A Data Mining Approach for Money Laundering Detection

Deepthi B C, M.tech Student, PES College of Engineering, Mandya, India, deepthibc31@gmail.com

Dr. Vinay S, Professor, PES College of Engineering, Mandya, India, vinaymanyan@gmail.com

Abstract: **Money laundering is the process of making huge amount of money generated by illegal activities such as drug trafficking, fraud or economic resources, etc. The money got from the criminal ways is also called as black money, the way of launder method makes it to look clean. The money laundering itself is a crime. The money Laundering has three stages: Placement, layering, integration. The existing work based on data mining uses hash based technique combining with graph theoretic approach to detect money laundering. The existing work is enhanced by considering previous year transaction data along with a specific threshold to predict whether the transaction is authorized or not.**

Keywords — Money laundering, Money laundering detection, Graph theory, Suspicious transaction, Anti-money laundering.

I. INTRODUCTION

Money laundering is conversion of unusable money into usable money. With the emergence of improved technology, booming e-commerce applications billions of transactions are happening online. Technology has simplified doing transaction with a single click which also makes it vulnerable to hackers. It not only effects the society and economic but it also effects the whole country. Hence identifying valid transactions is compulsory. The existing system has many drawbacks and its a manual process. This proposed system is automation for the prediction of suspicious accounts and transactions. It works based on transaction data and depending on analysis result of transaction data results will be given. The crime activity is getting highly complicated with the emergence of technology which has its positives as well as negatives.

Currently Reserve bank of India (RBI) gets all the data related to transactions which the bank deem as high valued or suspicious. This data is submitted to Financial Investigation Unit (FIU) for verification. This statistical approach of detecting suspicious account or transactions is cumbersome. Money Laundering is used in three stages such as placement, layering, and integration. Here, in placement stage, the criminal person giving illegal cashes to dealer. In layering, the dealer distributes money to multiple intermediaries. This is the challenging phase for money laundering detection since money is likely to be deposited as cash to accounts of multiple persons. Integration refers to aggregating money distributed to many and giving it to intended beneficiary and making it legal.

The proposed system detects suspicious accounts in money laundering. Data mining techniques will be of great use to detect such transactions. Propose system detects the suspicious accounts in money laundering process. Proposed system predicts the traversal path of money laundering and identifying the agent and integrator. The system satisfies the authorities in a better way by providing better results on prediction of suspicious accounts. This system makes the process of finding illegal account simpler by predicting suspicious accounts. In layering stage, the illegal cash spread into so many intermediaries, and also includes banks, financial institution etc,. Layering is the most challenging for fraud detection since it transfers money from one-to-one or one-to-many. The layering stage is difficult to trace all transaction.

In section 2 related work is discussed, section 3 introduces proposed system followed by experimental analysis in section 4. Section 5 describes the summary of the work.

DOI : 10.35291/2454-9150.2019.0087

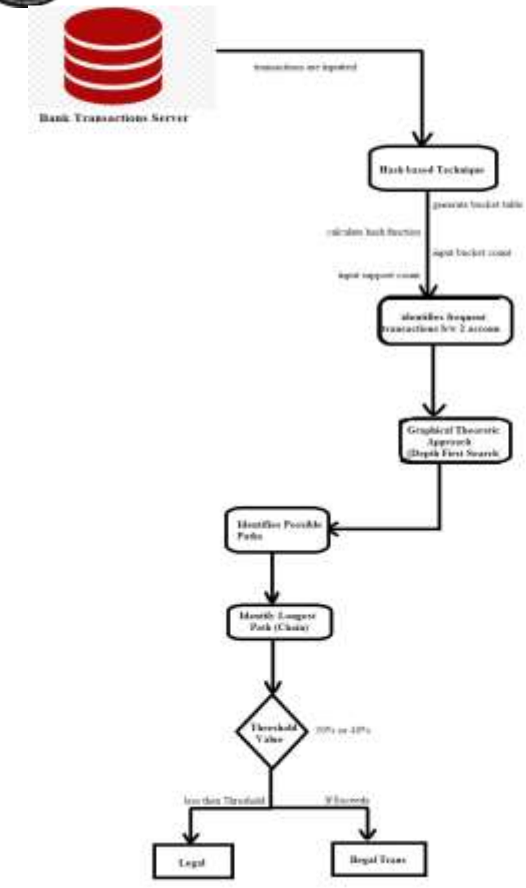

Fig 1: Fraud Detection Using Data Mining Techniques.

II. Related Work

Money laundering is a criminal activity to show the black money as white money. Money laundering is a criminal offense in which the large amount of funds and assets that has been obtained by the illegal activities will be converted into the legitimate assets and funds. There are three stages by which the money laundering happens namely, placement, layering and integration.

Money laundering will lead to criminal activity such has political corruption, terrorist funding, smuggling, financial frauds, and so on. In India efforts are on through Anti Money Laundering techniques to prevent the money laundering activity. Indian banks are following the guidelines that have been given by the Reserve Bank India. As per these guidelines if any suspicious transactions are found, this will be send to the Financial Intelligence Unit. But this is not the efficient way of detecting the money laundering since it is time consuming process and also the laundering done through the systems cannot be detected. [1] uses Hash based Association in the Anti money laundering system which is capable of identifying the traversal path of the laundered money. Also they have used the Graph theoretic approach to detect the money laundering in layering level.

Authors in [2] focus on using data mining techniques to understand client behaviour and profile. The propose a learning component in their agent based approach to money laundering.

Authors in [3] propose a knowledge based solution which is a combination of data mining and natural computing techniques for detecting money laundering. The work is part of a collaboration project with an investment bank.

Reza Sultana et al [4] propose a framework that focusses on reducing the input data set to a smaller size using reduction methods. The framework finds pairs of transactions with common attributes and behaviours that could be potentially suspicious transactions. A clustering method is then applied to detect potential ML groups.

Authors in [5] carry out a review of the Principal Indicators and Data Science Techniques used for the Detection of Financial Fraud and Money Laundering. It also serves as a guide to carry out future work using Data Science techniques for money laundering detection.

Authors in [6] do a comparison of Data Mining Techniques for Money Laundering Detection System. They have done two sets of experiment and the performance of different algorithms has been measured, compared and summarized.

Authors in [7] describe a framework based on rule base monitoring, behavior detection monitoring, cluster monitoring and link analysis based monitoring.

III. Proposed System

The money laundering is tougher task and it takes more number of transactions involved. In this proposed method to overcome suspicious transactions, a Hash based association mining for generating frequent transactional datasets and a graph theoretic approach for identifying the traversal path of the suspicious transactions is used. Proposed work is an extension of work done in [1] as shown in fig 1. Hash base Technique is to generate frequent accounts. The proposed system used synthetic transactional database for the experiment. The present banking system is considered each individual bank's data and is applied for same scenario.

Bank server provides transaction data that needs to be pre-processed. The pre-processing involves filtering and removes irrelevant data such as mobile, email id etc,. That data is input to the hash based technique. It generates frequent-2 item. The output of hash based technique is input to the graph theoretic approach. It is the longest path determines the agent and integrator.

It enhances the longest path predicted by the graphic theoretical approach. The longest path will be compared with previous year transaction data and a threshold is applied. Threshold could be scenario specific. For example if the threshold is 30%, and transactions which exceeds threshold is termed suspicious. The proposed system is shown in fig 2.

DOI : 10.35291/2454-9150.2019.0087

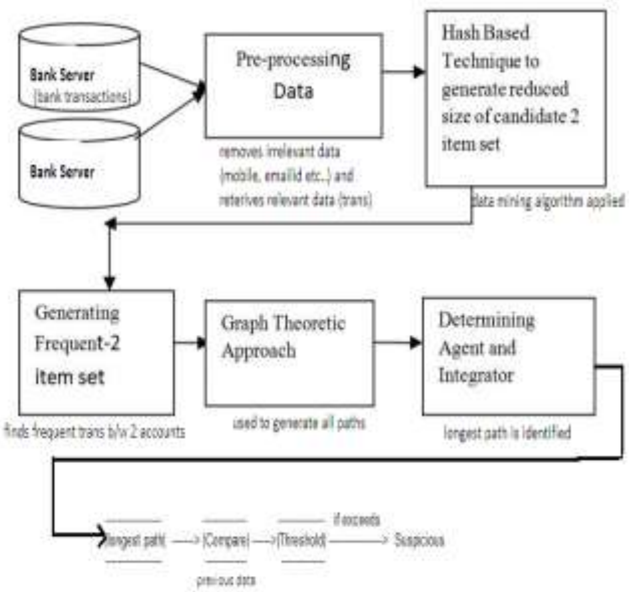
Bank Server
(bank transactions)
Bank Server
Pre-processing Data
removes irrelevant data (mobile, emailid etc..) and reterives relevant data (trans)
Hash Based Technique to generate reduced size of candidate 2 item set
Data mining algorithm applied
Generating Frequent-2 item set
finds frequent trans b/w 2 accounts
Graph Theoretic Approach
used to generate all paths
Determining Agent and Integrator
longest path is identified
longest path
Compare
previous data
Threshold
if exceeds
Suspicious

Fig 2: Proposed system enhancement model.

They are two steps in the proposed method

Step 1: Hash based technique help us to reduce the candidate k-items.

- Hash table uses the hash function to create hash table

h (x ,y)=((order of x *10)+order of y)mod 8

Step 2: Graphical theoretical approach shown in fig 3 helps to identify transition, it's illegal or not. A graph is generated by linking all transactions sequentially by considering each account in the frequent item set as a node. Weights are assigned for each link between the transaction. In-degree and out-degree of each node between agent and integrator are identified.

Table 1, 2, 3 and 4 shows the steps involved in the method.

- Table Generation of 2 item set using hash based method.

Trans. ID	Form-to transaction	2-item set	Trans. ID	Form-to transaction	2-item set	Trans. ID	Form-to transaction	2-item set
A	T1->T2	{1,2}	I	T4->T5	{4,5}	R	T1->T7	{1,7}
B	T2->T3	{2,3}	J	T3->T5	{3,5}	S	T3->T4	{3,4}
C	T3->T4	{3,4}	K	T5->T6	{5,6}	T	T3->T5	{3,5}
D	T1->T6	{1,6}	L	T1->T6	{1,6}	U	T6->T7	{6,7}
E	T2->T7	{2,3}	M	T3->T4	{3,4}	V	T6->T7	{6,7}
F	T3->T4	{3,4}	N	T5->T6	{5,6}	X	T1->T2	{1,2}
G	T4->T5	{4,5}	Q	T4->T6	{4,6}	Y	T4->T7	{4,7}
H	T5->T6	{5,6}	P	T1->T3	{1,3}	Z	T2->T3	2,3}
A	T2->T3	{2,3}						

A to Z transaction grouped in an index table. Now we calculate bucket count in each bucket. The less than minimum bucket will be deleted.

2. Table Bucket table with bucket counts.

Bucket Address	0	1	2	3	4	5	6	7
Bucket contents	1,6 1,6 5,6 5,6 5,6	1,7	3,4 3,4 3,4 3,4	2,7 3,5 3,5 6,7 6,7	1,2 1,2	4,5 4,5 1,3	4,6	2,3 2,3 2,3 4,7
Bucket Count	5	1	4	5	2	3	1	4

Minimum bucket count =2

3. Table Bucket count and item sets

Item set	Bucket count
1,6	5
5,6	5
1,7	1(*discarded)
3,4	4
2,7	5
3,5	5
6,7	5
1,2	2
4,5	3
1,3	3
4,6	1(*discarded)
2,3	4
4,7	4

4. Table Bucket count and Actual count.

Item set	Bucket count	Actual count
1,6	5	2
5,6	5	3
3,4	4	4
3,5	5	2
6,7	5	2
1,2	2	2
4,5	3	2
2,3	4	3

Minimum support count =2

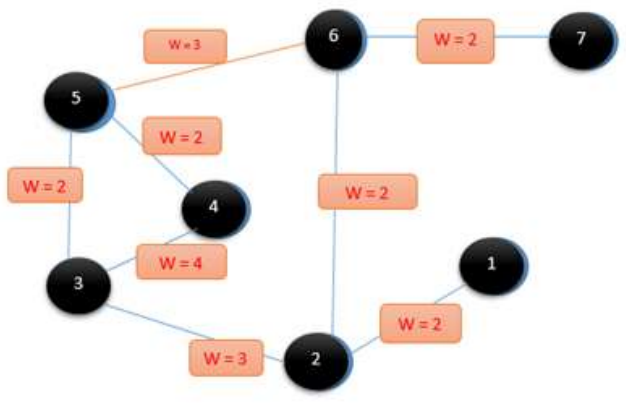

Fig 3: Graphical theoretical approach is identifying the longest path between agent and integrator.

DOI : 10.35291/2454-9150.2019.0087

IV. Experimental Analysis

In this system we are using 250-300 number of transactions of synthetic data set.

The hash base technique considers the first method of calculate hash function. The hash base technique calculative using hash functions and then shows the above result. It is created by a bucket table and it applying the minimum support count. Less than two bucket count will be rejected, next generate the item set and support count. Frequent two items are the output of hash based technique.

The output of the hash base technique is input to the graphical theoretical approach. This method can create a graph with number of transaction and data set size. This graph identifies the longest path and its suspicious show in previous paper. But now we are enhancing the graphical theoretical technique. The longest path is not suspicious. The longest path is compared with the previous year's data and put on some threshold 30% or 40%. It depends on type of the business. Here more than 40% exceeds, it should be a suspicious and the value of the threshold depends on type of the business and specific scenarios.

Money Laundering Suspicious Accounts Prediction

- Predict Longest Path

Source 1001
Destination 1006

Possible Paths and Longest Path (Directed Acyclic Graph)

Paths
1001, 1002,1003,1004,1006 (Amt: Rs.616000)
1001, 1002, 1003,1004,1005,1006 (Amt: Rs.616 000)
1001, 1002,1003,1005,1006 (Amt: Rs.400000)
1001,1002,1003,1006 (Amt: Rs.286000)
Longest Path: 1001, 1002,1003,1004,1006 (Amt: Rs.616000)
Previous Transactions Details (01/01/2016-04/01/2018): Rs. 840000 --> **Legal Transactions 1001 1006**

- Predict Longest Path

Source 1001
Destination 1006
Possible Paths and Longest Path (Directed Acyclic Graph)
Paths
1001, 1002,1003,1004,1006 (Amt: Rs.840000)
1001, 1002, 1003,1004,1005,1006 (Amt: Rs.876 000)
1001, 1002,1003,1005,1006 (Amt: Rs.723000)
1001,1002,1003,1006 (Amt: Rs.546000)
1001, 1004, 1006 (Amt: Rs.180000)
1001,1004,1005,1006 (Amt: Rs.216 000)
Longest Path: 1001, 1002, 1003,1004,1005,1006 (Amt: Rs.876000)
Previous Transactions Details (8/4/2014-2/6/2017): Rs.285000,
Difference: Rs.591000 --> **Suspicious Transactions 1001 1006**

V. Conclusion

The proposed method improves the existing hash based technique which is used with graphical theoretical approach by analysing the previous year transaction data. A threshold specified is applied on the longest path to determine if a transaction is suspicious or not. Hash base technique is used to generate frequent transaction and graphical theoretical approach identifying the traversal path of the suspicious transactions, using that identifying all the possible paths between agent and integrator.

The further enhancement could be carried out in identifying clusters where suspicious accounts can be grouped and patterns can be analysed to increase the accuracy of detecting money laundering.

References

[1] Ch. Suresh, Dr. K. Thammi Reddy, N. Sweta. A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques, I.J. Information Technology and Computer Science, 2016.

[2] Claudio Alexander and Joao Balsa. Integrating client profiling is an Anti-money Laundering Multi-Agent Based System, @springer International Publishing Swizerland 2016.

[3] Nhien An Le Khan, M.Teharkechadi. Anti-money Detection for using data mining application: A case study, IEEE International conference on data mining workshops 2010.

[4] Reza Sultana, Yuen Tran Nguyen, Yang Yong, Mohammad Afghani. A New Algorithm for Money Laundering Detection Based on Structural Similarity, Department of computer science published in 2016.

[5] Bronson Duhart, Neil Hernandez-Gress, Review of The Principle Indicators and Data Science Techniques Used For the Detection of Financial Fraud and Money Laundering, International conference on Computational science and Computational Intelligence, 2016.

[6]Rafał Drez˙ewski, Grzegorz Dziuban, Łukasz Hernik, Michał Pa˛czek, Comparison of Data Mining Techniques for Money Laundering Detection System, International Conference on Science in Information, Technology (ICSITech) 2015.

[7] Denys A. Flores, Olga Angelopoulos, Richard J. Self," Design of a Monitor for Detecting Money Laundering and Terrorist Financing", International Journal of Computer Networks and Applications 2014.

DOI : 10.35291/2454-9150.2019.0087