ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

Money Laundering Detection using Data Mining

Rohini Mohite[1], Jayraj Ghelani[2], Harshal Bhitre[3], Nikit Sawant[4], Anita .A. Lahane[5]

Rajiv Gandhi Institute Of Technology

Abstract: Money has given birth to numerous types of crime, one of which is Money Laundering. Money laundering is the method by which criminals disguise the illegal origins of their wealth and protect their asset bases, so as to avoid the suspicion of law enforcement agencies and prevent leaving a trail of incriminating evidence. To deal with this issue, the Reserve Bank of India (RBI), has introduced various guidelines to identify any sort of suspicious transactions. Once identified, they are forwarded to Financial Intelligence Unit (FIU). FIU studies the transaction and verifies its authenticity. However, this process is long, time consuming and not suitable to identify a certain type of transactions that occur in the system. To overcome this problems we propose an efficient Anti Money Laundering technique which is a composition of data mining approach and rule based engine. Data mining approach makes use of frequent association mining and graph analysis whereas Rule based engine works on the principle of rules getting fired whenever a suspicious transaction comes into main memory.

Keywords: Anti-money laundering, online payment, multi-agent, data mining, rule base agent.

I. INTRODUCTION

When criminals want to use their illicit money, usually obtained from criminal activities, they first need to convert it to make it appear legitimate. Money laundering is the process of creating the appearance that large amounts of such money, originated from a legitimate source. The process converts black money into white money and thus 'launders' it. This whole process done by criminals starts with layering, or concealing the original source of money by a series of complex transactions and other book keeping techniques. Integration refers to the act of acquiring that money in purportedly legitimate means.

One of the most widely used method is smurfing, it is a process where a person breaks up large chunks of cash into multiple small deposits, and disguises it using numerous accounts to avoid suspicion. Money laundering can also be done through the use of currency exchanges, wire transfers, and "rules" or cash smugglers, these smugglers smuggle large amounts of cash across borders and deposit them in accounts abroad in countries where money-laundering enforcement is considerably lesser strict. Some more money-laundering methods involve investing in assets commodities such as gold and gems that can be easily moved to other jurisdictions, investing in real estate, gambling, counterfeiting and creating shell companies.

Although the government has become increasing vigilant in its efforts to combat money laundering over the years by passing anti-money-laundering regulations, these regulations require financial institutions to have systems in place to detect and report suspected money-laundering activities. Numerous times it has been seen that these financial institutions are involved in various scams. To trace out the dirty proceeds immediately this proposed framework aims at developing an efficient tool for identifying the accounts, transactions and the amount involved in the layering stage of money laundering.

II. AIMS AND OBJECTIVES

To identify various Money Laundering cases.

- To study present Money Laundering detection techniques used by banks and FinancialIntelligence Unit (FIU).
- To study and implement various data mining techniques to detect suspicious moneylaundering transactions.
- To compare the results obtained from various mining techniques and the best onefor Money Laundering detection.

III. EXISTING SYSTEM

The main issue lies in underestimation or overestimation of load by the planner. This can be solved if the technique has a reasonable degree of accuracy. Therefore there is an increasing need to develop an optimised and accuracy based load forecasting models to improve (minimize) the forecast error. However, load forecasting is a difficult task because the consumption is influenced by many factors, including weather conditions, vacations, economy status, and idiosyncratic

DOI 10.17148/IJARCCE.2018.746

habits of individual customers. Inaccurate load forecasts may increase operating costs. Evidently, a poor load forecast misleads planners and often results in wrong and expensive expansion plans.

IV. PROBLEM STATEMENT AND SCOPE

Today, money laundering (ML) poses a serious threat not only to the financial institutions but also to the nation. The increasing amount of lack money lead to inflation and disrupts the whole cash flow and the economy. This criminal activity is becoming more and more sophisticated and seems to have moved from the niche of drug trafficking to enhancing terrorism and surely not forgetting personal gain. Most of the financial institutions internationally have been implementing anti-money laundering solutions (AML) to put a halt to investments in fraud activities. However, traditional investigative techniques consume numerous man-hours. Recently, data mining approaches have been developed and are considered as well-suited techniques for detecting ML activities. We survey recent data mining approaches for AML. Results obtained from various data mining approaches are compared. Approaches are compared based on three factors, their time complexity, space complexity and effectiveness of ending the suspicious transactions.

PROPOSED SYSTEM

Data Base Layer: The transaction database includes the user information and transaction information
a) User information: User Information contains Name of the Account Holder, Account Id, Name of Bank, and Name of Branch , Pan Card Number, Aadhar Card Number, Ration Card Number, IFSC code and many such important details.

b) Transaction information: Transaction information contains information about Sender and receiver of money, debit, credit, particulars, Cheque No, balance, Payment Timestamp and transaction result (successful or failing, failure reason, loss by the failure, complaint and the process result).

c)AML Data Ware House: AML data warehouse contains historical data of all the user and transaction database. It may contain information stored in aggregated form.

d) Knowledge Base: Knowledge base contains rules that are inferred from data analysis unit. Interaction between Database Layer and Data Analysis layer is two wayi. e. Data Analysis Layer not only uses Database to get the information, but the conclusions and rules that are obtained or formed in Data analysis layer is stored back in knowledge base.

Data Analysis Layer: Data Analysis Layer performs analysis on the information present in transaction information database and AML data warehouse. It uses Knowledge base to make new analysis, forms new rules and stores that information back to knowledge base. Rule Based Agent works on the principle of forward chain inferencing.

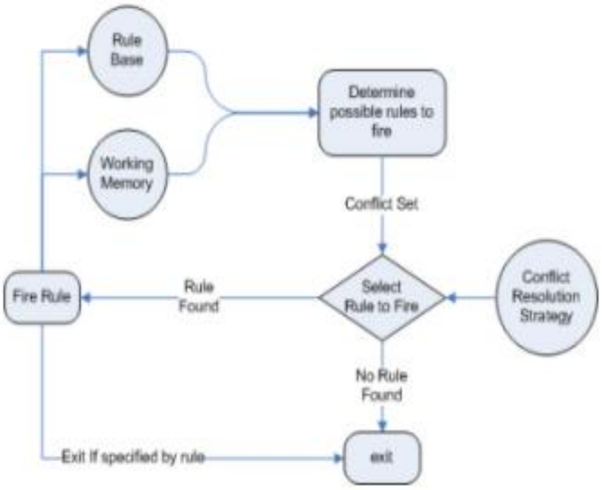

Figure 4.1: Rule Base Agent flow graph.

Application Service Layer: APIs provide an interface to the other authorities to interact with the system. API (Application Programming Interface) is nothing but a method which when called gives you the desired result. Various APIs that it can provide are as follows: Agent and Integrator Identification (): This interface will provide account Information of the Agent and Integrator. Account Information will be a structure containing user Information.

DOI 10.17148/IJARCCE.2018.746

Suspicious Transaction Identification():This interface will provide amount of money laundered. The interface will take Agents Account Id and Integrators Account Id as input parameters. Account Id of the agent and integrator can be obtained from Agent and Integrator Identification () method.

Interface Layer: Interface Layer provides interface to the user of the system. User will be provided with a dashboard. Interface Layer will provide all the information obtained from Application Service layer. Also it will show some graphs displaying some analysis.

V. METHODOLOGY

Methodology of the Data Mining Agent present inside the Data Analysis Layer is given below. Considering our two main aspects, to detect frequency of transactions and then to detect the path, It consists of two dependent modules. First one is frequent transactions detection and the other one is path detection and graph analysis module.

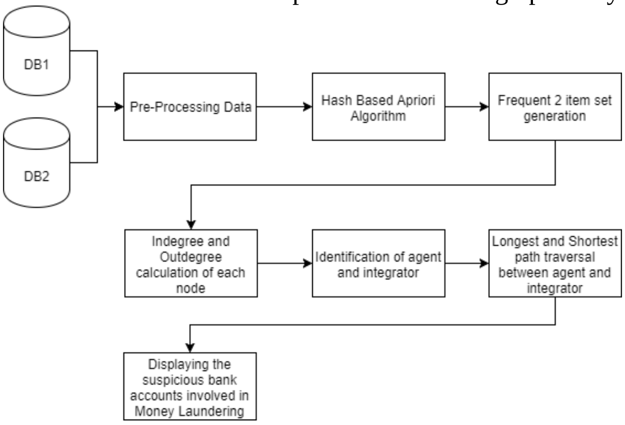
DB1
DB2
Pre-Processing Data
Hash Based Apriori Algorithm
Frequent 2 item set generation
Indegree and Outdegree calculation of each node
Identification of agent and integrator
Longest and Shortest path traversal between agent and integrator
Displaying the suspicious bank accounts involved in Money Laundering

Figure 5.1:hashing and pre-processing of data

Hash based Apriori implementation, uses a data structure that directly represents hash table. This algorithm proposes overcoming some of the weaknesses of the Apriori algorithm by reducing the number of candidate k-itemsets. In particular the 2-itemsets, since that is the key to improving performance. This algorithm uses a hash based technique to reduce the number of candidate item sets generated in the pass. It is claimed that the number of item sets in C2 generated using hashing can be reduced, so that the scan required determining L2 is more efficient. For example, when scanning each transaction in the database to generate the frequent 1-itemsets L1, from the candidate 1-itemsets in C1, we can generate all of the 2- item sets for each transaction, hash(I . e) map them into the different buckets of a hash table structure, and increase the corresponding bucket counts . A 2-itemset whose corresponding bucket count in the hash table is below the support threshold cannot be frequent and thus should be removed from the candidate set. Such a hash based Apriori may substantially reduce the number of the candidate k. Algorithm:

1. Scan all the transaction. Create possible 2-itemsets.
2. Decide a Hash Table of a particular size.
3. For each bucket assign a candidate pair using the ASCII value or Number associated with each item set.
4. Each bucket in the hash table has a count, which is increased by 1 each item an item-set is hashed to that bucket.
5. If the bucket count is equal or above the minimum support count, the bit vector is set to 1. Otherwise it is set to 0.
6. The candidate pairs that hash to locations where the bit vector bit is not set are removed.
7. Modify the transaction database to include only these candidate pairs. In this algorithm, each transaction counting all the 1- item-sets. At the same time all the possible2-itemsets in the current transaction are hashed to a hash table. It uses a hash table to reduce the number of candidate item sets. When the support count is established the algorithm determines the frequent item sets. It generates the candidate item sets as like the Apriori algorithm. But it is efficient in terms of space complexity. Frequent 2 item sets which are obtained after implementing the above method is then further used for finding relationship between frequent transactions. Graph Analysis Methodology.

1. Establishing a complete graph by means of frequent 2 item set obtained from hash based Apriori approach.
2. Edges between the graph can represent various attributes, like frequency of transaction between two accounts or the amount of money transferred between two accounts.
3. Finding out the suspicious path involves finding the agent and integrator at the layering stage.

DOI 10.17148/IJARCCE.2018.746

4. In order to do so, in degree and out degree of a node in the graph is found out. In a directed graph with vertices labelled (in-degree, out-degree). For a vertex, the number of head ends adjacent to a vertex is called the in-degree of the vertex and the number of tail ends adjacent to a vertex is its out-degree (called "branching factor" in trees).
5. Agent is a node with in-degree = 0 and integrator is a node with out-degree = 0. Once the agent and integrator are identified. Longest/Shortest Path traversal algorithm can be applied between the two nodes and find the suspicious account involved in money laundering.
6. The longest path problem for a general graph is not as easy as the shortest path problem because the longest path problem doesn't have optimal substructure property. In fact, the Longest Path problem is NP-Hard for a general graph. However, the longest path problem has a linear time solution for directed acyclic graphs. The idea is similar to linear time solution for shortest path in a directed acyclic graph, topological Sorting used.

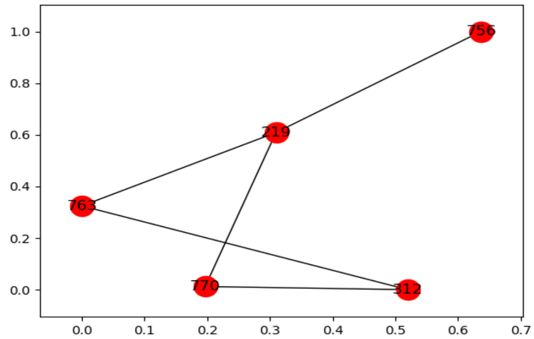

Figure 5.2:Generation of graph

VI. FUTURE SCOPE

Scope of the project is restricted to banking transactions wherein we consider only debait and credits made to an account. We don't take into consideration any other kind of banking transaction. With the chaining of accounts, we can further develop a system which identities the sure relation between these identified suspicious account using concepts like ontology. The relation between these accounts can give us additional information like whether the involved criminal people are belonging to same occupation or to the same location etc. The frequent accounts should not be the only criteria for ending out the suspicious transaction as there may be a case when the transaction does not occur frequently but even then they are illegal. To trace out such cases additional parameters have to be considered.

VII. CONCLUSION

This proposed system improve the efficiency of the existing anti money laundering techniques by identifying the suspicious accounts in the layering stage of money laundering process by generating frequent transactional datasets using Hash based Association mining. The generated frequent datasets will then be used in the graph theoretic approach to identify the traversal path of the suspicious transactions. We were successful in finding the agent and integrator in the transaction path. In our solution, we have considered the frequent accounts as the parameter and have obtained a chaining of accounts. These accounts have the highest possibility of being suspicious as there are involved in huge amount of transactions frequently. Also, Rule-based AML systems have replaced by artificial intelligent approach for AML and are found to better in terms of accuracy and time efficiency. The solution proposed here is highly advantageous over the existing anti-money laundering rules.

REFERENCES

[1] Ch. Suresh, Dr. K. Thammi Reddy, N. Shweta, A Hybrid Approach for Detecting Suspicious Accounts in Money Laundering Using Data Mining Techniques Proceedings, 2013.
[2] Ch. Suresh, Dr. K. Thammi Reddy, N. Shweta, Suspicious transaction detection for Anti Money Laundering Proceedings, 2013.
[3] Ma Bin, Analysis on the influence of the electronic money to the anti-money laundering and its solution, Fujian Financial Proceedings, 2006.
[4] Nhien-An Le-Khac, Sammer Markos and M-Tahar Kechadi, Proceedings of the 28th Annual ACM Symposium on Applied Computing, ACM New York, NY,USA , pp. 1852-1858 March 2013.
[5] Yang Qifeng, Feng Bin, Song Ping, A Heuristics Approach for Fast Detecting Suspicious Money Laundering Cases in an Investment Bank Proceedings, 2012.