Hybrid Approach for Credit Card Fraud Detection

Krishna Kumar Tripathi, Lata Ragha

Abstract - Due to a rapid growth in the e-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 normal 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. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques were proposed & implemented in literature, they have their own advantages and disadvantages. We proposed hybrid approach used in credit card fraud detection mechanism.

Index Terms - Credit card fraud, Credit Card Fraud detection methods, Electronic Commerce, Rule based filter, History database, Bayesian theorem, Dempster Shaper Adder.

I. INTRODUCTION

Today technology is basic mandatory need of human. Just look around and you will know why. Literally, at every instant of time, you are surrounded by technology. Today there is no such place where technology is not present. Due to technology communication is easy and quick, travel is fast, and movements are also fast. There are lots of advantages of technology, but with that it causes Fraud also. Fraud is behavior of human which is out of rule and causes crime.

One of the biggest facility provided by technology is that we can able do a shopping using various facility provided by bank e.g Credit Card, Debit Card, Internet Banking [1] etc. Here is major chance for fraud. Credit card becomes the most popular mode of payment for both online as well as regular purchase so mostly frauds happen in Credit Card System. A Credit Card Fraud is a transaction that is complete with your credit card by someone else.

Credit card fraud happens when someone steals your credit card, credit card information, or Personal Identification Number (PIN), and uses it without your permission to make purchases in stores, online or by telephone, or to withdraw money from an automated bank machine (ABM).

Many modern techniques[1] based on Artificial Intelligence, Data mining[3][4], Neural Network[2], Bayesian Network[6], Fuzzy logic[5], Artificial Immune System, K-nearest neighbor algorithm, Support Vector Machine[7][8], Decision Tree, Fuzzy Logic Based System, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. Each method is having its own pros & corns.

Manuscript received on September, 2013.

Krishna Kumar Tripathi, Computer Department, Terna Engineering College, Navi Mumbai, India.

Dr. Lata Ragha, Computer Department, Terna Engineering College, Navi Mumbai, India.

We propose a hybrid approach for credit card fraud detection & prevention, which combines evidences from current as well as past behavior. That consists of following modules.

- Shopping cart or Merchant web applications
- Checkout to payment gateway
- Rule based filter
- Belief Analysis by Bayesian Theorem[6] & Transaction history database
- Dempster–Shafer adder[6] (DSA)
- Security question
- One time password

II. SYSTEM DIAGRAM

Figure I show the block diagram of the flow of transaction.

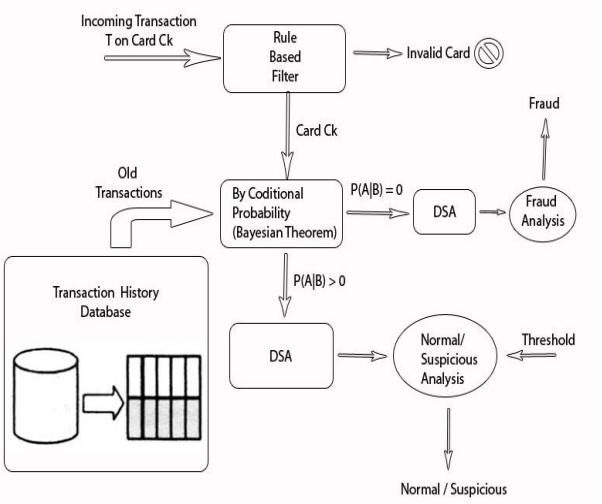

Figure I. System Diagram

The following steps are used for fraud detection.

1) Incoming transaction T and card details C_k will be given as input to rule based filter.
2) Rules based filter. Filter card number by Luhn algorithm. If card number is valid then moves to Bayesian analysis else it will display output as invalid card.
3) Bayesian analysis takes data from transaction history and analyzes current data with it.
4) If P(A|B) = 0 then moves to DSA, DSA make fraud analysis and mark transaction as **Fraud**.
5) If P(A|B) > 0 then moves to DSA, DSA make analysis based on threshold and mark transaction as **Normal** or **Suspicious**.

Where A may be Price, Category, Day or Time and B will be corresponding normal value.

III. SHOPPING CART APPLICATION & CHECKOUT

Its online store were user do shopping. It is part of electronic commerce or also called as part of online marketing. Here user shops products and add it to cart. After checkout of cart he/she redirect to payment gateway. At payment gateway

User need to select card i.e. Master card/Visa etc.

Here we are considering only Master card and Visa card in implementation. Table I. shows some of details of Master card and Visa which is useful for card validation.

Table I. CARD DETAILS

Card	Prefix	Length
Master card	51 - 55	16
Visa	4	13 or 16

These shopping cart applications typically provide a means of capturing a client's payment information, but in the case of a credit card they rely on the software module of the secure gateway provider (such as CCAvenew, Paypal , Billdesk or other bank payment gateway), in conjunction with the secure payment gateway, in order to conduct secure credit card transactions online.
Some setup must be done in the HTML code of the website, and the shopping cart software must be installed on the server which hosts the site, or on the secure server (https protocol) which accepts sensitive ordering information. Later at the process of finalizing the transactions, the information is accessed and an order is generated against the selected item thus clearing the shopping cart.

IV. RULE BASED FILTER

We are using Luhn Algorithm[9] for card number validation. It was designed to protect against accidental errors, not malicious attacks. Most credit cards and many government identification numbers use the algorithm as a simple method of distinguishing valid numbers from collections of random digits. Algorithm I describes Luhn methodology for card validation. Valid number for the same are listed in table II.

Algorithm I

1. First remove spaces / hyphens.
2. Find the length of card number (Input).
3. Find parity / Checksum / check digit
   Parity = Length % 2
4. Define total = 0 (Input)
5. Then we move as –

```
1:  For (I = 0; I < length; I ++)
2:  {
3:  Digit = number [I]
4:  If (I % 2 == parity)
5:  {
6:      Digit *= 2
7:      If (digit > 9)
8:      Digit - = 9
9:  }
10: Total + = digit
11: }
12: ((total % 10) == 0)? TRUE: FALSE
```

Example:

Card Number = 4181 5839 0000 0140
Length = 16
Total = 0
Parity = 16 % 2 = 0

If total % 10 = 0 then card is valid according to Luhn Algorithm else invalid card.

Table II. VALUE FOR CARD NUMBER VALIDATION

i	Card number	Digit
0	4	8
1	1	1
2	8	7
3	1	1
4	5	1
5	8	8
6	3	6
7	9	9
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	1	1
14	4	8
15	0	0

Here total = 50 so in current example card is **"valid"**.

V. BAYESIAN THEOREM

Bayes Theorem is a theorem of probability theory originally stated by the Reverend Thomas Bayes. It can be seen as a way of understanding how the probability that a theory is true is affected by a new piece of evidence.
Using this idea of conditional probability to express what we want to use Bayes Theorem to discover, we say that P(A|B), the probability that A is true given that B is true, is the *posterior probability* of A. The idea is that P(A|B) represents the probability assigned to A *after* taking into account the new piece of evidence, B. To calculate this we need, in addition to the prior probability P(A), two further conditional probabilities indicating how probable our piece of evidence is depending on whether our theory is or is not true.

$$P(A|B_n) = \frac{P(B_n|A)P(B_n)}{\sum P(B_n|A)P(B_n)}$$

Where n = 1, 2, 3, 4 …..

A. Cluster

For this we are considering four clusters like **"Price", "Category", "Day" & "Time"** as shown in figure II. Our work is carried out by considering last ten transactions and gets the probability of each cluster over **"Normal"** and **"Suspicious".**

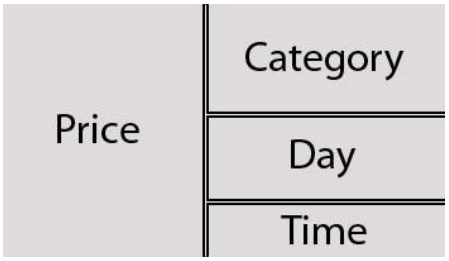

Figure II. Clusters

Cluster probability must give value one. The acquired space of clusters is as shown Table III.

Table III. CLUSTER PROBABILITY

Cluster	Percentage	Probability
Price	40 %	0.4
Category	30 %	0.3
Day	20 %	0.2
Time	10 %	0.1

B. Formulas for conditional probability

The following formulas are used –

$$P(price|normal) = \frac{P(normal|price)P(price)}{\sum P(normal)}$$

$$P(category|normal) = \frac{P(normal|category)P(category)}{\sum P(normal)}$$

$$P(day|normal) = \frac{P(normal|day)P(day)}{\sum P(normal)}$$

$$P(time|normal) = \frac{P(normal|time)P(time)}{\sum P(normal)}$$

C. Threshold for each cluster

$$\text{Threshold for price} = \frac{(0.5)P(price)}{\sum P(normal)}$$

$$\text{Threshold for catrgory} = \frac{(0.5)P(category)}{\sum P(normal)}$$

$$\text{Threshold for day} = \frac{(0.5)P(day)}{\sum P(normal)}$$

$$\text{Threshold for time} = \frac{(0.5)P(time)}{\sum P(normal)}$$

VI. TRANSACTION HISTORY DATABASE

A. Probability normal given price

For calculating probability normal for given price we take all transactions from database match current price with transactions in particular range .Range is calculated using following formula –

$$\theta = \frac{\min + \max}{4}$$

B. Probability normal given category

For this we take all transactions from database match current category with transactions.

C. Probability normal given day

For this we take all transactions from database match current day with transactions.

D. Probability normal given time

For this we take all transactions from database match current time with transactions. We match current time in between five hour before and five hour after.

VII. DEMPSTER–SHAFER ADDER

The role of the DSA is to combine evidences from the rules observation of Bayesian Network and by conditional probability and compute an overall belief value for each transaction.

Suppose from current transaction we get –
[Normal, Suspicious, Suspicious, Normal]
By using DSA we can make conclusion that overall result is **"Normal".**

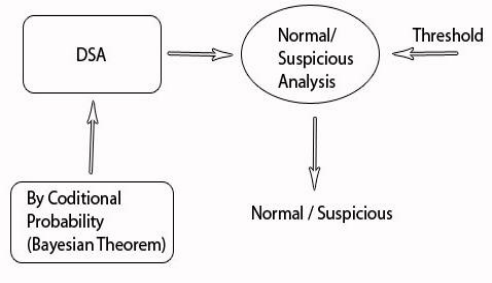

Figure III. DSA

RESULT

We demonstrate the effectiveness and usefulness of our FDS by testing it with large scale data. Due to unavailability of real life credit card data or benchmark data set for testing, we used dummy data that represent the behaviour of genuine cardholders as well as that of fraudsters.

Consider particular example, we demonstrate result by observation and by using mathematical analysis. Mathematical
Analysis is cross checked with threshold. These thresholds are calculated by using above threshold formulas.

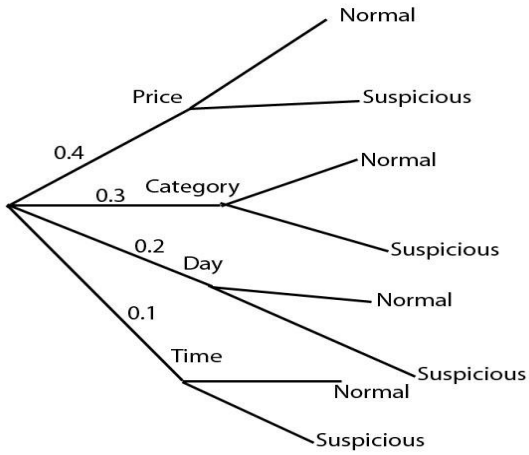

Figure IV. CLUSTER NETWORK

From above network we can draw result -

Table IV. RESULT

Cluster	Observation result	By Bayesian Theorem
Price	Normal	0.64 > 0.40 Normal
Category	Suspicious	0.06 < 0.30 Suspicious
Day	Suspicious	0.16 < 0.20 Suspicious
Time	Normal	0.14 > 0.10 Normal

From observation of cluster network shown in Figure IV. we get result as Normal, Suspicious, Suspicious and normal as indicated in Table IV. Also from mathematical calculation we get similar result.

We can represent result by bar chart.

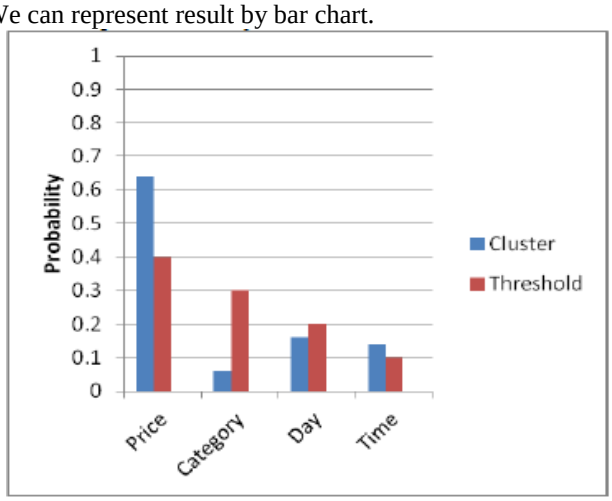

Figure V. Bar Chart Representation Of Reult

VIII. SECURITY QUESTION

A security question [11] is used as an authenticator by banks, many other companies as an extra security layer. They are a form of shared secret. Financial institutions have used questions to authenticate customers since at least the early 20th century. In a 1906 speech at a meeting of a section of the American Bankers Association, Baltimore banker William M. Hayden [12] described his institution's use of security questions as a supplement to customer signature records.

In the 2000s, security questions came into widespread use on the Internet. As a form of self-service password reset, security questions have reduced information technology help desk costs. By allowing the use of security questions online, they are rendered vulnerable to keystroke logging attacks. In addition, whereas a human customer service representative may be able to cope with inexact security answers appropriately, computers are less adept. As such, users must remember the exact spelling and sometimes even case of the answers they provide, which poses the threat that more answers will be written down, exposing them to physical theft.

IX. ONE TIME PASSWORD

- One time password [10] is one of those changes every time while using.
- One time password is very important for authentication because an intercept static password is useless because it cannot be reused.
- One time generated password SMS is sent to users register mobile number. User puts that password to our system and get authenticate.

X. CONCLUSION

The day by day credit card use is increasing online and offline. So according to that the fraud of credit card also increases. Every bank, finance company and others finance related institutes required this system. There are number of techniques present to implement this system. The objective of this survey was to review the major research in the area of intrusion detection using the Dempster-Shaffer theory of evidence. Most of the researchers have discussed of the resolution of various issues and intended future work in this area. It is very fast and effective method using minimum effort so widely prefer.

Credit card fraud detection system based on the integration of three approaches, namely, rule-based filtering, Dempster–Shafer theory and Bayesian learning. Dempster's rule is applied to combine multiple evidences from the rule-based component for computation of initial belief about each incoming transaction. The suspicion score is updated by means of Bayesian learning using history database of both genuine cardholder as well as fraudster.

Credit card fraud detection has drawn quite a lot of interest from the research community and a number of techniques have been proposed to count credit fraud. Bayesian learning takes place so that the FDS dynamically adapts to the changing behavior of genuine customers as well as fraudsters over time. Dempster–Shafer theory gives good performance, especially in terms of true positives, Bayesian learning helps to further improve the system accuracy.

Finally Fraud detection system gives more performance in terms of accuracy.

REFERENCES

[1] Linda Delamaire"Credit card fraud and detection techniques: a review". Bank and Bank Systems, Volume4, Issue2. (2009).
[2] Ghosh, D.L.Reilly, "Credit card fraud and detection with a Neural-Network". Proceeding of the International Conference on System Science ;(1994). (621-630).
[3] E.W.T.Ngai, Yong Hu, Y.H.Wong, Yijun Chen, Xin Sun "The application of data mining techniques in an academic review of literature". Elsevier-Decision Support System (2011).50; (559-569).
[4] Sherly K.K. (2012) "A comparative assessment of supervised data mining techniques for fraud prevention" TIST.Int.J.Sci.Tech.Res, Vol. 1; (1-6).
[5] S.Maes, K.Tuyls, B.Vanschoenwinkel, B.Manderick" Credit card fraud detection using Bayesian and neural networks". Processing of the First International NAISO Congress on Neuro Fuzzy Technologies ;(1993). (261-270).
[6] Suvasini Panigrahi, Amlan Kundu, Shamik Sural, A.K.Majumdar "Credit card fraud detection: A fusion approach using Dempster-Shafer Theory and Bayesian learning".Elsevier, Information Fusion 10 :(2009). (354-363).
[7] N. Cristianini, J. Shawe-Taylor "An Introduction to Support Vector Machines and Other Kernal-based Learning Methods". Cambridge University Press. (2000).
[8] Cortes, C. & Vapnik, V "Support vector networks Machine Learning". (1995).Vol.20 ;(273-297).
[9] http://en.wikipedia.org/wiki/Luhn_algorithm
[10] http://en.wikipedia.org/wiki/One-time_password
[11] http://en.wikipedia.org/wiki/Security_question
[12] William M. Hayden (1906), Systems in Savings Banks, *The Banking Law Journal*, volume 23, page 909.