Detecting Suspicious Accounts in Money Laundering

Supreeth T S[1], Sanath H S[2], Satish[3], Manoj Gowda C R[4], Padmini M S[5]

[1,2,3,4]BE, Department of CSE, NIE Mysore, Karnataka, India
[5]Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India

---***---

Abstract – *Money laundering could be a criminal activity to disguise black money as white money. It is a method by that outlawed funds and assets square measure born-again into legitimate funds and assets. Money Laundering happens in 3 stages: Placement, Layering, and Integration. It results in numerous criminal activities like Political corruption, smuggling, monetary frauds, etc. In Republic of India there's no sure-fire opposed concealment techniques that square measure accessible. The Federal Reserve Bank of Republic of India (RBI), has issued tips to spot the suspicious transactions and send it to monetary Intelligence Unit (FIU). FIU verifies if the dealing(transaction) is truly suspicious or not. This method is time intense and not appropriate to spot the outlawed transactions that happens within the system. To overcome this drawback we have a tendency to propose associate economical opposed concealment technique which might ready to determine the traversal path of the Laundered cash victimization Hash based mostly Association approach and sure-fire in distinctive agent and measuring instrument within the layering stage of cash wash by Graph theoretical Approach.*

Key Words: Data mining, Anti Money Laundering, FIU, Hash Based Mining, and Traversal Path.

1. INTRODUCTION

Money laundering is a process of converting unaccountable money in to accountable money. Day to day the technology is obtaining updated and during this quick dynamical technology several deserves similarly as demerits are associated. With the arrival of E-Commerce the globe has been thus globalized and additional the technology has created everything thus user friendly that with one click of a button, many transactions can be performed. Fraud Detection is necessary since it affects not solely to the financial organisation however additionally to the whole nation. This criminal activity is showing additional and additional refined and maybe this can be the main reason for the problem in fraud detection. This criminal activity results in numerous adverse effects starting from drug traffic to monetary coercion. Traditional investigative techniques consume numerous man-hours. Data Mining is within which large amounts of knowledge are analysed in numerous dimensions and angles and additional classified then eventually summarized in to helpful information. Data Mining is that the method of finding correlation or patterns among dozens of fields in giant databases. The governing bodies like Reserve Bank of India, Securities and Exchange Board of India have listed out various guidelines to the financial institutions. All the banks collect the list of transactions which is not in accordance with the Reserve Bank of India (RBI) and then submit it to Financial Investigation Unit (FIU) for further investigation. The FIU identifies the cash washing method from the applied math data obtained from numerous banks. This method is changing into additional and additional difficult since the count of suspicious transactions is increasing well and therefore the rules obligatory by tally alone isn't ample to observe this criminal activity. The 3 stages of cash washing embrace Placement, Layering and Integration. The placement stage is that the stage wherever within the actual criminal person disposes all the felonious money to a broker. This broker or agent is answerable for distributing cash. In the layering stage the money is unfold into multiple intermediaries which will embrace banks and different financial organisation.

The major issue lies during this layering stage of cash washing as a result of here the transfer of cash is also from one to at least one or one-to-many .The difficulty arises in tracing out all the chaining of transactions. In the integration stage all the money is transferred to a beneficiary typically known as planimeter.

At this stage all the transactions are created legal. 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. The rest of the paper is organized as follows; the literature survey is presented in section-2 of the paper, section3 of the paper deals with the proposed method. Section four the experimental analysis and in section five the conclusion and additional improvement has been explained.

2. LITERATURE SURVEY

R.cory Watkins et al. [4] has mentioned that from the time of layering stage criminals tries to faux that laundered cash feels like as funds from legal activities which can't be differentiated unremarkably. Traditional investigatory approaches use to uncover concealing patterns may be weakened into one amongst 3 categories: Identification, detection shunning and superintendence.

Nhien An Le Khac et al. [1][2][3] created a knowledge mining based mostly solutions for examining transactions to find concealing And prompt an investigation method supported completely different data processing techniques like call tree, genetic algorithm and fuzzy clustering. By merging natural computing techniques and data processing techniques; data based mostly answer were projected to find concealing. Different approaches were proposed for quick identification of customers for the purpose of application of Anti money laundering. In their paper implemented an approach where in they determined the important factors for investigating money laundering in the investment activities and then proposed an investigating process based on clustering and neural network to find suspicious cases within the context of cash wash. In order to enhance time period heuristics like suspicious screening were applied.

3. PROPOSED SYSTEM

Identifying concealing is incredibly troublesome task thanks to Brobdingnagian variety of transactions were concerned. To overcome this drawback we have a tendency to propose a technique that makes use of Hash primarily based association mining for generating frequent transactional datasets and a graph suppositional approach for distinctive the traversal path of the suspicious transactions, exploitation that all doable methods between agent and planimeter area unit known. This Graph suppositional approach looks to be attention-grabbing, as a result of they'll find advanced dependencies between transactions. It is also possible to take into account properties and relations of entities involved in sending and receiving the transfers.

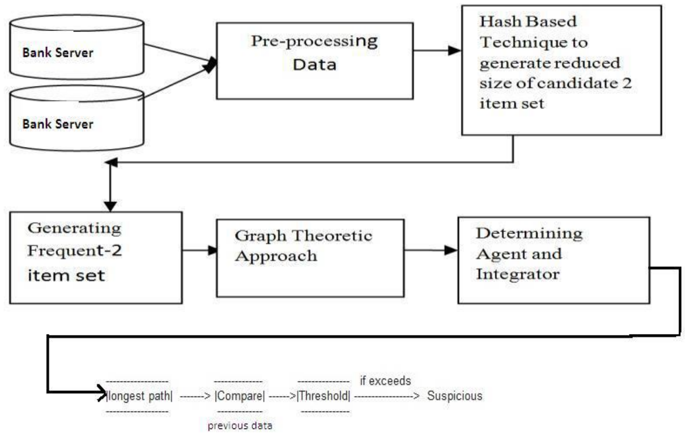


Fig.1. Proposed System Model for detecting money laundering using Hash Based Association Mining.

The projected system uses Hashing Technique to get frequent accounts.A synthetic transactional info was accustomed experiment the projected methodology a similar situation that is comparable to the current industry is taken into account every individual bank's knowledge that is stored in Databases say Data base1, Database 2...and so on are taken together and combined to form a single large database. Now the data of this large database has to be pre-processed in order to obtain data which is free from all null and missing or incorrect values.

A hash based mostly technique is applied on the transactional dataset to get a candidate set of reduced size .From this reduced size of candidate set we have a tendency to acquire Frequent-2 item set. This frequent-2 item sets additional forms the perimeters and nodes of the graph. On applying the 'Longest Path during a directed acyclic graph' formula we have a tendency to acquire the trail during which great amount has been transferred. On the idea of in-degree and out-degree of every node, we determine agent and integrator.

The proposed system is represented in two steps

Step1: Applying hash based mostly technique to get frequent two item set.This technique is employed to cut back the candidate k-items, Ck, for k >1.

The formula for hash function used for creating Hash Table is

$$h(x, y) = ((order\ of\ x * 10) + order\ of\ y)\ mod\ 7$$

Step2: characteristic suspicious transactions path mistreatment graph metaphysical approach

• Linking all the transactions consecutive and generating a graph by considering every account within the frequent item set as a node.

• for every link between the dealings, assign weights to reflect the multiplicity of the occurrence and hence the strength of the path.

• Finding the in-degree and out-degree of every node and determining agent and integrator.

The Hashing Technique that is adopted however a similar unvaried level wise approach of apriori formula is followed. This means that K-item sets square measure accustomed explore (k+1) item sets.

Hash Based Technique over Apriori Algorithm:

A hash based mostly technique will be wont to cut back the scale of the candidate k-item sets, Ck, for k>1. This is as a result of during this technique we have a tendency to apply a hash perform to every of the item set of the dealings.
Suppose for equation (1), we have an item set , then x=1 and y=4.Hence h (1, 4) = ((1*10) +4) mod 7=14 mod 7=0. Now we place in bucket address 0. Likewise we have a tendency to fill the hash table and record the bucket count. If any bucket has count but the minimum support count, then that whole bucket (i.e. its entire contents) is discarded) All the

undeleted bucket counts now form elements of candidate set. Thus currently we've a candidate item set that is smaller in size and thus we want to scan the info less range of times to search out the frequent item sets thereby up the potency of apriori formula.

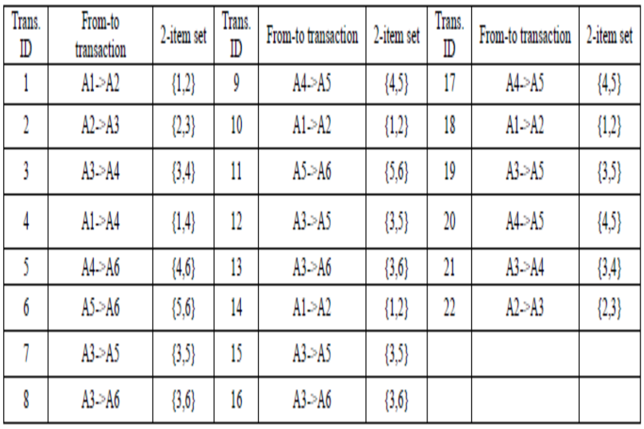
Thus now we have a candidate item set which is smaller in size and thus we want to scan the info less range of times to search out the frequent item sets thereby up the potency of apriori formula. Candidate 2-item set generation: All the contents of the undeleted hash table contents are derived so the duplicate transactions are eliminated.

Then we obtain candidate 2 item set.

Transitivity relation: As at a time only 2 accounts are involved in a transaction, to find the chaining of accounts, we have used the mathematical transitivity relation, i.e., if A->B and B->C, then A->B->C.

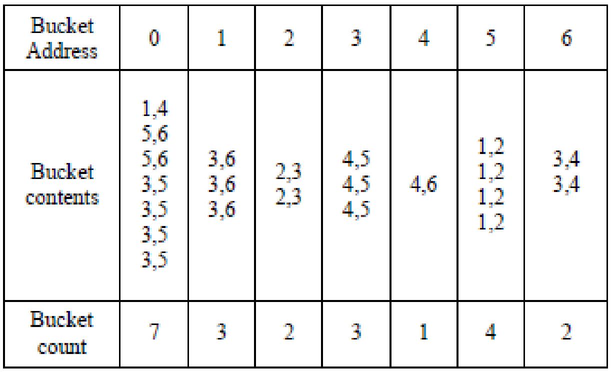
Frequent three item sets: From the transitivity relation we have a tendency to get three item sets.
These item sets have the quantity related to it.

Table 1. Generation of 2 item set using hash based approach



Trans. ID	From-to transaction	2-item set	Trans. ID	From-to transaction	2-item set	Trans. ID	From-to transaction	2-item set
1	A1->A2	{1,2}	9	A4->A5	{4,5}	17	A4->A5	{4,5}
2	A2->A3	{2,3}	10	A1->A2	{1,2}	18	A1->A2	{1,2}
3	A3->A4	{3,4}	11	A5->A6	{5,6}	19	A3->A5	{3,5}
4	A1->A4	{1,4}	12	A3->A5	{3,5}	20	A4->A5	{4,5}
5	A4->A6	{4,6}	13	A3->A6	{3,6}	21	A3->A4	{3,4}
6	A5->A6	{5,6}	14	A1->A2	{1,2}	22	A2->A3	{2,3}
7	A3->A5	{3,5}	15	A3->A5	{3,5}			
8	A3->A6	{3,6}	16	A3->A6	{3,6}			

Table 2. Bucket table with bucket counts



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

Enter the minimum bucket count (say 2).

Then all the bucket whose total count is a smaller amount than the minimum bucket are deleted with all its contents. Here bucket 4 is deleted.

Generating a sequent traversal path: From the frequent accounts, we are able to produce the perimeters of the graph and conjointly the load of every edge is up to the quantity transferred between those two accounts.

Longest path during a directed acyclic graph: There square measure several methods within the graph.
Now to find the most suspicious path, we are applying this algorithm and getting the path with the total amount.

The entire implementation will be understood by considering a tiny low example of twenty 2 transactions.

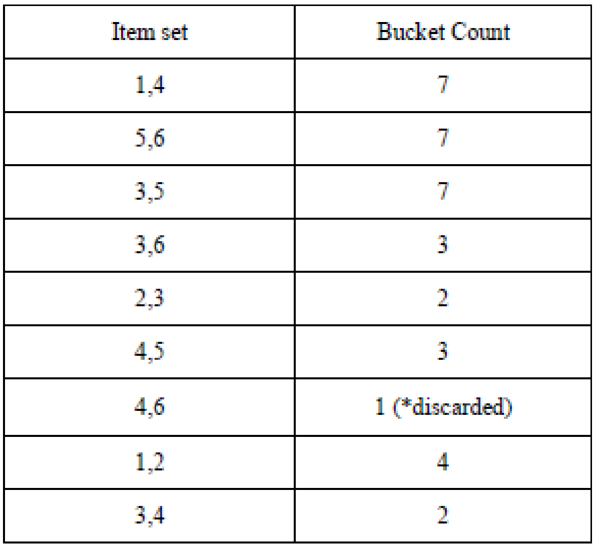
- Step 1: Generating frequent accounts victimization hashing: think about a tiny low dealing dataset of twenty-two transactions.

On this set of twenty-two transactions hash formula is applied equ (1) is applied.

Here x= from_acc_id and y=to_acc_id
Now all these 22 transactions are grouped in to different indexes in hash table.
Now the bucket count is calculated for every bucket.

Minimum Bucket Count=2

Table 3. Bucket table for item sets and minimum support count



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

Now the left over dealings in the buckets are taken and then their actual count in the database is recorded.

Table 4. The bucket count and actual count are recorded

Item sets	Bucket Count	Actual Count
1,4	7	1 (*discarded)
5,6	7	2
3,5	7	4
3,6	3	3
2,3	2	2
4,5	3	3
1,2	4	4
3,4	2	2

Minimum Support Count =2

Now all the dealings which have occurred 2 or more no of times are taken in to Frequent -2 item sets. Thus obtained frequent-2 Transactions(dealings) are:

Table 5.Frequent 2 accounts with their support count

Frequent-2 Item set	Support Count
5,6	2
3,5	4
3,6	3
2,3	2
4,5	3
1,2	4
3,4	2

- Step 2: *Finding the traversal path*

Various paths are identified by connecting all frequent accounts as nodes.

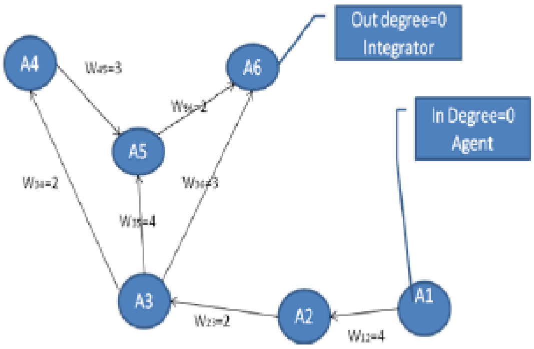


Fig.2. Identifying agent and integrator using graph theoretic approach

Enhancement

The algorithm is further enhanced by considering previous year's transactions data. This is done by comparing the obtained total amount transacted along the longest paths with the previous year's transacted amount along respective paths .

Set a threshold percentage hike for eg: say threshold hike = 40%
in the transacted amount along the paths. If there is a significant hike in the amount transacted and exceeds the threshold hike percentage then that longest path is considered to be the suspicious path and the accounts involved in that path are considered suspicious accounts.

Experimental Analysis

The proposed system is accessed using a 17000. We have taken an artificial dealing dataset from multiple banks over a amount of one hundred twenty days. We think about here dealing datasets up to seventeen thousand sizes. For different dataset we select different tables and enter different values of support/threshold count. We first apply hashing technique to come up with frequent two item sets.

Table 6. Experiments carried over different Data sets

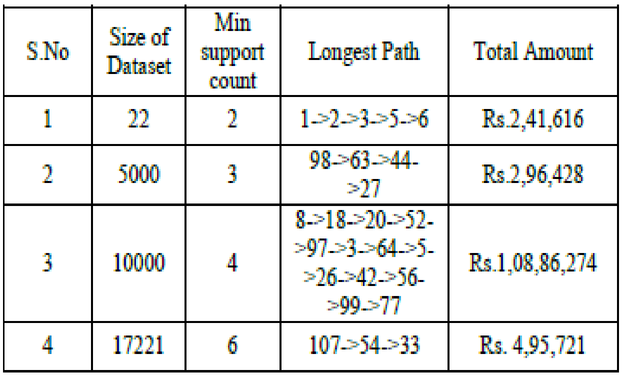
S.no	Size of dataset	Minimum bucket count	Minimum support count	No of frequent transactions
1	22	2	2	7
2	5000	700	3	77
3	10000	1400	4	103
4	17221	2500	6	36

This is the result obtained when applying Hashing Technique.
We observe that when the no of transactions in the data set increases, the no of frequent account also increases. However if we have a tendency to increase the support count price to an outsized price then a couple of no of frequent transactions are obtained as within the last case of our experiment.

The algorithmic rule to search out the longest path during a directed acyclic graph is applied on these frequent transactions in every of the higher than four cases.
We get the following results. Similarly we will realize a path between any 2 frequent accounts mistreatment this algorithmic rule.

For Eg: we will realize the trail between thirty one and forty nine in ten thousand dataset group action giving minimum bucket count as 1400 and minimum support count as four.
31->96->86->36->49.

Table 7.Identifying the longest path by varying the size of datasets and support count



S.No	Size of Dataset	Min support count	Longest Path	Total Amount
1	22	2	1->2->3->5->6	Rs.2,41,616
2	5000	3	98->63->44->27	Rs.2,96,428
3	10000	4	8->18->20->52->97->3->64->5->26->42->56->99->77	Rs.1,08,86,274
4	17221	6	107->54->33	Rs. 4,95,721

Enhancement done to the IEEE paper :

By comparing the total amount with previous year's data and checking for a significant hike , there is much improvement in the results obtained as the suspicious accounts are more likely to be involved in such paths.
By the above enhancement we are further narrowing it down and there is much better chances that suspicious accounts are found out.

4. CONCLUSIONS

The projected system improve the potency of the prevailing anti money laundering techniques by distinguishing the suspicious accounts within the layering stage of cash laundering method 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 to find the agent and measuring device within the group action path. In our answer, we have considered the frequent accounts as the parameter and have obtained a chaining of accounts. These accounts have the very best chance of being suspicious as there square measure concerned in large quantity of transactions often. The solution projected here is extremely advantageous over the prevailing anti-money lavation rules.

The algorithm is further enhanced by considering previous year's transactions data. This is done by comparing the obtained total amount transacted along the longest paths with the previous year's transacted amount along respective paths . As a result of this enhancement done , the user of this system can be almost as sure whether the money laundering is happening or not along the transacted paths.

Further enhancement: The frequent accounts mustn't be the sole criteria for locating out the suspicious dealing as there is also a case once the dealing doesn't occur often but even then they are illegal. To trace out such cases further parameters ought to be thought of.

REFERENCES

[1] NhienAn Le Khac, SammerMarkos, M. O'Neill, A. Brabazon and M-TaharKechadi. An investigation into data processing approaches for anti concealment. In International conference on pc Engineering & Applications 2009.

[2] Nhien An Le Khac, M.Teharkechadi. Application of information mining for Anti-money Detection: A case study. IEEE International conference on Data mining workshops 2010.

[3] Nhien An Le Khac, SammerMarkos,M.Teharkechadi,. A data mining based mostly answer for sleuthing suspicious concealment cases in AN investment bank. IEEE Computer society 2010.

[4] R.Corywatkins, K.Michaelreynolds, Ron Demara. Tracking Dirty Proceeds: Exploring data processing Techniques on Investigate concealment. In police practice and research 2003.