ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

# Prediction of the Malicious Activities on Online Transaction Based on User's Usage Behavior

Ms.M.Dhanalakshmi, M.E, (2nd year) [1] Mrs.V.Umadevi, M.E,
Computer science and Engineering
Arunai Engineering College, Tiruvannamalai.

*Abstract*— **Now a days people are using the credit card, debit card, mobile banking and internet banking for their business transactions, online shopping, bill payments like electricity charges, insurance payments and also booking a railway tickets, etc. The E-banking will do a vital role on this online transaction, increase of these online transactions many fraudulent activities are also increased in our account information. In existing system of fraud detection using the technique called as GLT and TRGSM in this fraudulent transaction can be identified once the fraud has happened. It is difficult to find out fraudulent transaction once it has done. The main objective of my proposed work is to prevent the fraudulent activity done in customer accounts. To do this work I have proposed the technique called as forward algorithm, the purpose is to prevent the customer from fraudulent transaction by using specific technique i.e. increasing number accessing questionnaire based on the user's usage behavior. The forward algorithm will increase security more compare than the K-mean algorithm in existing system. If any fraudulent activity observed means then SIM and SPC will be generated and the entering pattern will be known by only the genuine users so the unauthorized used cannot access the authorized user accounts. The objective is to identify the customer behavior by using forward algorithm able to stop the fraudulent activity before accessing to the transactions. (i.e) can stop them from the authentication itself.**

***Keywords***— SIM, SPC, Usage Behavior,Fraud Detection,Secure Transaction,Conformity,Non-conformity

## I. INTRODUCTION

Big data is a term for any collection of data sets which are large, complex and unstructured so that it becomes difficult to process using traditional data processing applications. It usually includes data sets with size beyond the ability of commonly used software tools to analyze, capture, curate, manage and process data within a tolerable elapsed time. In big data 'size' is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data. There are various kind of challenges in big data such as analyse, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.

In 2001 research report and related lectures, META GROUP analyst DOOG LANEY defined data growthchallenges and opportunities as being in three dimension (i.e)increasing volume(amount of data),velocity(speed of data) and variety(range of data types and sources). Additionally a new V 'Veracity' is added by some organizations expanded to
other complementary characteristics of big data. There are three kind of data used in big data such as structured, un-structured and semi-structured.

The term big data often refers simply to the use of predictive analytics, user behaviour analytics or certain other advanced data analytic methods that extract value from data and seldom to particular size of data sets. Accuracy in big data may lead to more confident decision making and better decision can result in greater operational efficiency, cost reduction and reduced risk. We are using big data for many purposes like social media analysis, web display advertising, science and research, call centre optimization sensor analysis, fraud risk management, finance, internet search and many other things. Most of these analytical solution were not possible previously because data technology were unable to store such huge size of data or processing technologies were not capable of handling large volume of workload or it was too costly to implement the solution in timely manner.

In 2000, seisint Inc., developed C++ based distributed file sharing framework for data storage and querying. Data are stored in multiple servers and querying is done by C++ called ECL. In 2004 LexisNexis acquired seisint Inc., and 2008 acquired ChoicePoint, Inc and their high speed parallel processing platform. These two platforms merged into HPCC systems and in 2011 was open sourced under

Apache V2.0 License. In 2004, google published a paper on a process called MapReduce that used such an architecture and it provide parallel processing model and associated implementation to process huge amount of data. Data can be processed in Map step and results delivered and gathered in Reduce step.

Based on online transaction eBay uses two data warehoused at 7.5PB and 40PB hadoop cluster for search, consumer recommendations and merchandising. Amazon handles million of back-end operations every day as well as queries from more than half-a million third party sellers. walmart and retail banking handles more than one million customer transaction every hour, which are imported into database estimated to contain more than 2.5PB of data. Clemons etal (2014) identify different ways to characterize privacy and online invasion of privacy one of the way is an uninvited intrusion into a user's personal space. This includes online marketing, spam advertising, pop-ups and sponsored sites around the edges of a web page. The most serious threats are fraudulent e-commerce transactions and identify theft.

The credit card ownership and usage behaviour significantly depends on demographic characteristics of user such as sex, age, profession, religious believe, education level, income marital status, culture and attitudes toward debt. According to the planned behaviour theory the credit card consumption behaviour depends on earning capabilities, toward risk tolerance attitude and the planning horizon. The user habits are classified as logit models, associations were discovered between personal characteristics and credit card habits that involve financially risky behaviour. The usage behaviour can be classified as following,

1. Always paying a credit card balance fall.
2. Carrying over a credit card balance and being charged interest.
3. Making only minimum payment on a credit card balance.
4. Being charged a fee for a late payment.
5. Being charged a fee for exceeding a credit card limit.

Specifically credit derived from cards has been studied in terms of one's attitude toward credit as well as the correlation between one's attitude toward credit card and their usage of credit. Kim and Devanely liked various factors such as education, credit card interest rates, credit limits and attitudes as positively correlated to having outstanding balance on credit cards.

### A. *Methods to steal personal information:*

- ⦿ Hacking
- ⦿ Phishing
- ⦿ Spoofing
- ⦿ Spyware
- ⦿ Shoulder Surfing
- ⦿ Dumpster Diving

### B. *Some security mechanism to prevent from fraud:*

- ⦿ AVS(Address Verification System)
- ⦿ Fraud Rates
- ⦿ Relocation
- ⦿ Chip and PIN
- ⦿ 3D Secure
- ⦿ Bio-tech
- ⦿ OTP

## II. LITERATURE SURVEY

An outlier is defined as a data point which is very different from the rest of the data based on some measure. Such a point often contains useful information on abnormal behavior of the system described by the data. Outliers are a well-known problem in all experimental scientific and industrial fields. Outlier detection can usually be considered as a pre-processing step for locating, in a data set, those objects that do not conform to well-defined notions of expected behavior. To detect the outlier data points we are using the logistic sigmoid functions related to hyperbolic tangent will be used weightage function for find outlier data.[1]

Competitiveness in the retail industry is continuing and it is becoming increasingly aggressive as revealed by recent events in the sector and specialist studies. One of the leading businesses in this sphere is hire purchase and one of the main commercial strategies is the emission of department store credit cards to clients. In order to extract knowledge so that normal behavior patterns may be obtained in unlawful transactions from transactional credit card databases in order to detect and prevent fraud. To prevent this fraud in this proposed work we are using the association rules and fuzzy association rule techniques. Another possibility is to give the values a certain degree of imprecision and to extract fuzzy association rules.[2]

In the last few years, there has been a rapid increase in the number of card issuers, card users and online merchants, giving very little time for technology to catch-up and prevent online fraud completely. This has made it easier for fraudsters to indulge in new and abstruse ways of committing credit card fraud over the Internet. This paper focuses on real-time fraud detection and presents a new and innovative approach in understanding spending patterns to decipher potential fraud cases. It makes use of self-organization map to decipher, filter and analyze customer behavior for detection of fraud. Fraud detection techniques involving sophisticated screening of transactions to tracking customer behaviour and spending patterns are now being developed and employed by both merchants as well as card issuer banks. Some of the recently employed techniques include transaction screening through Address Verification Systems (AVS), Card Verification Method (CVM), Personal Identification Number (PIN) and Biometrics.[3]

The most substantial field of data mining is discovery of frequent objects like sequential pattern and item sets. It is well known that it requires less number of database scanning, reduced number of rules, time and space efficiency. To do this the efficient transaction reduction approach named TR-BC(transaction reduction and bitmap class lables) which is used to reduces the rule generation by counting the item support and class support instead of only item support. The rules are reduced by horizontal and vertical transaction and then finally combined rules are generated by eliminating the redundancy. The existing algorithm is CCARM (combined and composite association rule mining) produces high number of rules and multiple database scans that negatively affect the mining performance by consuming more system resources. The TR-BC is used to mine the frequent pattern based on bitmap and class lables.[4]

Performance measured in banking industry is normally involved with various qualitative as well as quantitative criteria, which leads to the implementation of multiple criteria decision making techniques. This proposed method presents a hybrid grey relational analysis and k-means to cluster and measure the performance of banking system. Performance measurement plays essential role on improving business units performance and their efficiencies. Today, the ability to detect the profitable customers, building a long-term loyalty in them and expanding the existing relationships is the primary key and competitive factors for a customer oriented organization. The prerequisite for having such competitive factors is the existence of a very powerful customer relationship management (CRM). RFM is a technique which scrutinizes three properties namely recency, frequency and monetary for each customer and scores based on these properties presented a method, which obtains the behavioural traits of customer using the extended RFM approach. It then classifies the customer based on k-means algorithm and finally scores the customer in terms of their loyalty in each cluster.[5]

E-Commerce has become a dynamic force, changing all kinds of business operations world-wide, new forms of fraud based on Internet has been invented. In comparison with traditional fraud, online-fraud poses more challenges in terms of prevention and detection. The algorithms focus only on very specific types of their applications, and do not try to implement an extensible approach to the prevention of different kinds of online fraud. This paper proposes an improved approach of E-fraud detection by using logging data sets which can help e-commerce better understand the issues and plan the activities involved in a systemic approach to E-fraud and it is focused on logging data since many systems have logging for accounting purposes. In order to obtain sufficient quality data for fraud detection system, a Video-On-Demand system is used for this purpose. This paper proposes a new hybrid model for online fraud detection of the Video-on-Demand System, which is aimed to improve the current Risk Management Pipeline (RMP) by adding Artificial Immune System (AIS) based fraud detection for logging data. The AIS based model combines two artificial immune system algorithms with behavior based intrusion detection using Classification and Regression trees (CART).[6]

We develop a method which improves a credit card fraud detection solution currently being used in a bank. With this solution each transaction is scored and based on these scores the transactions are classified as fraudulent or legitimate. The proposed method is used to improve a credit card fraud detection solution currently being used in banks. With this solution each transaction is scored and based on these score the transaction are classified as fraudulent or legitimate. Fraud detection has been usually seen as a data mining problem where the objective is to correctly classify the transaction as legitimate or fraudulent. After analyze the main characteristics of popular meta-heuristic algorithms, for our problem we decided to use the Genetic Algorithm and Scatter Search in combined manner called hybrid solution as GASS.[7]

The new target of database marketing in banking and financial service is to provide the right product to the right customer at right time. Data mining improve decision making by searching for relationship and patterns from extensive data collected by organization, it also reduce information overload and it is often applied to extract and uncover the hidden truths being very large quantities of data. The RFM is one of the most important elements in clustering it is a 3-dimensional way of ranking customer to determine top 20% or best customer. It is based on 80/20 principle where 20% of customer brings 80% of revenues. It uses 3 features like Recency, Frequency, Monetary. It is also uses the two clusters traditional statistical method and neural network approach. In this approach we need to find the customer behavior.[8]

Fraud detection is generally viewed as a data mining classification problem, where the objective is to correctly classify the credit card transactions as legitimate or fraudulent. Fraud detection involves monitoring the behavior of users in order to estimate, detect, or avoid undesirable behavior. Due to the rise and rapid growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. In an era of digitalization, credit card fraud detection is of great importance to financial institutions. In this paper, we analyze credit card fraud detection using different techniques: Bayesian Learning, BLAST-SSAHA Hybridization, Hidden Markov Model, Fuzzy Darwinian detection, Neural Networks, SVM, K-Nearest Neighbour and Naïve Bayes. After analyzing through each technique, our aim is to compare all the techniques based on some parameters.[9]

With the developments in the information technology, fraud is spreading all over the world, resulting in huge financial losses. Though fraud prevention mechanisms such as CHIP&PIN are developed for credit card systems, these mechanisms do not prevent the most common fraud types such as fraudulent credit card usages over virtual POS (Point Of Sale) terminals or mail orders so called online credit card fraud. a new cost-sensitive decision tree approach which minimizes the sum of misclassification costs while selecting the splitting attribute at each non-terminal node is developed and the performance of this approach is compared with the well-known traditional classification models on a real world credit card data set. We have developed and implemented a number of cost-sensitive decision tree approaches to be used in credit card fraud detection and show that it outperforms the models built using the traditional data mining methods such as decision trees, ANN and SVM.[10]

## III.EXISTING SYSTEM

In existing Fraud detection system(FDS) to prevent the customer from the online transaction fraudulent activities uses the geographical location technique(GLT) based on the data mining concept and in this method risk score(RS) is calculated by using Bayesian learning approach. In geographical location based method the current location is compared with the previous geographical location. Based on this comparison the risk score is calculated if the risk score is less than 0.5 then it is genuine transaction otherwise it is fraudulent transaction. In this technique the shipping address is compared with the delivery address if the address is mismatched then risk score is generated. For example if any customers perform the transaction in India at 9am and the same time user perform the transaction in London at 11am then it is treated as fraudulent transaction. By using k-mean algorithm first find the machine address and trace the IP address to find the geographical location then compare the current location with the previous location. Then calculate the distance between two location find actual distances of user if the calculated distance is greater than the actual distance then the transaction is fraudulent transaction.

### Problem identified:

In this geographical location based technique it is easy to made fraudulent transaction because the frauduster easily hake the customer information. Then in this method sometimes the genuine users is also treated as a fraudulent user and we need to continuously update the user's activities. In this the malicious activities are found once it will happened.

## IV.PROPOSED SYSTEM

To make the online transaction more secure compare than the previous method I have proposed a new technology by using the forward algorithm. In this proposed method we need analyze the user's usage behaviour(UUB) to find fraudulent activities. The main objective of my work is to prevent the fraudulent activities done in a customer account by increasing the number of questionnaries based on timing. The user's usage behaviour is continuously monitored and stored in a bank database. While online transaction any non-conformity of user's usage behaviour is found then first the Security Information Module(SIM) will be generated it have number of security questions and each will be filled with time based manner. After that the Security Pattern Code(SPC) will be send to the customer mobile and the entering pattern of the code will be known by the authenticated customer only so the frauduster can't access the customer account and don't make transaction. By using these two ways the forward algorithm can able to stop the fraudulent activities before the transaction done. (i . e) can stop them in authentication itself. If the forward algorithm confirms the transaction is fake, it detects the intrusion and issuing bank declines the transaction then the customer alerted by the bank administrator.

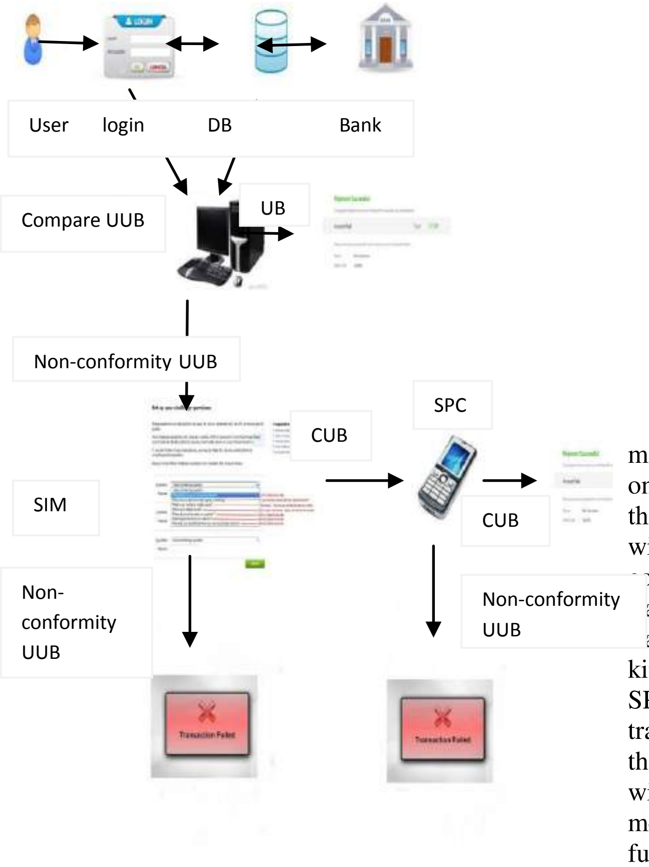
### Advantages:

1. It is more secure compare than the geographical location method.
2. Number of states observation increases security level more accurate.

3. Stop fraudulent activities authentication itself.
4. More efficient compare than GLT.
5. Bank itself decline transaction.

## V.ARCHITECTURE DIAGRAM

The architecture diagram is shown in the given figure. In that the customers make online transaction means then they need to login after that login the customer's usage behavior will be stored on bank's database. Continuously the customer behavior will be stored for minimum 10 transactions. If any non-conformity of that usage behavior will be found means then the SIM module will be generated in that it will have set of security questions we need to answer those questions if any wrong answer given means the transaction will be failed. Otherwise the SPC module will be generated in the 5digit code will be send to customer's mobile the entering pattern(i.e) addition of those numbers, subtraction of those numbers, reverse order of those numbers,etc., only known by genuine customer. The correct pattern entered means the transaction is done otherwise failed.



## VI.ALGORITHM

In this proposed work I use a forward algorithm it have following steps,

- We use the forward algorithm to find the probability of an observed sequence of user spending behavior.
- It exploits recursion in the calculations to avoid the necessity for exhaustive calculation of all paths through the execution trellis.
- Given this algorithm, it is straightforward to determine which of a number of HMMs best describes a given observation sequence - the forward algorithm is evaluated for each, and that giving the highest probability selected.
- We use the forward algorithm to calculate the probability of a T long observation sequence;
- where each of the y is one of the observable set. Intermediate probabilities ( 's) are calculated recursively by first calculating for all states at t=1.

❖ **STEPS:**

- To use the example follow these steps :
- Enter a number of valid observed states in the input field.
- Press 'Set' to initialise the matrix.
- Use either 'Run' or 'Step' to make the calculations.
    - 'Run' will calculate the 's for each and every node and return the probability of the HMM.
    - 'Step' will calculate the value for the next node only. Its value is displayed in the output window.
- When you have finished with the current settings you may press 'Set' to reinitialise with the current settings, or you may enter a new set of observed states, followed by 'Set'.

## VII.CONCLUSION

In this paper the purpose is to propose a security mechanism using forward algorithm that can detect online fraud and give more secure transaction. To give this security I analyze the customer's usage behaviour with minimum 10 transaction and if any non-onformity of those transaction is found while online ansaction then it is concluded as fraudulent ansaction. In this the proposed method have two kind of modules in that first is the SIM and second is SPC after completion of these modules only the transaction will be allowed otherwise it is denied and the customer is alerted by the bank that is your card will be stolen or misused. This proposed method is more secure compare than the existing technique. In future work many kind of activities related to the customer behaviour and actions will be added to give secure transaction.

## REFERENCES

[1]. Wu.S and Wang.S , "Outlier Detection using Weighted Holoentropy with Hyperbolic Tangent Function", International Journal of Advances in Engineering Science and Technology ISSN: 2319-1112,2013.

[2]. Sanchez.D, Cerda.L, Serrano.J.M and Vila-.M.A, "Association Rules applied to Credit Card Fraud Detection", Science Direct Expert System with applications 36 pp 3630-3640, 2009.

[3]. Quah.J.T.S and Sriganesh.M, "Real-time credit card fraud detection using Computational Intelligence", Science Direct Expert System with applications 35 pp 1721-1732, 2007.

[4]. Cao.L, Zhang.H, Zhao.Y, Luo.D and Zhang.C, "Combined Mining: Discovering Informative Knowledge in complex data", IEEE Transactions Vol. 41 No.3 pp 699-712, 2011.

[5]. Mohammad Emami and Farshad Faezy Razi, "A hybrid grey based K-means and feature selection for bank evaluation", Decision Science Letters 3 (2014) 269–274,2014

[6]. Huang.R, Tawfik.H and Nagar.A.K, "A Novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection", International Conference on Computer Science, Science Direct, pp 2733-2742, 2012.

[7]. Duman.E. and Ozcelik.H.M , "Detecting credit card fraud by genetic algorithm and scatter search", Science Direct, Expert System with Applications 38 , pp 13057-13063,2011.

[8]. Bhattacharya.S , Jha.S, Tharakunnel.K and Westland.C.J, "Data mining for credit card fraud", Science Direct, Decision Support System pp 602-613, 2010.

[9]. Kundu.A,Suvasini.P,Sural.S and Majumdar.A.K,"BLAST-SSAHA Hybridization for Credit Card Fraud Detection" IEEE Transactions On Dependable and Secure Computing, VOL. 6, NO. 4, Dec 2009.

[10]. Sahin.Y, Bulkan.S and Duman.E, "A cost-sensitive decision tree approach for fraud detection", Science Direct, Expert System with Applications 40 pp-5916-5923, 2013.