Anomaly Detection Approach Using Hidden Markov Model

[1,]Sonali N.Jadhav, [2,]Kiran Bhandari
[1]*M.E Scholar,* [1,2]*,Assitant Professor, Thakur College Of Engineering And Technology, Mumbai, India*

ABSTRACT:

This paper proposes an empirical method of Anomaly detection by analyzing the spending habit of vendee. Proposed system models the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and shows how it can be used for the detection of frauds. In the existing credit card fraud detection business processing system, fraudulent transaction will be detected after transaction is done. It is difficult to find out fraudulent and regarding loses will be barred by issuing authorities. Hidden Markov Model is the statistical tools for engineer and scientists to solve various problems. It is shown that credit card fraud can be detected using Hidden Markov Model during transactions. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate. Does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. 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. Hence HMM is an ideal choice for addressing this problem. The objective of the system is to detect the Anomaly during the transaction only and confirm the fraud by asking some security code. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate.

KEYWORDS: *Anomaly detection, Hidden Markov Model, Online Transactions, Data Mining, Clustering,Probability,Security.*

I. INTRODUCTION

In day to day life credit cards are used for purchasing goods and services with the help of virtual card for online transaction or physical card for offline transaction. In physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carryout 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 online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.).In this purchase method, mainly transactions will be done through Internet or 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 analyze 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 [1]. Since humans tend to exhibit specific behaviorist 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. Deviation from such patterns is considered as fraud.

A Hidden Markov Model [5] is a finite set of states; each state is linked with a probability distribution. Transitions among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are ``hidden" to the outside; hence the name Hidden Markov Model. Hence, Hidden Markov Model is a perfect solution for addressing detection of fraud transaction through credit card. One more important benefit of the HMM-based approach is an extreme decrease in the number of False Positives transactions recognized as malicious by a fraud detection system even though they are really genuine.

In this prediction process, HMM consider mainly three price value ranges such as.1) Low (l),2) Medium (m) and,3) High (h).First, it will be required to find out transaction amount belongs to a particular category either it will be in low, medium, or high ranges.Initially the HMM is trained with the normal behavior of a card holder then spending patterns of user can be determined with the help of K-means clustering algorithm. If an incoming transaction is not accepted by the HMM with sufficient probability then it can be detected as fraud for further confirmation security question module will be activated that contains some personal questions that are only known to authorized customer and if the transaction is fraudulent then verification code is asked for further confirmation. Hidden Markov model works on Markov chain property in which probability of each subsequent state depends on the previous state, which consists of observation probabilities, transition probabilities and initial probabilities. 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 Marko process rather than independent of each other.

II. RELATED WORK

Ghosh and Reilly [3] have proposed credit card fraud detection with a neural network. They have built detection system, which is trained on a large sample of labeled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, andnonreceived issue (NRI) fraud.Stolfo et al [4] suggest a credit card fraud detection system (FDS)using meta-learning techniques to learn models of fraudulent credit card transactions. Metalearningis a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A meta-classifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection [11]. They use Java agents for Meta-learning (JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like TP-FP (True Positive – False Positive) spread and accuracy have been defined by them.Aleskerov et al. [12] present CARDWATCH, database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Fan et al. [13] suggest the application of distributed data mining in credit card fraud detection. Braise et al. Stolfo. use an agent-based approach with distributed learning for detecting frauds in credit card transactions. It is based on artificial intelligence and combines inductive learning algorithms and metal earning methods for achieving higher accuracy. Phua et al. [14] suggest the use of met classifier similar to in fraud detection problems.

They consider naive Bayesian, and Back Propagation neural networks as the base classifiers. A met classifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use credit card fraud detection as the target application, their approach is quite generic.Vatsa et al.[15] have recently proposed a game-theoretic approach to credit card fraud detection. They model the interaction between an attacker and an FDS as a multi stage game between two players, each trying to maximize his HMM-based applications are common in various Areas such as speech recognition, bioinformatics, and Genomics. In recent years, Joshi and Phoba [16] have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. Cho and Park [17] suggest an HMM-based intrusion detection system that improves the modeling time and performance by considering only the privilege transition flows based on the domain knowledge of attacks. Ours ton etal. [18] have proposed the application of HMM in detecting multistage network attacks. Hoang et al. present a new method to process sequences of system calls for anomaly detection using HMM. The key idea is to build a multilayer model of program behaviors based on both HMMs and enumerating methods for anomaly detection. Lane has used HMM to model human behavior. Once human behaviorism correctly modeled, any detected deviation is a cause for concern since an attacker is not expected to have a behavior similar to the genuine user. Hence, an alarm is raised in case of any deviation.

III. PROPOSED WORK

3.1 Problem Definition and Outline

In proposed system, by using Hidden Markov Model (HMM) which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. 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. Hence HMM is an ideal choice for addressing this problem. To complete the transaction Vendee should answer the security questions. Fraud is confirmed by asking some security code which is sent by email transaction proceed only

when verification code is correct otherwise transaction cancelled. Fraud is detected using the probability difference that is in between old observation sequence and new observation sequence.

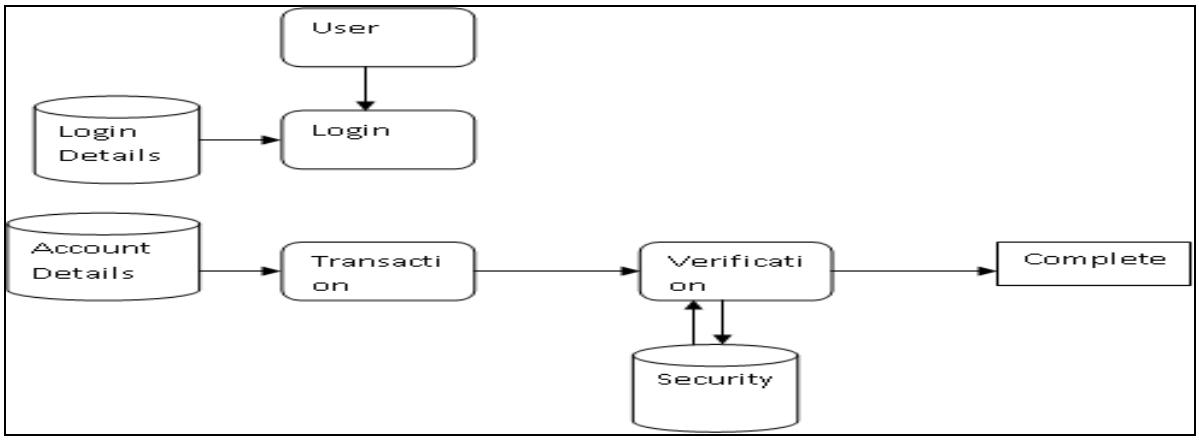


Fig.1: Outline of the proposed method

3.2. Hidden Markov Model

HMM is basically a model consisting of sequence of states that works on Markov chain property [5]. Name Hidden here indicates that observer does not know in which state it is but having a probabilistic insight on where it should be. Input to HMM is observation sequence and output is probability [6] of a sequence. 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

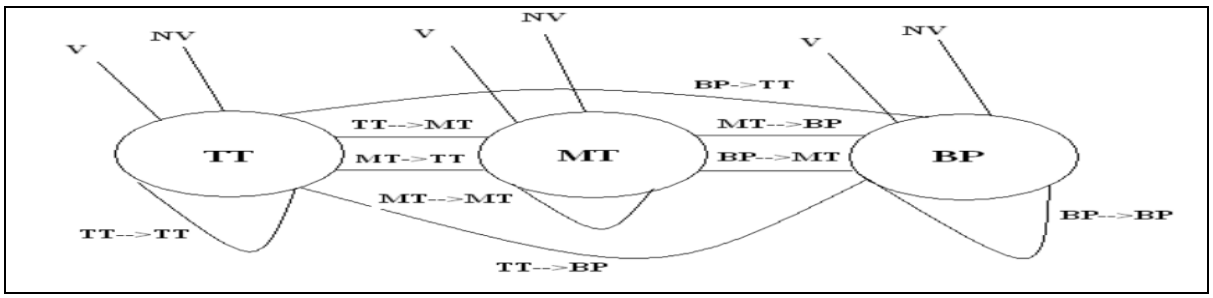


Fig 2: HMM model for fraud detection

Here in above Figure 2 HMM Model for fraud detection is considered [8]. 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 on-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. to use HMM we need to calculate the HMM parameters such as state and transition probabilities. Those parameters are calculated using Baum-Welch algorithm [9].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
3. Use the forward-backward probabilities to estimate the Expected frequencies
4. Use the expected frequencies to estimate the parameters.
5. Repeat 2 to 4 until the parameters converge.

3.3. K-Means Clustering Algorithm

By using K-MEANS clustering algorithm which divides the spending profile of a user into low medium and high cluster and accordingly generates observation symbols that are further given to HMM for training as well as detection purpose K-means clustering algorithm first divides the transaction amount into different clusters. Consider example 10 transactions,)

Tabel 1.Transaction amount

Transaction	1	2	3	4	5	6	7	8	9	10
Amount	40	25	15	5	10	25	15	20	10	80

Table 2. Output of K-means clustering

Cluster mean/centroid name	Cl	Cm	Ch
Observation symbol	V1=l	V2=m	V3=h
Mean value	8.3	20	60
Percentage of total transactions	30	50	20

By applying K-Means clustering algorithm on above transactions output is shown in table 2.Considering M= 3, if we execute K-means algorithm on the example transactions in Table 3.2, we get the clusters, as shown in Table 3.3, with cl, cm, and ch as the respective centroids. It may be noted that the dollar amounts 5, 10, and 10 have been clustered together as cl resulting in a centroid of 8.3. The percentage (p) of total number of transactions in this cluster is thus 30 percent. Similarly, dollar amounts 15, 15, 20, 25, and 25 have been grouped in the cluster cm with centroid 20, whereas amounts 40 and 80 have been grouped together in cluster ch. cm and ch, thus, contain 50 percent and 20 percent of the total number of transactions. When the FDS receives a transaction T for this cardholder, it measures the distance of the purchase amount x with respect to the means cl, cm, and ch to decide the cluster to which T belongs and, hence, the corresponding observation symbol. As an example, if x= $10, then in Table 2 the observation symbol is V1= l.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results include identifying the spending habit of customer and detect the fraud by analyzingthe spending habit of customer and then precede the transaction. Implementation is done using java and MySQL. Some of the snapshots for the implementation given below:

Step1: Vendee will select the product from list and add it to cart.initialy cardholder will select the product and add it to the cart.

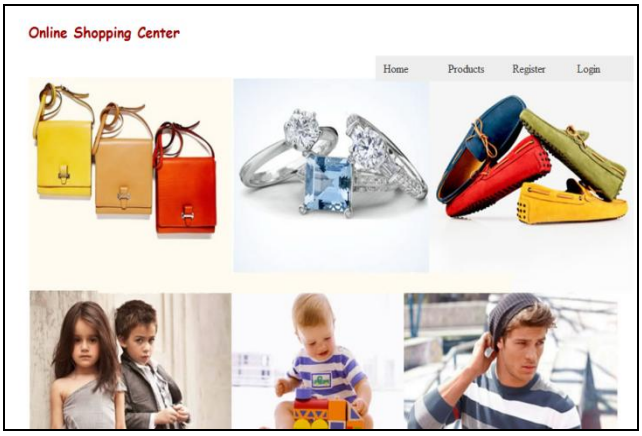


Fig 3: Selection of product

Fig 4: Purchase the product

Step2: Enter Details after selecting the product.

Fig 4: Card Details

Fig 5: States of HMM

As shown in figure 4 vendee will enter the Details when he submit the details using k-means clustering algorithm observation sequence is generated Low Medium High further Hidden Markov Model starts calculating the spending habit of customer figure 5 shows states of Hidden Markov Model.

Step3: comparing previous observation sequence with existing and calculates the probability difference if it is >0 then fraud is detected and mail sent to Vendee for confirmation shown in figure 7, else the transaction is completed.

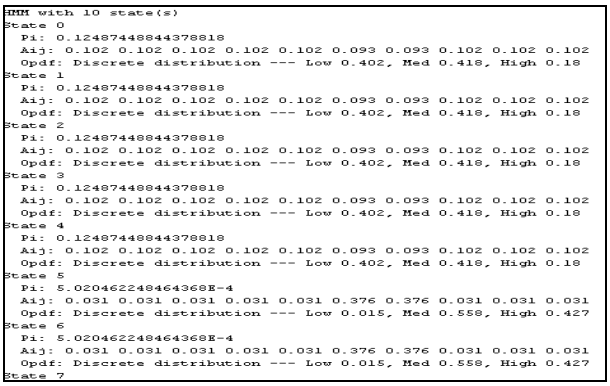


Fig 6: Probability sequence generated by HMM

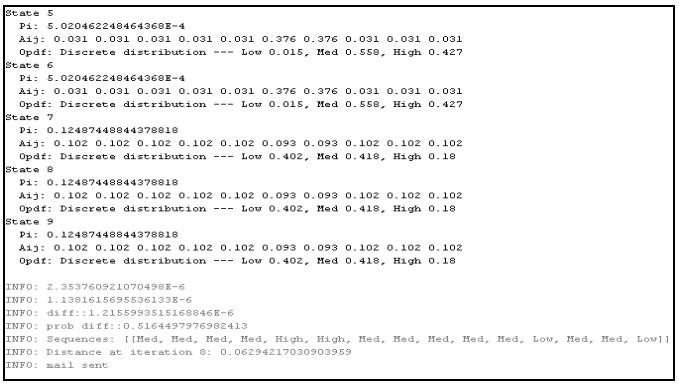
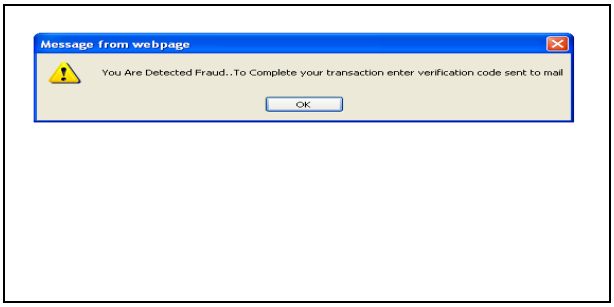


Fig 7: Anomaly Detection by HMM

Step4: Fraud detection confirmed by asking verification code if user will enter the correct verification code then transaction is completed else fraud is confirmed.



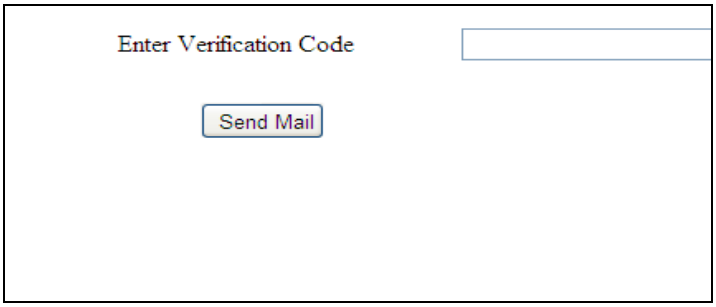


Fig 8: Fraud detection and confirmation.

Table 3. List of all the most recent transaction

Transaction	1	2	3	4	5	6	7	8	9	10
Amount	140	125	15	5	10	125	15	120	10	280
Transaction	11	12	13	14	15	16	17	18	19	20
Amount	210	550	800	110	35	118	20	148	141	6

In Table 3, list of all the most recent transaction is placed at the first position and correspondingly first transaction is placed at the last position in the table. The pattern of spending profile of the card holder is shown in Figure 9 based on all transactions done. Transactions are shown.

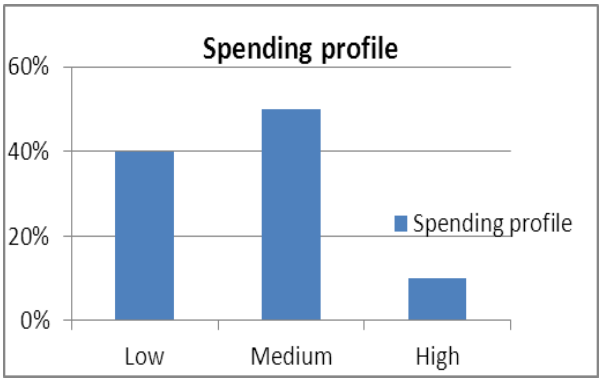


Fig. 9: Spending profile of all transaction

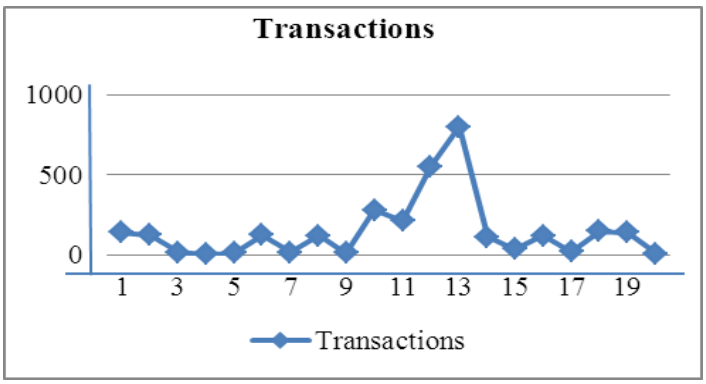


fig.10 Percentage of each spending profile

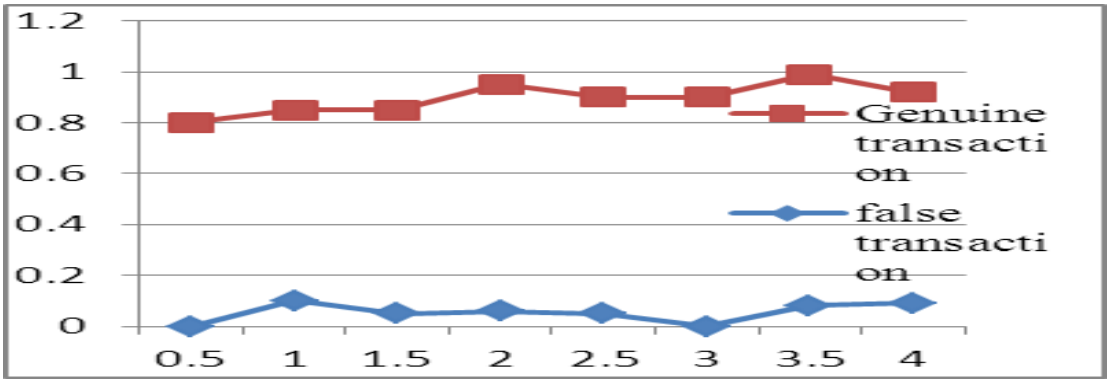


Fig:11 Probability of False Alarm compared with Fraud Transaction Mean Distribution

Fraud detection mean distribution is shown in Figure 11, where probability of false transaction compared with that of genuine transaction. Hence, it is expected that when the probability of false alarm will be more than threshold probability, then it should generate an alarm for fraudulent and also decline the transaction.

V. CONCLUSION.

We proposed system which is an application of HMM in Anomaly Detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. The ranges of transaction amount can be used as the observation symbols, whereas the types of item have been considered to be states of the HMM. Also proposed system suggests 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 them can detect whether an incoming transaction is fraudulent or not. The system is also scalable for handling large volumes of transactions. The proposed method can be enhanced to achieve more accuracy and better algorithms for clustering.

REFERENCES

[1] "Study on Fraud Risk Prevention of Online Banks" By Qinghua Zhang. 2010 International Conference onNetworks Security, Wireless Communications and Trusted Computing.

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

[3] Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with aNeural-Network, 27th Hawaii International l Conference on InformationSystems, vol. 3 (2003), pp. 621- 630.

[4] Stolfo, S. J., Fan, D. W., Lee, W., Prodromidis, A., and Chan, P. K., 2000.Cost-Based Modeling for Fraud and Intrusion Detection: Results from theJAM Project, Proceedings of DARPA Information SurvivabilityConference and Exposition, vol. 2 (2000), pp. 130-144.

[5] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in SpeechRecognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[6] K. S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Second Edition, John Wiley and Sons, New York, 2001.

[7] Iyer, Divya.; Mohanpurkar, Arti; Janardhan, Sneha; Rathod, Dhanashree; Sardeshmukh, Amruta," Credit Card Fraud Detection" Proceedings of the IEEE, pp. 1062-1066, 2011.

[8] Jiawei Han and Michelin Kamber,"Data mining concepts and Techniques",Elsevler publication second edition,pp 598.

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

[10] S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows using Hidden Markov Model," Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.

[11] R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining for CreditCard Fraud Detection," roc. IEEE Int'l Conf. Tools with ArtificialIntelligence, pp. 103-106, 1999.

[12] Aleskerov, E., Freisleben, B., and Rao, B., 1997. CARDWATCH: ANeural Network Based Database Mining System for Credit Card FraudDetection, Proceedings of IEEE/IAFE: Computational Intelligence forFinancial Eng. (1997), pp. 220-226.

[13] W. Fan, A.L. Prodromidis, and S.J. Stolfo, "Distributed Data Mining inCredit Card Fraud Detection," IEEE Intelligent Systems, vol. 14, no. 6,pp. 67-74, 1999.

[14] C. Phua, D. Alahakoon, and V. Lee, "Minority Report in FraudDetection: Classification of Skewed Data," ACM SIGKDD ExplorationsNewsletter, vol. 6, no. 1, pp. 50-59, 2004.

[15] V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-eoretic Approach toCredit Card Fraud Detection," Proc. First Int'l Conf. Information SystemsSecurity, pp. 263-276, 2005

[16] S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov ModelsCapabilities in Anomaly Detection," Proc. 43rd ACM Ann. SoutheastRegional Conf., vol. 1, pp. 98-103, 2005.

[17] S.B. Cho and H.J. Park, "Efficient Anomaly Detection by ModelingPrivilege Flows Using Hidden Markov Model," Computer and Security,vol. 22, no. 1, pp. 45-55, 2003.

[18] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications ofHidden Markov Models to Detecting Multi-Stage Network Attacks,"Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344,2003.