ISSN 2394-739X
International Journal for Research in Science Engineering and Technology
Approved by National Science Library (NSL),
National Institute of Science Communication And Information Resources (NISCAIR),
Council of Scientific and Industrial Research (CSIR), New Delhi, India .

# Fraud Detection in Credit Card Using DataMining Techniques

Mr.P.Matheswaran[1],Mrs.E.Siva Sankari ME[2],Mr.R.Rajesh[3]
[1]P.G. Student, Department of CSE, Govt.College of Engineering, Thirunelveli, India.
[2]Assistant Professor, Department of CSE, Govt.College of Engineering, Thirunelveli, India.
[3]P.G. Student, Department of CSE, Govt.College of Engineering, Thirunelveli, India.

**Abstract**

**Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card ordebit card, as a fraudulent source of funds in a transaction. The purpose may be to obtain goods without paying, or to obtain unauthorized funds from an account. The most accept mode is credit card for both online and offline in today world. It provide cashless shopping at every shop in all countries. So as credit card is becoming popular mode for online financial transactions, at the same time fraud associated with it are also rising. In this paper Hidden Markov Model (HMM) is used to model the sequence of operation in credit card transaction processing. HMM is trained using Baum-Welch algorithm with normal behaviour of cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent.**

***Key Words*—Hidden Markov Model, credit card, fraud detection, Hidden Markov Model, Internet, online shopping.**

## I. INTRODUCTION

The internet becomes most popular mode of payment for online transaction. Banking system provides e-cash, e commerce and e-services improving for online transaction. Credit card is one of the most conventional ways of online transaction. In case of risk of fraud transaction using credit card has also been increasing. Credit card fraud detection is one of the ethical issues in the credit card companies, mortgage companies, banks and financial institutes.Due to a rapid advancement in the electronic commerce technology, the use of credit cards has increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of credit card fraud also rising. Financial fraud is increasing significantly with the development of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. The fraudulent transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. The concept of paying for goods and services electronically is not a new one. Since

late 1970 and early 1980, a range of Methods has been initiated to accept payment to be resulted across a computer network. After a period of rapid expansion, 1.5 billion populations have internet access globally as of 2008. The e-commerce began at the beginning of the year 1997, an enormous selection of diverse payment techniques developed by the researched some of these were instigated some of these were instigated on the market and unsuccessful to arrive at a critical mass. The e commerce is a process of value exchange in electronic e-commerce; where the amount is transferred online on internet, other computer network. Normally existing fraud detection system for online banking will detect the fraudulent transaction after completion of the transaction.Credit-card-based purchases can be categorized into two types:

1. Physical card
2. Virtual card

*A. Physical 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.

*B. Virtual card*

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. 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. 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. Deviation from such patterns is a potential threat to the system.

*C. Credit Card Fraud*

Credit card fraud can be defined as "Unauthorized account activity by a person for which the account was not intended. Operationally, this is an event for which action can be taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future".

Credit Card Fraud is defined as when an individual uses another individual s credit card for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. And the persons using the card has not at all having the connection with the cardholder or the issuer and has no intention of making the repayments for the purchase they done.

Virtual credit card is where card holder in not present, internet baking is part of virtual credit card. Online baking is challenging part of traditional baking system. The credit card is use of modern society day by day. Prevalent of credit card fraud is difficult task when using online transaction.

## II. RELATED WORK

Dorronsoro and others developed a neural network based fraud detection system called Minerva. This system'smain focus is to imbed itself deep in credit card transaction servers to detect fraud in real-time. It uses a novel nonlineardiscriminant analysis technique that 43 combine the multilayer perceptron architecture of a neural network with Fisher'sdiscriminant analysis method. Minerva does not require a large set of historical data because it acts solely on immediate
previous history, and is able to classify a transaction in 60ms. The disadvantage of this system is the difficulty in determining a meaningful set of detection variables and the difficulty in obtaining effective datasets to train with [6].

Ghosh and Reilly [1] have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labeled credit card account transactions. These transaction contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and nonreceived issue (NRI) fraud.

Recently, Syeda et al. [2] have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection.

Alekerov et al. [4] present CARDWATCH, a 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. [6] suggest the application of distributed data mining in credit card fraud detection. Brauset al. [7] have
developed an approach that involves advanced data mining
techniques and neural network algorithms to obtain high
fraud coverage.

Chiu and Tsai [8] have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. 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.

Phuaetal.[9] have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report.

Fan et al. [10] suggest the application of distributed data mining in credit card fraud detection. Brause et al. [11] have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage.

Chiu and Tsai [12] have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. 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. [15] suggest the use of meta classifier similar to [6] in fraud detection problems. They consider naïve Bayesian, C4.5, and Back Propagation neural networks as the base classifiers. A meta 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. [16] 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 payoff.

## III. PROPOSED METHODOLOGY

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.

An FDS runs at a credit card issuing bank. Each incoming transaction is submitted to the FDS for verification. If the FDS confirms the transaction to be of fraud, it raises an alarm, and the issuing bank declines the 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. Since the number of genuine transactions is a few orders of magnitude higher than the number of malicious transactions, an FDS should bedesigned in such a way that the number of FPs is as low as possible.

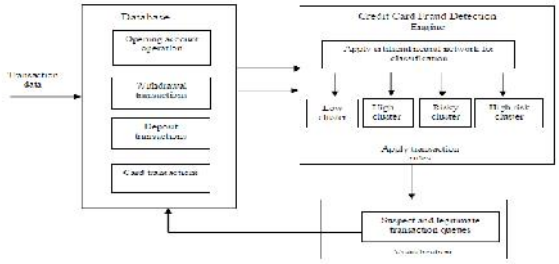

Figure 1.1 Architecture of the Credit Card Fraud

## IV. HIDDEN MARKOV MODEL

A Hidden Markov Model 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. 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.

An HMM is a double embedded stochastic process with two hierarchy levels. It can be used to model complicated stochastic processes as compared to a traditional Markov model. An HMM has a finite set of states governed by a set of transition probabilities. In a particular state, an outcome or observation can be generated according to an associated probability distribution. It is only the outcome and not the state that is visible to an external observer.

To map the credit card transaction processing operation in terms of an HMM, we start by first deciding the observation symbols in our model. Wequantize the purchase values x into M price ranges V1; V2; . . . VM, forming the observation symbols at the issuing bank. The actual price 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. We use Vk, k ¼ 1; 2; . . .M, to represent both the observation symbol, as well as the corresponding price range.

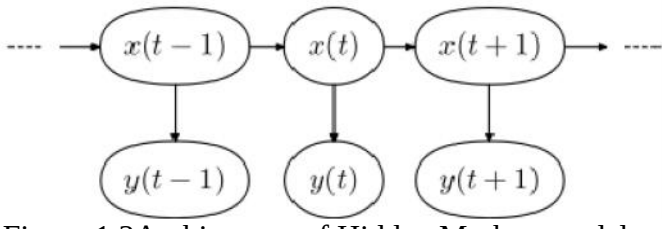


Figure 1.2Architecture of Hidden Markov model

Figure 1.2 shows a general architecture of hidden markov model. Each oval shape represents a random variable that can adopt any number of values. The x(t) and y(t) represents the random variables. The x(t) is hidden state and y(t) is a observation at time 't'. The markov property states that the hidden variables x(t) at all times depends on the values of the hidden variables x(t-1). A hidden markov models are one of the most popular models in machine learning and provides statistics for modeling sequences. A probability distribution over sequences of observations is defined by using HMM. HMM maintains a log of several user transactions which provides a proof for the bank. HMM reduces substantial work of an employee since it maintains a log.

The elements of a discrete-time hidden markov model will now be summarized. These elements will be used throughout the thesis:

Number of states N. although the states are hidden, for many practical applications there is often some physical significance attached to the states or to sets of states of model [17]. For instance in the urn and ball model, the states corresponds to the urns.

The labels for the individual states are {1, 2, 3… N}, and the states at time t denoted qt. Model parameter M. if discrete observation densities are used, the parameter M is the number of class or cells that should be used, e.g. M equals the number of colors in the urn and bell example. If continuous observation densities are used, M is represented by the number of mixtures in every state.

Mathematically an HMM can be defined as below

1. N is number states in the model and set of state is $S=\{S_1, S_2, S_3, \ldots S_N\}$,Where $S_i$, i=1,2,…..N are individual states. State at any time t is denoted by$q_t$.
   - ✓ M is number of distinct observation symbols. Observation symbols correspond to physical output of system being modeled. We denote set of observation symbols $V=\{V_1, V_2, V_3, \ldots V_M\}$,Where $V_i$, i=1,2,3,…M are individual observation symbols.
   - ✓ State transition probability matrix $A=[a_{ij}]$ .where is transition probability from state *i* to *j*.$a_{ij}=P(q_{t+1}=S_{ij}|q_t=S_i)$.
   - ✓ The observation symbol probability matrix B=$b_j$(k).Where $b_j$(k)is the probability distribution of observation symbol *k* at state *j*.
   - ✓ Initial state distribution $\pi=[\pi_i]$,Where $\pi_i=P(q_1=s_1)$. The observation sequence $O=O_1,O_2,O_3,\ldots\ldots O_R$where each observation sequence$O_t$, one of the observation symbols from V, and R is the number of observations in the sequence.

## V. CREDIT CARD FRAUD DETECTION USING AN HMM

HMM uses cardholder's spending behavior to detect fraud. In our Implementation, three behavior of cardholder are taken into consideration.

- ✓ Low spending behavior
- ✓ Medium spending behavior
- ✓ High spending behavior

Different cardholders has their different spending behavior (low, medium, high).Low spending behavior of any cardholder means cardholder spend low amount, medium spending behavior of any cardholder means cardholder spend medium amount, high spending behavior of any cardholder means cardholder spend high amount.

### *A. Generating Observation Symbols*

For each cardholder, we train and maintain an HMM. To find one of the three observation symbols corresponding to individual cardholder's transactions, we run K-means clustering algorithm on past transactions. We use random numbers as spending amounts in transactions. With clustering algorithm we get three clusters and clusters represent observation symbols. We then calculate clustering probability of each cluster, which is percentage of number of transaction in each cluster to total number of transactions.

Transactions in red forms low spending group, transactions in green form medium spending group, and transactions in blue form high spending group. These groups are observation symbols in our implementation. Indicates that clustering probability of each observation symbol.

In this clustering probability of high spending is highest among three. It can be said that spending profile of given cardholder is high spending. Following equation calculates spending profile.

SP=MAX $(P_i)$, $P_i$percentage of number of transaction those belongs to cluster $i$, $1 \leq i \leq M$.

### *B. HMM Training*

Training of an HMM is an offline process. We use Baum-Welch algorithm to train an HMM. Baum-Welch algorithm uses observation symbols generated at the end of k-means clustering. At the end of training phase we get an HMM corresponding to each cardholder. Baum-Welch algorithm is as follow

Particular observation sequence is $O=O_1,O_2,O_3,\ldots\ldots O_T$.

### *C. Initialization*

Set $\lambda=(A, B, \pi)$ with random initial conditions. The algorithm updates the parameters of $\lambda$ iteratively until convergence.

### *D. Forward procedure*

We define: $\propto_i(t)=P(O_1,O_2,O_3,\ldots\ldots O_t,S_t=i \mid \lambda)$, , which is the probability of seeing the partial sequence $O_1,O_2,O_3,\ldots\ldots O_T$and ending up in state *i* at time *t*.

### *E. Fraud Detection*

Let initial sequence of observation symbols of length *R* up to time *t* is $O=O_1,O_2,O_3,\ldots\ldots O_R$. In our implementation we have taken 50 as length of sequence. We calculate the probability of acceptance of this sequence by HMM, let $\propto_1$be the probability of acceptance

$$\propto_1 = P\,(O_1,O_2,O_3,\ldots\ldots O_R, \mid \lambda).$$

At time $t+1$ sequence is $O_1,O_2,O_3,\ldots\ldots O_{R+1}$ let $\propto_2$be the probability of acceptance of this sequence

$$\propto_2 = P\,(O_1,O_2,O_3,\ldots\ldots O_{R+1} \mid \lambda)$$

Let $\Delta\propto=\propto_1-\propto_2$

If $\Delta\propto>0$ it means new sequence is accepted by an HMM with low probability, and it could be a fraud. The new added transaction is determined to be fraudulent if percentage change in probability is above threshold, that is

$$\text{Threshold} \leq \Delta\propto/\propto_1$$

The threshold value can be learned empirically and Baum-Welch algorithm calculates it automatically. If $O_{R+1}$ is malicious, the issuing bank does not approve the transaction, and the FDS discards the symbol.

### *F. K-Means*

Choose K vectors from the training vectors, here denoted $\chi$, at random. , vectors will be the centroids $\mu k$ which is to be found correctly. For each vector in the training set, let every vector belong to cluster k. this is done by choosing the cluster closest to the vector:

$$k^* = \arg\min_k \left[ d(x, \mu_k) \right]$$

From this clustering (done for one state j), the following
$(c_{jm}, \mu_{jm,}|\sum_{jm})$ parameters have found.

Table 1.1Notation

| Notation | Meaning |
|---|---|
| $c_{jm}$ | Weighting coefficients |
| $\mu_{jm,}$ | Mean vectors |
| $\sum_{jm}$ | Covariance matrices |
| O | Observation symbols |
| N | Number of hidden states |
| L,M,H | Low, Medium, High |
| FDS | Fraud Detection System |

It is very difficult to do simulation on real time data set that is not providing from any credit card bank on security reasons. We calculate probability of each spending profile high, low, medium (h, l, and m) Fraud detection of incoming transaction will be checked on last 10 transactions.

Table 1.2: all transactions happened until date

| Transaction No ` | Category | Amount in |
|---|---|---|
| 1 | h | 65 |
| 2 | l | 10 |
| 3 | m | 40 |
| 4 | m | 75 |
| 5 | l | 28 |
| 6 | h | 115 |
| 7 | l | 54 |
| 8 | m | 110 |
| 9 | m | 140 |
| 10 | h | 125 |
| 11 | m | 180 |
| 12 | l | 119 |
| 13 | h | 145 |
| 14 | m | 240 |
| 15 | h | 180 |
| 16 | l | 430 |
| 17 | h | 355 |
| 18 | l | 520 |
| 19 | m | 280 |
| 20 | h | 560 |

We have simulated several large data sets; one is shown in Table 1.2, in our proposed fraud detection system and found out probability mean distribution of false and genuine transactions. When probability of genuine transaction is going down, correspondingly probability of false transaction going up and vice versa. If the percentage change in probability of false transaction will be more than threshold value, then alarm will be generated for fraudulent transaction and credit card bank will decline the same transaction.

According to this table 5.2 we propose:

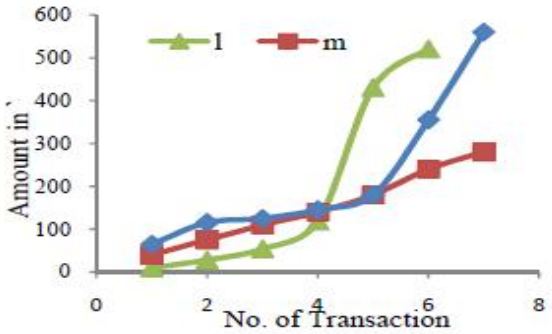


Figure 1.3 Different transaction amount in a Category

The HMM based credit card fraud detection system is not having complex process to perform fraud check like the existing system. Proposed fraud detection system gives genuine and fast result than existing system. The Hidden Markov Model Makes the processing of detection very easy and tries to remove the complexity.

## VI. CONCLUSION

Efficient credit card fraud detection system is an utmost required for card issuing bank or all type of online transaction that through using credit card. In this report, we have implemented of hidden markov model in credit card fraud detection. The very easily detect and remove the complexity forusing in this hidden markov model. It has also explained the hidden markov model how can detect whether an incoming transaction is fraudulent or not comparative studies reveal that the accuracy to the system is also 92 % over a wide variationin the input data. We are dividing the transaction amount in three categories that is grouping high, medium & low used on different ranges of transaction amount each group show the aberration symbols. In hidden markov model methods is very low compare techniques using fraud detection rate. The different steps in credit card

transaction processing are represented as the underlying stochastic process of an HMM. I 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 also have been explained low the hidden markov model can detecting whether an incoming transaction is fraudulent or not. The system is also scalable for handling large volumes of transactions.

## VII.REFERENCES


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

[2] Syeda, M., Zhang, Y. Q., and Pan, Y., 2002 Parallel Granular Networks for Fast Credit Card Fraud Detection, Proceedings of IEEE International Conference on Fuzzy Systems, pp. 572-577 (2002).

[3] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc.Int'l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

[4] Aleskerov, E., Freisleben, B., and Rao, B., 1997. CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection, Proceedings of IEEE/IAFE: Computational Intelligence for Financial Eng. (1997), pp. 220-226.

[5] "Statistics for General and On-Line Card Fraud,"

http://www.epaynews.com/statistics/fraud.html, Mar. 2007.

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

[7] R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. Tools with Artificial Intelligence, pp. 103-106, 1999.

[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.

[9] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research," http://www.bsys.monash.edu.au/people/cphua/, Mar. 2007.

[10] D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams, "Data Mining for Fun and Profit," Statistical Science, vol. 15, no. 2, pp. 111-131, 2000.

[11] S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," ACM Trans. Information and System Security, vol. 3, no. 3, pp. 186-205, 2000.

[12] 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.

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

[14] Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344, 2003.

[15] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, vol. 77, no. 2, pp. 257-286, 1989.

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

[17] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications of Hidden

[18] L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley Series in Probability and Math. Statistics, 1990. J. Banks, J.S. Carson II, B.L. Nelson, and D.M. Nicol, Discrete-Event System Simulation, fourth ed. Prentice Hall, 2004.

[19] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "Credit Card

Fraud Detection Using Meta-Learning: Issues and Initial Results," Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90, 1997.

[20] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

[21] V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-theoretic Approach to Credit Card Fraud Detection," Proc. First Int'l Conf. Information Systems Security, pp. 263-276, 2005.