CREDIT CARD FRAUD DETECTION

Credit card fraud detection has drawn a lot of research interest and a number of techniques. Gosh and Reilly have proposed credit card fraud detection with a neural network.



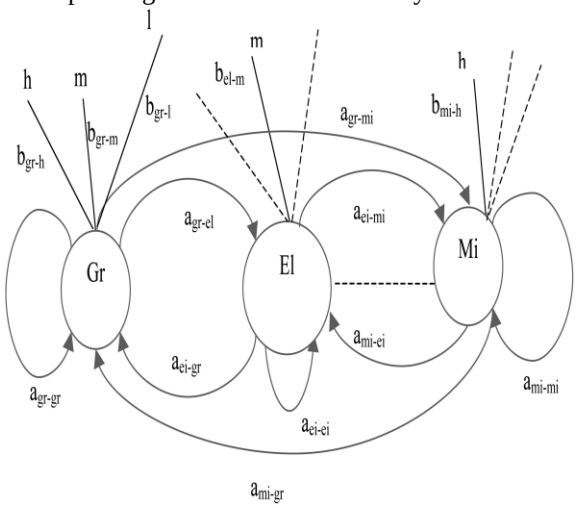
With this scheme, participating banks share knowledge about the fraud patterns in a heterogeneous and distributed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. Phua et al. have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report .In contrast, we present a Hidden Markov Model (HMM)-based credit card FDS, which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. We model a credit card transaction processing sequence by the stochastic process of an HMM. The details of items purchased in individual transactions are usually not known to an FDS running at the bank that issues credit cards to the cardholders. This can be represented as the underlying finite Markov chain, which is not observable. The transactions can only be observed through the other stochastic process that produces the sequence of the amount of money spent in each transaction. Hence, we feel that HMM is an ideal choice for addressing this problem. Another important advantage of the HMM-based approach is a drastic reduction in the number of False Positives (FPs)—transactions identified as malicious by an FDS although they are actually genuine.

III. HMM MODEL

To map the credit card transaction processing operation in terms of an HMM, we start by first deciding the observation symbols in our model. We quantize the purchase values x into M price ranges V1; V2; . . . VM, forming the observation symbols at the issuing bank. The actual price

SEEK DIGITAL LIBRARY

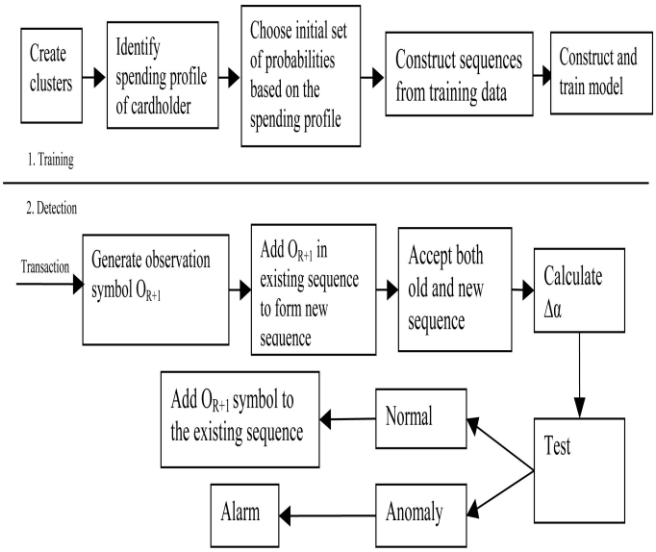
range for each symbol is configurable based on the spending habit of individual cardholders. These price ranges can be determined dynamically by applying a clustering algorithm on the values of each cardholder's transactions. In this work, we consider only three price ranges, namely, low (l), medium (m), and high (h). For example, let l= (0, $100], m = ($100, $500], and h= ($500, credit card limit]. If a cardholder performs a transaction of $190, then the corresponding observation symbol is m.



A credit cardholder makes different kinds of purchases of different amounts over a period of time. One possibility is to consider the sequence of transaction amounts and look for deviations in them. However, the sequence of types of purchase is more stable compared to the sequence of transaction amounts. The reason is that, a cardholder makes purchases depending on his need for procuring different types of items over a period of time. This, in turn, generates a sequence of transaction amounts. Each individual transaction amount usually depends on the corresponding type of purchase. Hence, we consider the transition in the type of purchase as state transition in our model. The type of each purchase is linked to the line of business of the corresponding merchant. This information about the merchant's line of business is not known to the issuing bank running the FDS. Thus, the type of purchase of the cardholder is hidden from the FDS. The set of all possible types of purchase and, equivalently, the set of all possible lines of business of merchants forms the set of hidden states of the HMM. It should be noted at this stage that the line of business of the merchant is known to the acquiring bank, since this information is furnished at the time of registration of a merchant. Also, some merchants may be dealing in various types of commodities (For example, Wal-Mart, K-Mart, or Target sells tens of thousands of different items). Such types of line of business are considered as Miscellaneous, and we do not attempt to determine the actual types of items purchased in these transactions. Any assumption about availability of this information with the issuing bank and, hence, with the FDS, is not practical and, therefore, would not have been valid.

IV. FRAUD FINDING

An FDS runs at a credit card issuing bank. Each incoming transaction is submitted to the FDS for verification. FDS receives the card details and the value of purchase to verify whether the transaction is genuine or not. The types of goods that are bought in that transaction are not known to the FDS. It tries to find any anomaly in the transaction based on the spending profile of the cardholder, shipping address, and billing address, etc. If the FDS confirms the transaction to be malicious, it raises an alarm, and the issuing bank declines the transaction. The concerned cardholder may then be contacted and alerted about the possibility that the card is compromised. In this section, we explain how HMM can be used for credit card fraud detection.



After the HMM parameters are learned, we take the symbols from a cardholder's training data and form an initial sequence of symbols. Let o1, o2----or be one such sequence of length R. This recorded sequence is formed from the cardholder's transactions up to time t. We input this sequence to the HMM and compute the probability of acceptance by the HMM. Let the probability be $\alpha 1$, which can be written as follows:

$\alpha 1$=P (o1, o2, ----or)

Let or+1 be the symbol generated by a new transaction at time t+1. To form another sequence of length R, we drop O1 and append or+1 in that sequence, generating o2, o3----or+1as the new sequence. We input this new sequence to the HMM and calculate the probability of acceptance by the HMM. Let the new probability,

$\alpha 2$=P (o2, o3, ----or+1)

Let $\Delta \alpha = \alpha 1 - \alpha 2$ .If $\Delta \alpha > 0$, it means that the new sequence is accepted by the HMM with low probability, and it could be a fraud. The newly added

SEEK DIGITAL LIBRARY

transaction is determined to be fraudulent, otherwise the transaction is genuine.

V. EXPERIMENTAL RESULTS

We show performance of the proposed system as we vary the number of fraudulent transactions and also the spending profile of the cardholder. We compare performance of our approach (denoted by OA below) with the credit card fraud detection technique proposed by Stolfo etc. (denoted by ST below). For comparison, we consider the metrics
TP and FP, as well as TP-FP and Accuracy Here, (a ,b ,c) profile represents a ls profile cardholder who has been found to carry out a percent of his transactions in the low, b percent in medium, and c percent in the high range. It may be noted that for cardholders in the other two groups, namely, hs and ms, will show similar performance as only the relative ordering of a, b, and c will change. We also vary the mean value μ of malicious transactions. Thus, every sequence of transaction that we use for testing is a mixed sequence containing both genuine, as well as malicious, transactions. For each combination of spending profile and malicious transaction distribution, we carried out 100 runs. And report the average result. The same set of data was used to determine the performance of both OA and ST. Fig. 5a shows the variation of TP and FP for the two approaches using the spending profile (95 3 2). Variation of TP-FP and Accuracy is shown in Fig. 5b. It is seen from the figures that TP of the proposed approach is very close to that of Stolfo et al's approach. Also, both the approaches have almost similar values of FP. As a result, the two systems have comparable accuracies and average TP-FP spread. Further, the two approaches exhibit similar trend with variation in _. Next, we show how the two systems behave as we mix the transaction amounts. Percentages of low value, medium value, and high value transactions are changed from (95 3 2), as shown above to (70 20 10) in Figs. 6a and 6b and the advantages are,
1) The detection of the fraud use of the card is found much faster that the existing system.
2) In case of the existing system even the original card holder is also checked for fraud detection. But in this system no need to check the original user as we maintain a log.
3) The log which is maintained will also be a proof for the bank for the transaction made.
4) We can find the most accurate detection using this technique.
5) This reduce the tedious work of an employee in the bank

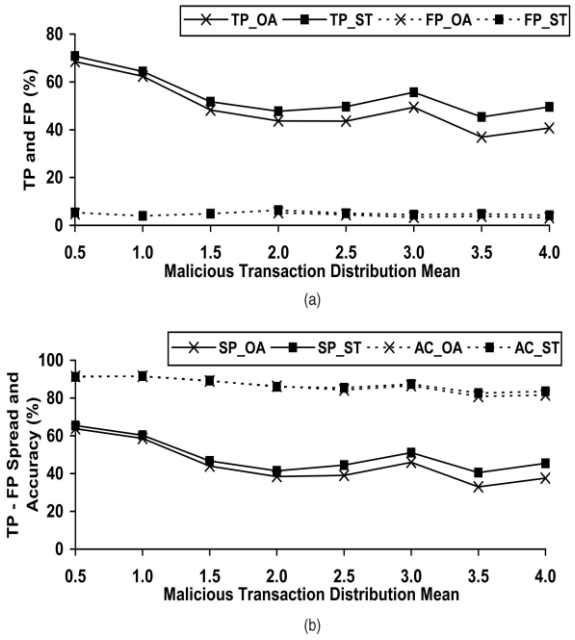


Fig 5(a) Performance variation of the two systems (OA) and ST for the spending profile (95 3 2). (a) TP and FP. (b) TP-FP spread (SP) and Accuracy (AC).

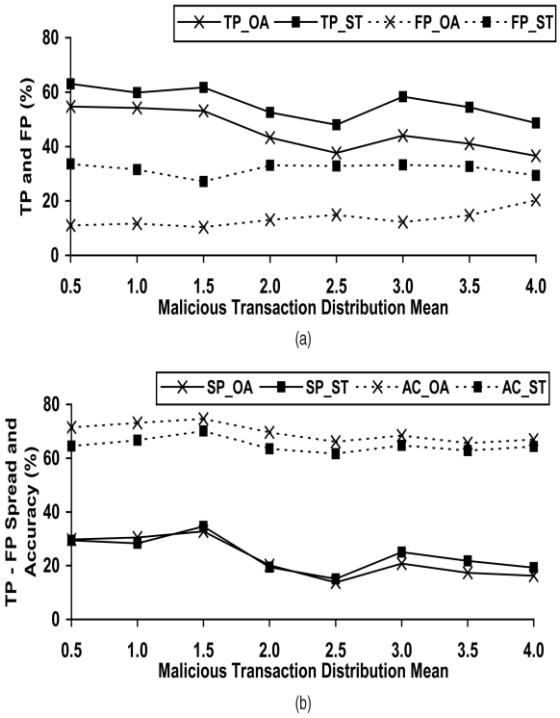


Fig 6 Performance variation of the two systems (OA and ST) for the spending profile (70 20 10). (a) TP and FP. (b) TP-FP spread (SP) and Accuracy (AC).

SEEK DIGITAL LIBRARY

VI. CONCLUSION

The different steps in credit card transaction processing are represented as the underlying Stochastic process of an HMM. We 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. We 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.

REFERENCES

[1] "Global Consumer Attitude towards On-Line Shopping,"http://www2.acnielsen.com/reports/documents/2005_cc_onlineshopping.pdf

[2] "Statistics for General and On-Line Card Fraud," http://www.epaynews.com/statistics/fraud.html.

[3] S. Gosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems: Decision Support and Knowledge-Based Systems

[4] M. Syeda, Y.Q. Zhang, and Y. Pan, "Parallel Granular Networks for Fast Credit Card Fraud Detection"

[5] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit Card Fraud Detection Using Meta-Learning"

[6] Simple Object Access Protocol, http://www.w3.org/tr/soap
[7] Web Services Description Language, http://www.w3.org/tr/wsdl

[8] C. Chiu and C. Tsai,"A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e-Service, pp. 177-181, 2004.

SEEK DIGITAL LIBRARY