International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 5, May-2014, pp: (291-295), Impact Factor: 1.252, Available online at: www.erpublications.com

Identifying Money Laundering Groups in Multi Mode Network Using Data Mining

G. Krishnapriya[1], Dr. M. Prabakaran[2]
[1]M.C.A., M.Phil, Research Scholar, Bharathidasan University, Trichy, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science Government Arts College, Ariyalur, Tamil Nadu, India

Abstract: The growth of internet technology opens the gate for money laundering where the internet user can easily transfer enormous amount to any suspicious account in the world. The user need no confirmation or approval from the destination account holder and can transfer the amounts instantly which can travel through set of account and become untraceable at the end. There exist many approaches but suffers with the accuracy of money laundering identification. We propose a new technique which identifies the group of accounts which involved in money laundering and identify the source account also. The proposed method uses multi mode network, where each account is considered as a node and the attributes, beneficiary becomes the mode of network. We cluster the accounts with more similar properties and links to form a group, from which we identify the set of accounts involved in money laundering. Earlier the money laundering approaches involved in identifying only the source or destination account but the proposed method identifies the money laundering groups.

Key Terms: Multi mode networks, Money Laundering, Clustering, Data Mining.

1. INTRODUCTION

Similar to social networks, the transactional data set can be considered as multi mode networks, where each account has various properties like list of beneficiary , list of accounts linked, location and etc. Each property of the accounts plays a vital role in forming groups between accounts. For example, a account holder may have N number of accounts linked of his own and has many number of other user accounts linked to his accounts and so on. Similarly the money laundering groups has many relations and transaction between them. The problem is the dimension of trace and number of traversal occurring between source and destination, so that the source of laundering could not be identified. Money laundering is the process of sending and receiving money between accounts which is unaccountable. The money laundering affects the country economy and stability of banking sectors heavily. For instance, a malicious user can send enormous amount to one destination account of any bank and at some time he may transfer the complete fund to some other account without any reason. What the banking organization will suffer with this is, the share or value of the bank will go down instantly and the investing companies or persons will question them. Also the same fund may be used for some terrorist activities later. Identifying money laundering is the most important task for the enforcement directors and finance ministry also. Through money laundering, criminals try to convert monetary proceeds derived from illicit activities into “clean” funds using a legal medium such as large investment or pension funds hosted in retail or investment banks. This type of criminal activity is getting more and more sophisticated and seems to have moved from the cliché of drug trafficking to financing terrorism and surely not forgetting personal gain. Today, ML is the third largest “Business” in the world after Currency Exchange and Auto Industry.

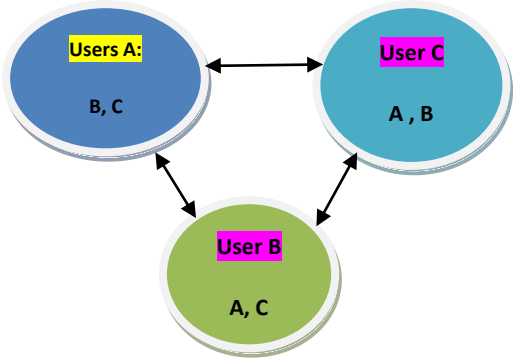


Figure 1: Shows sample scenario of account network

The figure 1, shows the account graph, where account A,B,C has linked to each other. Similarly the accounts and transactions of the bank can be constructed as a graph or network where each node can be connected with other nodes only if they have linked in any way. Data mining can be applied for ML identification using various approaches. The process of mining information from large set of transactional data set would help the task and we propose such a methodology using which we can identify the set of malicious accounts. The nodes of the network can be grouped or clustered by means of relations. The relation between any two node is achieved by computing number of transactions and beneficiary present in between them. The clustering approach is used to identify money laundering groups of the network and can easily trace the malicious accounts

.

2. RELATED WORKS

There has been many number of approaches discussed earlier and we discuss few of them here according to our problem statement. Statistical Methods for Fighting Financial Crimes [2], focuses on two important types of financial crimes: fraud and money laundering. It discusses some of the traditional statistical techniques that have been applied as well as more recent machine learning and data mining algorithms. The goal of the article is to introduce the subject and to provide a survey of broad classes of methodologies accompanied by selected illustrative examples. Laundering Sexual Deviance: Targeting Online Pornography through Anti-money Laundering [3], concentrates on cyber-pornography/obscenity, which encompasses online publications or distribution of sexually explicit material in breach of the English obscenity and indecency laws. After examining the major deficiencies of the attempts to restrict illegal pornographic representations, the authors aim to highlight that the debate regarding their availability in the Internet era neglects the lucrative nature of the circulation of such material, which can be also targeted through anti-money laundering. Rising profits fuel the need to recycle the money back into the legal financial system, with a view to concealing their illegal origin. Anti-money laundering laws require disclosure of any 'suspicion' related to money laundering, thus opening another door for law enforcement to reach the criminal.

Combining Digital Forensic Practices and Database Analysis as an Anti-Money Laundering Strategy for Financial Institutions [4], propose an anti-money laundering model by combining digital forensics practices along with database tools and database analysis methodologies. As consequence, admissible Suspicious Activity Reports (SARs) can be generated, based on evidence obtained from forensically analyzing database financial logs in compliance with Know-Your-Customer policies for money laundering detection. Event-based approach to money laundering data analysis and visualization [6], proposes crime specific event patterns are crucial in detecting potential relationships among suspects in criminal networks. However, current link analysis tools commonly used in detection do not utilize such patterns for detecting various types of crimes. These analysis tools usually provide generic functions for all types of crimes and heavily rely on the user's expertise on the domain knowledge of the crime for successful detection. As a result, they are less effective in detecting patterns in certain crimes. In addition, substantial effort is also required for analyzing vast amount of crime data and visualizing the structural views of the entire criminal network. In order to alleviate these problems, an event-based approach to money laundering data analysis and visualization is proposed in this paper.

A Fistful of Bit coins: Characterizing Payments Among Men with No Names [7], explore this unique characteristic further, using heuristic clustering to group Bit coin wallets based on evidence of shared authority, and then using re-identification attacks (i.e., empirical purchasing of goods and services) to classify the operators of those clusters. From this analysis, we characterize longitudinal changes in the Bit coin market, the stresses these changes are placing on the system, and the challenges for those seeking to use Bit coin for criminal or fraudulent purposes at scale. Zerocoin: Anonymous Distributed E-Cash from Bitcoin [9], uses standard cryptographic assumptions and does not introduce new trusted parties or otherwise change the security model of Bitcoin. We detail Zerocoin's cryptographic construction, its integration into Bitcoin, and examine its performance both in terms of computation and impact on the Bitcoin protocol.

A Framework on Developing an Intelligent Discriminating System of Anti Money Laundering [15], based on support vector machine (SVM) in order to take the place of traditional predefined-rule suspicious transaction data filtering system. It could efficiently surmount the worst forms of suspicious data analyzing and reporting mechanism among bank branches including enormous data volume, dimensionality disorder with massive variances and feature overload. Money Laundering Detection using Synthetic Data [17], present an analysis of the difficulties and considerations of applying machine learning techniques to this problem. We discuss the pros and cons of using synthetic data and problems and advantages inherent in the generation of such a data set. We do this using a case study and suggest an approach based on Multi-Agent Based Simulations (MABS). All these above discussed approaches has concentrated on identifying money laundering and we propose a new method which is intended to identify the source and groups of money laundering.

3. PROPOSED METHOD

The proposed multi mode money laundering identification framework has three phases namely Preprocessing, MultiMode Clustering, Money Laundering Group Evaluation. At the first phase i.e. in preprocessing, the transactional data are converted into the form of multi mode network , at the second phase the network is clustered according to their similarity and finally money laundering has been identified.

3.1 Preprocessing:

The input transactional data set Ts, is preprocessed to identify missing items or noisy data points. If there exist any noisy records then that will be excluded from further processing and removed. Then, we identify distinct actors or attributes of the transaction, i.e. we point attribute as the account numbers, transaction type, date , branch, amount, and etc. The proposed method creates a node for each of the distinct account identified and initialize with the properties of its own. Second, each node is identified with set of accounts it has linked with and it creates the link between the nodes.

Input: Transactional Data Set Ts.
Output: Network Graph NG.
For each transaction T_i from Ts
If any missing items Then
Remove T_i from Ts =Ts × Ti
end
Identify Number of modes M.
Create Node N_i = {AccNumber, AccName, Benificiaries, amounts, branch, ttype}.
Identify set of Nodes has linked and create connections.
$N_i(links)= \sum_{i=0}^{i=n} \in Nj.Benificiaries$
end.

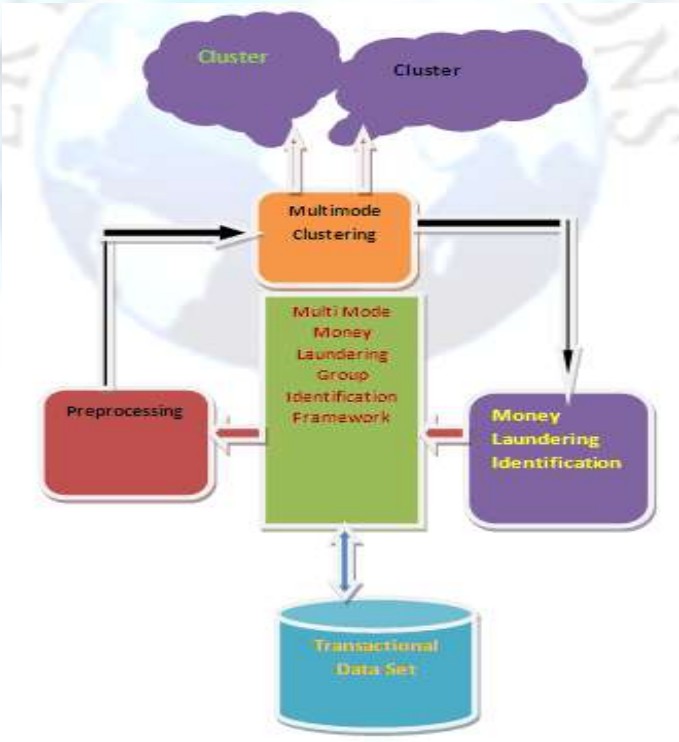


Figure 2: Proposed system architecture.

3.2 Multimode Clustering:

Multimode clustering is performed on the preprocessed transactional data set to group similar accounts and set of linked accounts and so on. For each Node of the network , we identify the presence of any of the beneficiary in its list and if so we

include those node also into the group to form the initial cluster. Now we have set of nodes to form the cluster, here even if there is no transaction between any two nodes and the third node , then also they were brought into the group. This will be removed at the next level of the clustering. The cluster validation is performed to remove the unwanted nodes from the groups. For each of the transaction, we start from first node and identify the sequence of nodes passed through the transaction. Before moving to the next transaction , we identify whether there exist any other transaction which has pass through the account. This will be iterated for each distinct source account and will form separate cluster for each source account.

Input: Preprocessed Ts.
Output: Clusters Cs.
Initialize Clusters Cs.
Select Distinct Source Accounts SA.
For each SA_i of SA
Identify first target accounts Ta = $\sum_{i=0}^{i=N}$ Ti ∈ Ta
Identify next level accounts and transactions T_n = $\sum_{i=0}^{i=N}$ Ti ∈ Tn
Identify presence of transaction to pass through.
if yes then
Add to cluster CS_i = Ta+Tn.
end
end.

3.3 Money Laundering Group Evaluation:

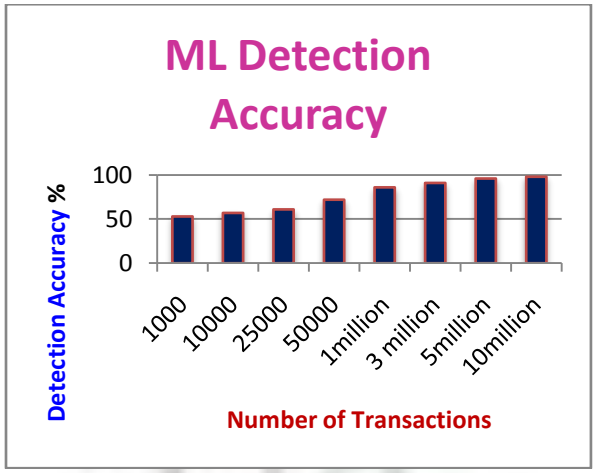
The money laundering identification and group evaluation is performed using the clusters generated at the previous stage. For each transaction T_i, we identify and check set of accounts the amount pass through from the cluster. We first identify the cluster where the source account resides, from that cluster we identify the accounts through which the fund transferred. We compute the laundering weight for each of the destination account , over each distinct middle node will be computed weight against transaction. The weight is computed based on Total amount transferred towards a single account, Number of transactions for each distinct account participated. We compute the cumulative weight by summing all the intermediate weights for distinct destination account. If the overall amount transferred and frequency is higher then , the particular account and the source account will be considered as performed money laundering. The other nodes in the participating list , which has weights more than a threshold will be considered as group node and the group is named as money laundering groups.

Input : Multi Mode Cluster Cs.
Output:Money Laundering groups MLG.
For each transaction T_i
Identify accounts pass through Ng = Σaccounts∈ Ti
For each N_i of Ng
Compute laundering weight lw =(NT/(T_i∈ Ni)× (NT/(ΣT_i(Amounts)>ATh)).
ATh – Amount Threshold.
if lw> LTh then
Add to list Al.
End.
If $T_{i(Amount)}$> ATh Then
Cs= ΣCs+ N_i
end.

End.

4. RESULTS AND DISCUSSION

The proposed method has been evaluated using various transactional set collected from different banking sectors and we have separated the accounts which are linked through different banks. Finally we have collected 5000 accounts from different banks having 10 million transactions. The proposed method has produced efficient results and detection accuracy is also higher.



Graph 1: shows the efficiency of identifying money laundering

The graph1 shows the efficiency of identifying money laundering with respect to number of transaction used. It is clear that the efficiency is increased if the size of transaction is increased. The proposed methodology produces efficient result by increasing the size of transaction.

CONCLUSION

We analyze various methodologies to identify money laundering crime. We identify that all methods have scalable in accuracy and efficiency. We proposed a multi mode network clustering model which uses time variant transactional data. The proposed method has produced higher efficient results and with accurate findings. The proposed method has produced results with less time complexity.

REFERENCES

[1]. Diane J. Cook, Lawrence B. Holder, Jeff Coble and Joseph Potts. (2005). Graph-based Mining of Complex Data. Advanced Methods for Knowledge Discovery from Complex Data, Springer, 2005, Part I, pp.75-94.
[2]. Agus Sudjianto, Sheela Nair, Ming Yuan, Aijun Zhang, Daniel Kern, and Fernando Cela-Daz. Statistical Methods for Fighting Financial Crimes. Technometrics, 52(1):5{19, February 2010.
[3]. Odense, Laundering Sexual Deviance: Targeting Online Pornography through Anti-money Laundering, European Intelligence and Security Informatics Conference, 2012.
[4]. Bucharest, Combining Digital Forensic Practices and Database Analysis as an Anti-Money Laundering Strategy for Financial Institutions, Third International Conference on Emerging Intelligent Data and Web Technologies, 2012.
[5]. Jason Hong, The State of Phishing Attacks, Communications of the ACM, Vol. 55 No. 1, Pages 74-81, 2012.
[6]. Tat-Man Cheong, Event-based approach to money laundering data analysis and visualization, Proceedings of the 3rd International Symposium on Visual Information Communication, ACM , 2010.
[7]. Sarah Meiklejohn, A Fistful of Bitcoins: Characterizing Payments Among Men with No Names, ACM 2013.
[8]. E. Androulaki, G. Karame, M. Roeschlin, T. Scherer, and S. Capkun. Evaluating User Privacy in Bitcoin. In Proceedings of Financial Cryptography 2013, 2013.
[9]. I. Miers, C. Garman, M. Green, and A. D. Rubin. Zerocoin: Anonymous Distributed E-Cash from Bitcoin. In Proceedings of the IEEE Symposium on Security and Privacy, 2013.
[10]. T. Moore and N. Christin. Beware the Middleman: Empirica Analysis of Bitcoin-Exchange Risk. In Proceedings of Financial Cryptography 2013, 2013.
[11]. Tao Jiang and Ah-Hwee Tan. (2005). Ontolopy-Assisted Mining of RDF Documents. Advanced Methods for Knowledge Discovery from Complex Data, Springer, 2005, Part II, pp.231-252.
[12]. Carlo Batini, Monica Scannapiece. (2006). Data Quality (Concepts, Methodologies and Techniques). First Edition, Springer, 2006.
[13]. Vicenc Torra. (2003). Trends in Information fusion in Data Mining. Information Fusion in Data Mining, Springer, 2003, pp. 1-6.
[14]. Mohammed J. Zaki. (2005). TreeMiner: An Efficient Algorithm for Mining Embedded Ordered Frequent Trees. Advanced Methods for Knowledge Discovery from Complex Data, Springer, 2005, Part I, pp.123-152.
[15]. J. Tang. A Framework on Developing an Intelligent Discriminating System of Anti Money Laundering, International Conference on Financial and Banking, Czech Rep., 2005.