Analytics for Detection of Money Laundering

Technical Report · April 2014

3 authors:

Girish Palshikar
Tata Consultancy Services Limited
136 PUBLICATIONS **724** CITATIONS
SEE PROFILE

Manoj M. Apte
Tata Consultancy Services Limited
17 PUBLICATIONS **123** CITATIONS
SEE PROFILE

Sriram Baskaran
University of Southern California
4 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Analytics for Detection of Money Laundering

Girish Keshav Palshikar
Tata Research Development and Design Centre (TRDDC)
Tata Consultancy Services Ltd.
54B Hadapsar Industrial Estate,
Pune -411013
91-20-66086400
gk.palshikar@tcs.com

Manoj Apte
Tata Research Development and Design Centre (TRDDC)
Tata Consultancy Services Ltd.
54B Hadapsar Industrial Estate,
Pune -411013
91-20-66086414
manoj.apte@tcs.com

Sriram Baskaran
Tata Research Development and Design Centre (TRDDC)
Tata Consultancy Services Ltd.
54B Hadapsar Industrial Estate,
Pune -411013
91-20-66086303
sriram.baskaran@tcs.com

ABSTRACT

Money Laundering (ML) is a serious problem for the economies and financial institutions around the world. Financial institutions get used by organized criminals and terrorists as vehicles of large-scale money laundering, which presents the institutions with challenges of regulatory compliance, maintaining financial security, preserving goodwill and reputation and avoiding operational risks like liquidity crunch and lawsuits. Hence prevention, detection and control of ML are crucial for the financial security and risk management of financial institutions. In this paper, we begin with an overview of the problem of ML, discuss some commonly used methods of ML, study a real life example where one of the premier banks was fined for insufficient AML controls, look at challenges faced by banks and financial institutions, discuss the analytic techniques which can be used for detection of ML and Money Laundering Detection Program. Further this paper reports the results of experimentation done on two ML patterns to determine the effectiveness of the RRS and angle based outlier detection algorithms and show that the algorithms are reasonably accurate in detecting ML accounts.

1. INTRODUCTION

Money Laundering (ML) is a serious problem for the economies and financial institutions around the world. Financial institutions get used by organized criminals and terrorists as vehicles of large-scale money laundering, which presents them with challenges such as regulatory compliance, maintaining financial security, preserving goodwill and reputation and avoiding operational risks like liquidity crunch and lawsuits. With its connections to organized crimes as well as terrorist financing, ML has become a serious issue worldwide and has been receiving considerable attention from national governments and international bodies such as the United Nations (UN), International Monetary Fund and the World Bank [1-4].

TACTiCS – TCS Technical Architects' Conference 2014

ML refers to activities performed with the aim of enabling the use of illegally obtained ("dirty") money for legal purposes, while hiding the true source of the money from government authorities (Figure 1).

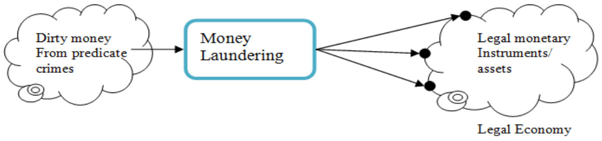

Figure 1. Money Laundering

Dirty money often comes from predicate (underlying) crimes such as drug trafficking, illicit arms trades, smuggling, prostitution, gambling, corruption and bribery, fraud, piracy, robbery, currency counterfeiting and other organized crimes. In some cases, it may also come from incomes of legal businesses which need to be hidden for evading taxes. ML enables the conversion of cash from the underground (shadow) economy into monetary instruments of the legal economy.

Key issues in anti-money laundering (AML) are: identifying the modus operandi, identifying monetary instruments and institutions involved, identifying parties and beneficiaries involved, proving the unlawful origins of the money, tracing the transactions, proving liabilities and intent, proving violations of particular laws, preventive actions (audits, employee training) and investigative measures (e.g., suspicious transactions report) etc.

Many nations, including India, have AML laws and even the UN has issued guidelines for AML. Compliance to such AML rules and regulations is mandatory for major banks. However the main focus for such laws till now has been on compliance to Know Your Customer (KYC) norms and then tracking the suspicious entities based on linkage with the suspicious entities. Detecting ML instances based on analyzing the transactions moving through banks and financial institutions has not yet reached a level so that AML tools can use the same effectively. More details on the efforts of estimating the extent of ML and applications of Data Mining in ML detection can be found in [5].

2. TYPOLOGIES

Understanding the specific patterns used for ML activities is very crucial for AML regime as it will help devise strategies to identify the occurrence of such patterns. Some of the patterns that have been identified as suspicious to Money Laundering [1-2] are as follows:

- Transactions are conducted in bursts of activities in a short period of time, especially in previously dormant accounts.

- Extremely regular ATM withdrawals particularly from an unlikely account.

- Prior conversion of cash to checks through formal or informal check-cashing services

- Simple Banking Operation where cash deposits are broken down to sums below the reporting threshold and deposited in multiple banks under different aliases and then use cashier checks to consolidate in a different account.

- Securing an exemption from reporting

- Company, engaged in international trade in goods and services (which is, by definition, a wholesale operation) deposits large sums of cash in its domestic accounts.

- There can be multiple bank transfers, again from country to country, where each transfer is protected by secrecy laws that must be breached one at a time.

- Withdrawing cash from a bank in one place, re-depositing it in a bank somewhere else and then wiring it to yet a third location.

- Money is transferred from the criminal’s offshore bank account to a casino in some tourist centre abroad. The casino pays the money in chips; the chips are then cashed in; and the money is repatriated via bank check, money-order or wire transfer to the criminal’s domestic bank account where it can be explained as the result of good luck during a gambling junket

- Shell corporations are created which have outward appearance of legitimate businesses and bank accounts opened which accept deposits in cash/transfers outside the jurisdiction and then consolidate the money in a company’s account for different services rendered which might not have been delivered.

- Over-invoicing method where goods and services are charged at a higher rate (multiple times) and proceeds are transferred legitimately to a different company.

3. REAL LIFE EXAMPLE

AML efforts have met increasing success as the cooperation amongst international monitoring institutes has increased. With this evident cooperation it becomes easier to detect, track and identify potential money laundering occurring in banking and other financial sectors. There are many real life examples of money laundering which were brought to light due to the continuous efforts of the monitoring institutes.

In 2010, HSBC (Hong Kong Shanghai Banking Corporation) was cited by its federal regulator, the Office of the Comptroller of the Currency (OCC) [6], for multiple severe AML deficiencies, including a failure to monitor $60 trillion in wire transfer and account activity; a backlog of 17,000 un-reviewed account alerts regarding potentially suspicious activity; and a failure to conduct AML due diligence before opening accounts for HSBC affiliates. Subcommittee investigators found that the OCC had failed to take a single enforcement action against the bank, formal or informal, over the previous six years, despite ample evidence of AML problems.

The Subcommittee investigation focused on five areas of abuse:

1) Servicing High Risk Affiliates: HSBC’s U.S. Bank offered correspondent banking services to HSBC Bank Mexico, and treated it as a low risk client, despite its location in a country facing money laundering and drug trafficking challenges, high risk clients like casas de cambio, high risk products like U.S. dollar accounts in the Cayman Islands, a secrecy jurisdiction, and weak AML controls. The Mexican affiliate transported $7 billion in physical U.S. dollars to HSBC's US Bank from 2007 to 2008, outstripping other Mexican banks, even one twice its size, raising red flags that the volume of dollars included proceeds from illegal drug sales in the United States.

2) Circumventing OFAC Safeguards: Foreign HSBC banks actively circumvented U.S. safeguards at US Bank, designed to block transactions involving terrorists, drug lords, and rogue establishments. In one case examined by the Subcommittee, two HSBC affiliates sent nearly 25,000 transactions involving $19.4 billion through their HSBC's US Bank accounts over seven years without disclosing the transactions’ links to Iran.

3) Disregarding Terrorist Financing Links: HSBC's US Bank provided U.S. dollars and banking services to some banks in Saudi Arabia and Bangladesh despite links to terrorist financing.

4) Clearing Suspicious Bulk Travelers Checks: In less than four years, HSBC cleared $290 million in obviously suspicious U.S. traveler’s cheques for a Japanese bank, benefiting Russians who claimed to be in the used car business.

5) Offering Bearer Share Accounts: HSBC offered more than 2,000 accounts to bearer share corporations, despite the high risk of money laundering and illicit conduct that results since their ownership can be readily transferred without a trail.

HSBC has agreed to pay a fine of $1.92 Billion to settle the multi-year probe by the US prosecutors, for failing to impose strict Anti-Money Laundering (AML) rules to prevent laundering of cash from criminal activities.

4. CHALLENGES FOR EFFECTIVE ML CONTROL PROGRAM

It is mandatory for Banks to comply with AML rules. However Banks also have some challenges in that respect. Some of the challenges faced by the banks for effective compliance are as follows:

- Finding adequate resources for such activity keeping in mind the high costs for compliance and enforcement of AML initiatives. This makes the AML activities a cost centre and increases the risk of noncompliance.

- Lack of commitment to AML risk management among senior management and key AML staff.

- Making the information available to identify the potential links between seemingly unrelated entities and events is difficult.
- ‘Just enough’ and ‘Just in time’ approaches to meet the regulatory demands and to avoid penalties
- Loose internal controls on KYC aspects
- Isolation of operations to back office i.e. decentralization
- Involvement of Bank of Financial institution staff

Since the challenges include availability of manpower, commitment, some process related aspects and possible involvement of staff, alternative approaches involving analytics would help the banks in overcoming some of these challenges.

5. ANALYTIC TECHNIQUES FOR DETECTION OF MONEY LAUNDERING

A number of data mining and statistical techniques have been used for detection of ML instances. The input data is usually either the various suspicious reports (CTR, SAR etc.) or the dataset of all transactions within a financial institution. The output is the set of highly suspicious transactions or highly suspicious entities (e.g., persons, organizations or accounts). Supervised classification techniques (such as support vector machines) are not that suitable because of general unavailability of reliably proven ML instances as labeled training data as well as severe class imbalance, since the number of known ML instances are likely to be far fewer than normal transactions. Unsupervised techniques such as clustering, anomaly detection, community analysis and red flags (based on expert knowledge) have been used for ML detection.

5.1 Anomaly Detection

Anomalies are patterns in the data that do not confirm to a well-defined notion of normal behavior. A distance measure is assumed to be available which represents the proximity of the data point with respect to other data points in numeric terms. Anomalies are also known as outliers and represent data points which are away from host of other data points representing unexpected behavior of the data. Since anomalies are few in number and different from the normal data they generally are considered to be actionable information in wide variety of application domains.

5.2 Clustering / Grouping

Clustering/Grouping divides the records into groups of clusters such that the records in a cluster are similar to each other and records belonging to different clusters are dissimilar. Each group represents a certain class or a concept in a domain. A distance measure is assumed to be available which represents the similarity/dissimilarity in numeric terms. The exercise of clustering or grouping does not need any training data to be provided. Moreover the number, size and shape of clusters are unknown. In general different clustering algorithms discover different clusters even for the same database since the underlying distance measure could be different. There is no unique grouping of records in clusters and this makes it an ideal unsupervised technique for experimentation.

Many times Anomaly detection and Clustering are used with one another.

5.3 Community Detection

Community detection is grouping the different entities in a graph based network. However in a graph based environment it is difficult to arrive at a distance measure since the network data is discrete. So a different set of algorithms which use the graph properties like cliques, vertex/edge-betweenness are used. Entities can coexist as part of multiple communities whereas the communities by themselves might not be related. This results in linking different communities in sparse way. This technique is mainly used in analyzing patterns in social media and in fraud detection where all the links of suspicious entities are carefully investigated.

5.4 Red Flag

Red flag technique is a rule based method to find cases that are out of the ordinary. The rules define the conditions for normal and abnormal behavior of the entities and any special behavior of one or more entities will be flagged for further investigation. Each entity will be evaluated for statistical deviance and known abnormal patterns. The evaluation is done by scanning the data set and comparing each entity with the rest of the group and the behavior of the entity across the time scale along with the patterns of any special behavior. This method requires domain knowledge to define the rules of normal and abnormal behavior. Example of red flag technique would be the alarm systems in heat and thermal plants if the temperature goes above a certain level. In the field of fraud detection, accounts/person having contrasting transaction pattern from others or over the time scale can trigger a red flag which can be put under investigation for further knowledge on the behavior.

5.5 Profiling

A set of summary features (profile) is computed for each entity (based on domain knowledge). These profile features are usually nonlinear functions of the data under study and are designed to be highly representative of the behavior of the entities. (e.g., based on withdraw/deposit frequencies, transaction amount deviations, transaction volumes and velocities etc.). A suspiciousness index for entities is found out by using the analytic techniques based on which further investigations can be carried out.

A common approach for ML detection as used within a financial institute (e.g., a bank) is to first the segment the entities (e.g., accounts) into clusters and anomalies, using a suitable similarity measure and business knowledge. Finally, the entities are prioritized on the basis of their profile features and top-*k* (few) are selected for in-depth investigations.

6. EXISTING AML PRODUCTS

There are various AML Products which provide the mechanism to facilitate the mandatory AML regulations. We have explained the features of some widely used tools below.

6.1 SAS Money Laundering Detection

SAS Anti Money Laundering tool [7] has been one of the popularly used Money Laundering detection tools at present. It includes banking-specific specific data model that maps transaction records to the customer and includes core scheme for preparing data for nightly batch analysis. The data that SAS supports can be of any type, which can be nonmonetary event data, geographic data, risk lists, third-party data, associate data and a variety of customer information data etc. It provides high performance analytics and visualization. It also monitors and reports suspicious activities and generates alerts based on the risk analysis. The risk analysis is carried out by assigning risk scores to the customer, account or transaction based on the factors like, number of transaction, the origin country of the transaction, previous suspicious activity, velocity of transaction etc. It integrates the Customer details and generates risk scores for the same. It allows the user to specify periodic reassessment of the risk classification. It combines functionality with the Dow Jones Watch list service to provide leading compliance risk information in a format designed for automated screening and risk management. Once the suspicious accounts, customers or transactions are flagged, it is the work of the investigation agency to carry out further investigations on the same. This includes an investigation management which provides a web-based interface that supports the management, investigation and reporting needs of analysts and investigators.

6.2 Oracle Anti Money Laundering System

Oracle Financial Services [8] has got a comprehensive financial crime and Compliance Management System that enables financial services customers to deploy multiple compliance applications on a common platform one of which is the Anti Money Laundering tool. The data for this can be of any format that covers all aspects of trading and account activity. The correlation information that is used to identify illegal activities can be user defined based on the information the user wants to track. This provides a configurable and flexible definition layer where institutions can define their own correlation rules and criteria. Once the correlation and rules are defined, the scenario evaluates the flow of funds through an account, looking at relative incoming to outgoing and comparing to the retained net worth of the account. The parameters can be set separately based on whether the account is new or seasoned and based on the risk levels associated with the account and with the other parties on the wire. A single threshold set evaluates each threshold considering these dimensions in a single scenario run. The tools Oracle has developed allow firms to find potential violations that would go undetected in the simpler compliance systems developed and used by most firms. This tool enables the user to generate timely reports after flagging the account, customer or transaction for further investigation.

6.3 TCS BaNCS AML Program

TCS BaNCS AML program [9] aims at a simple single view of enterprise risk exposure which provides a holistic view of critical lines of business at an enterprise level to access the type, risk associated with the customer and the severity of the alert in comparison with each of the Lines of Businesses (LOBs) with in the firm. This system manages to centralize the risk management view which enables the user to have a consolidated view across various LOBs for effective KYC and transaction surveillance. The single point of access provides an easy and simple way to access the KYC details of the customer and assign risk scores for the based on the customer transaction patters across various LOBs.

7. MONEY LAUNDERING DETECTION PROGRAM

Key issues in *Anti Money Laundering* (*AML*) are: identifying the modus operandi, identifying monetary instruments and institutions involved, identifying parties and beneficiaries involved, proving the unlawful origins of the money, tracing the transactions, proving liabilities and intent, proving violations of particular laws, preventive actions (audits, employee training) and investigative measures (e.g., suspicious transactions report) etc.

While there are mandatory AML regulations which are to be adhered to by the banks, the regulations themselves focus more on knowing the real identity end users (the Know Your Customer (KYC) business process) and most banks focus more on compliance to the KYC regulations. This creates a need for an independent analysis of the transactions of the various entities in a bank or financial institution and detect any suspicious activities. This is possible by using the process which is outlined in Figure 2.

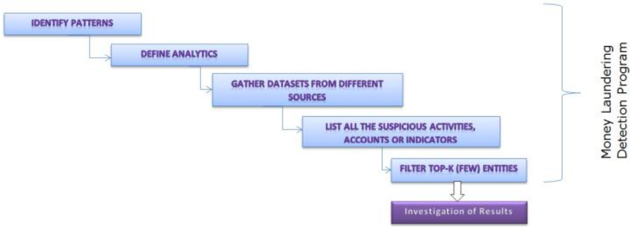
IDENTIFY PATTERNS
DEFINE ANALYTICS
GATHER DATASETS FROM DIFFERENT SOURCES
LIST ALL THE SUSPICIOUS ACTIVITIES, ACCOUNTS OR INDICATORS
FILTER TOP-K (FEW) ENTITIES
Investigation of Results
Money Laundering Detection Program

Figure 2. The Analytics Process for ML Detection.

We focus on the problem of identifying potential ML cases in the database consisting of all transactions, including cash and other types of transactions, in a bank. This data is necessarily incomplete (partial), because only some part of a particular ML activity may happen through this particular bank. The goal is to identify entities or transactions that are potentially involved in ML. Any further investigations need legal and investigative expertise and hence are necessarily manual, and so we do not focus on them here.

7.1 Research Component

Detection of incidences of ML is the basis of any AML activity. One major approach for this involves establishing "discordance" between the expected and observed business transaction profiles of the parties involved.

First step in this direction is to develop effective *unsupervised* data analytics algorithms for ML detection in all the transactions of a bank by applying specific techniques to specifically identified and formalized sub problems. We focus on this step. However, in real life, the next steps would integrate/fuse multiple databases including data from different banks, financial reports, cash flows, outside information, public information. In this paper, we investigate the use and adaptation of anomaly (or outlier, novelty) detection algorithms, for the purpose of detecting suspicious ML transactions or entities.

7.2 Banking Transactions Simulator

Since data is very crucial for any initiative related to Analytics, we have started building a Banking Transactions Simulator (BTS). The simulator generates data for "normal" banking transactions carried out by specific categories of Individuals/Businesses. The overview of the BTS is given in Figure 3.

The Banking Transactions Simulator has been created in three layers

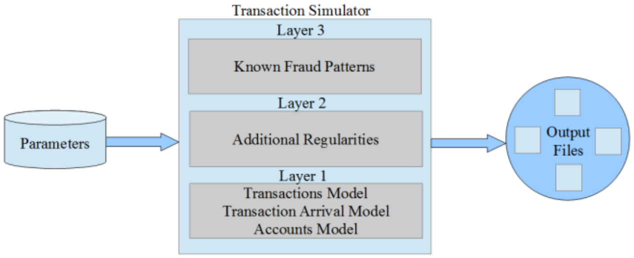

Figure 3. Banking Transaction Simulator

a. Layer1: Accounts Model, Transactions Arrival Model and Transactions Model

This layer first generates account details based on the configuration parameter values, decides the number of transactions of the account on a specific day, and then generate specific transactions with the details like date of the transaction, source account, destination account, amount and instrument. The simulator supports individual accounts of various type (e.g., salaried, student, retired) and also business accounts of various type (e.g., small, medium, and large businesses). The "normal" behavior for each type of account is defined through various parameters given above and BTS randomly generates transactions data for each account based on them.

b. Layer2: Additional Regularities

Every account has regularities other than the normal behavior. Some of the examples for the same are salaries, regular bills like telephone/electricity, loan repayments etc. Layer 2 generates such transactions based on the configuration parameters.

c. Layer 3: Known Fraud Patterns

Some known Fraud (ML) Patterns identified are built into this layer. The user is allowed to select and configure specific fraud patterns. The BTS simulator then marks specific accounts as ML related accounts, generates ML related transactions so that the same can be appended to the transactions along with additional regularities. The final transactions data includes generated data for both the "normal" accounts and speciallyconfigured ML accounts. This data is given as input to the ML (or anomaly) detection algorithms discussed later, and we examine the accuracy with which these algorithms detect the ML accounts. Of course, no information is provided to the algorithms about which accounts are involved in ML.

7.3 Money Laundering Detection Toolkit

We are building a Money Laundering Detection (MLD) Toolkit (Figure 4. Money Laundering Detection Toolkit) which would work on the available banking transactions data and detect the possibility of ML transactions and then generate a list of suspicious accounts and suspicious activity.

Ideally, the MLD Toolkit would work on data from various sources like the Banking Transaction Data from Banks, Available Financial Data like annual reports, fund flow statements from different entities and public domain data of specific entities. Connectivity with multiple databases and data warehouses would be provided for accepting the data. Facility of fusing the data will also be provided so as to unify the data from various sources and various formats. Analytics engine would use the various analytical algorithms identified above to report a list of Suspicious Activities, Accounts and Indicators which would indicate Risk of ML happening inside a bank.

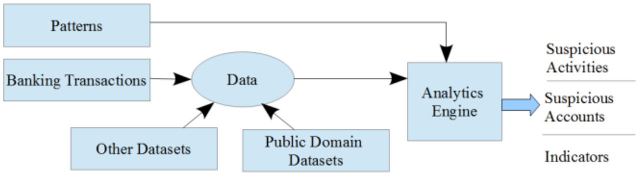

Figure 4. Money Laundering Detection Toolkit

8. EXPERIMENTS

As discussed earlier, we have implemented and adapted some well-known anomaly detection algorithms for the task of ML detection. In this paper, we report the results of our experiments on the accuracy of the two Outlier Detection algorithms, which are implemented in Java.

8.1 RRS Outlier Detection

Ramaswamy, Rastogi and Shim's (RRS) algorithm [10] defined a distance-based notion of an outlier as follows. Given an integer *k* (denoting number of nearest neighbors to be used), and another integer *N*, we find *N* possible candidate outlier records as follows. Let *D(P,k)* denote the distance of the k-th nearest neighbor of point *P*. For each point P in the given database, find *D(P, k)*. Sort the points in descending order of their *D(P, k)* values. The top *k* points in this sorted order are possible outliers.

8.2 Angle Based Outlier Detection

Using concepts like distance or nearest neighbor has some limitations. A novel approach using variance of angles between pairs of data points is proposed to counter the dimensionality problem in [11]. The intuition is that the variance of the angles subtended by a point with many other points is much smaller for outliers compared to other points in the data. Figure 5 Angle between different pairs of data points shows the representation of the concept.

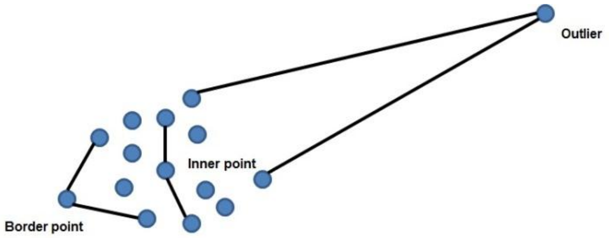

Figure 5 Angle between different pairs of data points

However the naive algorithm for computing the angle based outliers based on variance has complexity in $O(dn^3)$. A near linear time approximation algorithm FastVOA is proposed which uses the AMS sketching approach in [12].

8.3 Data Generation

We have used the Transactions model as given in Layer1 of the Banking Transactions Simulator. We have generated data for 10000 “normal” personal accounts.

The transaction arrival model for “normal” personal accounts uses the Poisson distribution with mean arrival rate represented by the parameter lambda (λ). Period under consideration for transactions generation is 1 year, i.e. 365 days.

Each “normal” Personal account can “interact” with a number of other accounts, which is randomly chosen and can be up to 30. The type of interaction is either “Cash Deposit” or “Cash Withdrawal”. The type of transaction i.e. Deposit/Withdrawal is chosen randomly with the probability of Withdrawal as 0.7.

Amount of each transaction of a “normal” Personal account is normally distributed with mean μ and standard deviation, σ.

In addition to the 10000 “normal” Personal” accounts, we generate data for 5 “dirty” accounts which are involved in ML. The transactions data for these “dirty” accounts is generated using two well-known Money Laundering Patterns.

8.3.1 Money Laundering Patterns

We have considered couple of Money Laundering Patterns which are explained below.

8.3.1.1 Regular Cash Withdrawals

One well-known way in which ML accounts behave is to have extremely regular (e.g., daily) cash withdrawals of amounts at or near to the legal limits. Hence in the Regular Cash Withdrawals ML pattern, we generate a series of very regular withdrawals per month which we have configured to be 20 for this experimentation. The withdrawal amount is considered to be normally distributed with mean 20000 and standard deviation 2000 keeping in mind the reporting threshold of 25000.

We generated 5 accounts with this ML pattern and used it along with the data generated in the previous step for “normal” Personal accounts.

8.3.1.2 Muling

A money mule or sometimes referred to as a "smurfer" is a person who transfers money acquired illegally (e.g., stolen) in person, through a courier service, or electronically, on behalf of others.

In this pattern typically accounts with low activity and low volumes are chosen to launder the money in spurts of transactions at chosen intervals.

In this ML pattern we generate 5 accounts with specifically low activity (where λ and μ for muling accounts is given separately based on specific values representing the “low activity”). We generate muling activity in a number of muling sessions/frequency (ψ). Each session would be conducted in a time window in the range of 2-5 days. The amount to be “muled” in each muling session is normally distributed with the mean of 50000 and standard deviation of 10000. The deposits in the accounts will be in cash with the amount chosen as under reporting threshold specified and the withdrawals will be into a list of accounts specified.

8.3.2 Summary Generation

Each record in the generated transactions has the following fields: ID, Timestamp, Source-Account, Destination-Account, Type-of-Transaction, Monetary-Instrument, Amount etc. Thus the transactions for a particular account forms a time-series. For the purposes of anomaly detection, we summarize this time-series and create a single record for each account that summarizes all its transactions. This summary record has several fields like

- Number of withdrawals by each account, mean and standard deviation for a period
- Total amount withdrawn by each account, mean and standard deviation for a period
- Number of deposits by each account, mean and standard deviation for a period
- Total amount deposited by each account, mean and standard deviation for a period
- Average Amount withdrawn per transaction
- Average Amount deposited per transaction
- Total number of transactions and standard deviation for a period
- Total amount involved in transactions and standard deviation for a period
- Maximum of “total amount involved in transactions per a given period”.
- Maximum of “total number of transactions per a given period”
- Total number of accounts that each account transacts with

8.3.3 Choice of Summary Variables

We have experimented with all the summary variables for each ML Pattern but found out that the accuracy of the algorithms degrades due to presence of noise introduced by irrelevant attributes (specifically for muling). So we chose a subset of the summary variables for experiments on both the ML patterns. For Regular Cash Withdrawals the subset of Summary Variables that we have considered includes No. of Withdrawals, No. of Deposits, Total Amount Withdrawn and Total Amount Deposited.

Similarly for Muling we have chosen Std. Deviation of Amount Deposited per period, Std. Deviation of Amount Withdrawn per period, Average Amount Withdrawn per period, Average Amount Withdrawn per transaction, Average Amount Deposited per transaction, Maximum Amount per transaction.

8.4 Experimental Results

For RRS we used series of 2nd Nearest, 3rd Nearest, 4th Nearest and 5th Nearest Neighbors for finding outliers. For FastVOA We used the parameter values, $s_1 = 1600$, $s_2 = 10$, $t = 100$ where t is number of random projections, s_1 and s_2 are the parameters which boost the accuracy of AMS sketching.. The selection for the parameter values was done based on the results reported for the FastVOA. For Regular Cash Withdrawals we varied λ, μ and σ and the results are shown below. For Muling we varied λ(Mule Accounts), μ(Mule Accounts) and ψ.

For each experiment we have captured two metrics. First is the number of ML accounts reported by the algorithm in the top 5 outliers (which is the number of ML accounts injected by us.). This can lead to the value of false positives/true negatives reported by the algorithm. Second metric is the maximum rank amongst the outlier accounts reported by the algorithm. This shows us the level where all the 5 ML accounts are reported by the algorithm.

However here we have reported only the first metric due to limited availability of space.

Experiments are carried out 10 times and mean values of number of ML accounts reported are shown as results.

8.4.1 *Regular Cash Withdrawals*

8.4.1.1 *Experiment 1*

We kept μ = 10000, σ = 1000 and changed λ from 0.2 to 0.4.

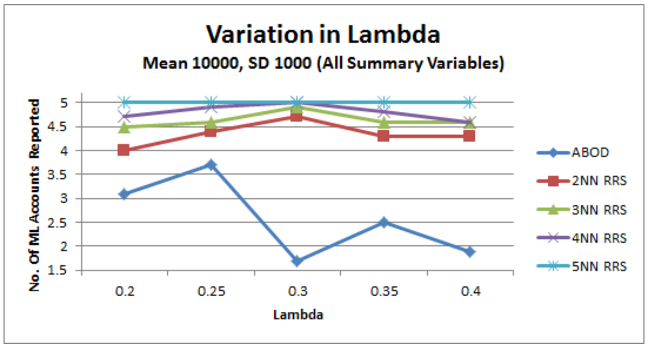

Figure 6 No. of ML Accounts Reported for Variation in Lambda for Regular Cash Withdrawals for All Summary Variables

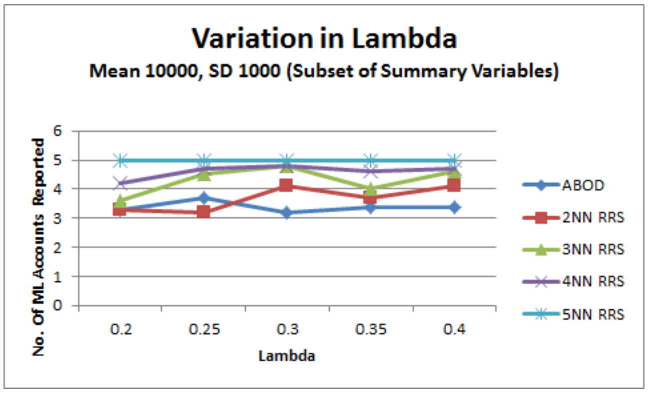

Figure 7 No. of ML Accounts Reported for Variation in Lambda for Regular Cash Withdrawals for Subset of Summary Variables

Despite the number of the transactions for normal accounts changing from 6 per month to 12 per month both the algorithms catch the ML accounts fairly accurately with RRS outperforming the FastVOA algorithm.

8.4.1.2 *Experiment 2*

We kept λ = *0.2,* σ = 1000 and changed μ from 10000 to 18000.

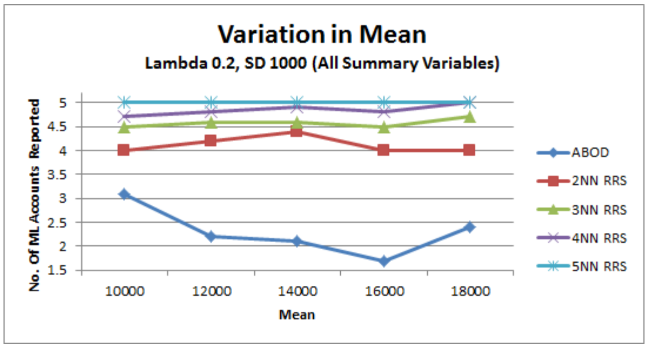

Figure 8 No. of ML Accounts Reported for Variation in Mean for Regular Cash Withdrawals for All Summary Variables

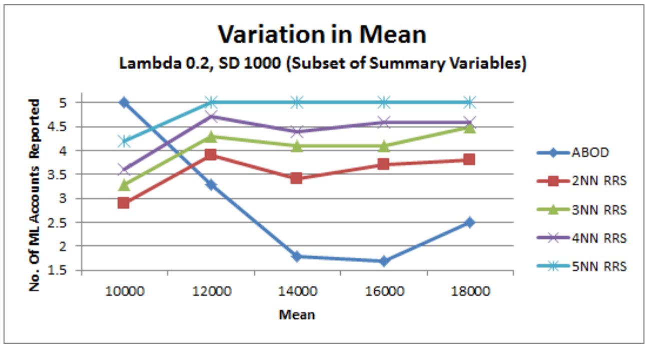

Figure 9 No. of ML Accounts Reported for Variation in Mean for Regular Cash Withdrawals for Subset of Summary Variables

For change in mean for normal accounts transaction amount varying from 10000 to 18000 (which is closer to the mean of RCW ML account transaction amount), the RRS algorithm catches the ML accounts effectively. However we are not able to explain reduction and increase in the performance of FastVOA.

8.4.1.3 *Experiment 3*

We kept λ = *0.2,* μ = 10000 and changed σ from 500 to 4500.

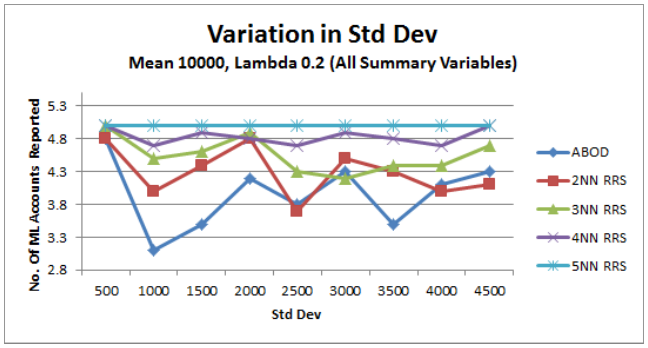

Figure 10 No. of ML Accounts Reported for Variation in Standard Deviation for Regular Cash Withdrawals for All Summary Variables

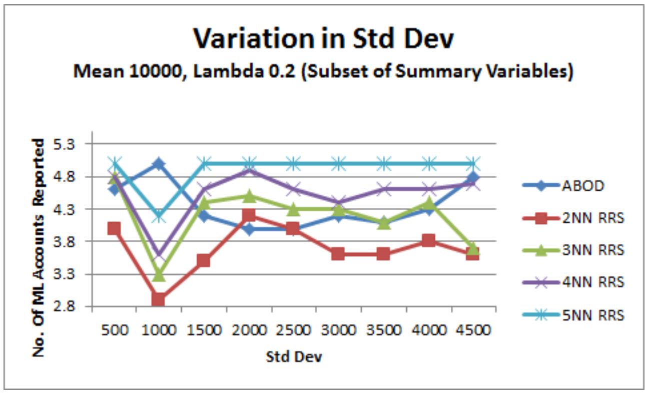

Figure 11 No. of ML Accounts Reported for Variation in Standard Deviation for Regular Cash Withdrawals for Subset of Summary Variables

For change in standard deviation for normal accounts transaction amount varying from 500 to 4500, both the algorithms are able to report the ML accounts as outliers effectively.

8.4.2 *Muling*

8.4.2.1 *Experiment 1*

We kept *λ (Muling)* = 0.05, ψ = 4 and changed *μ(Muling)* from 1000 to 2500.

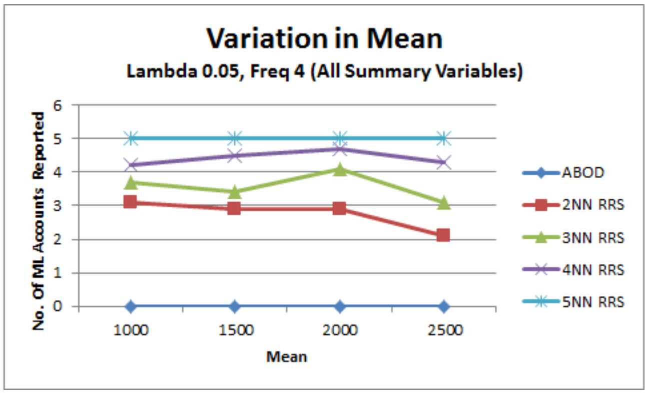

Figure 12 No. of ML Accounts Reported for Variation in Mean(Muling) for Muling for All Summary Variables

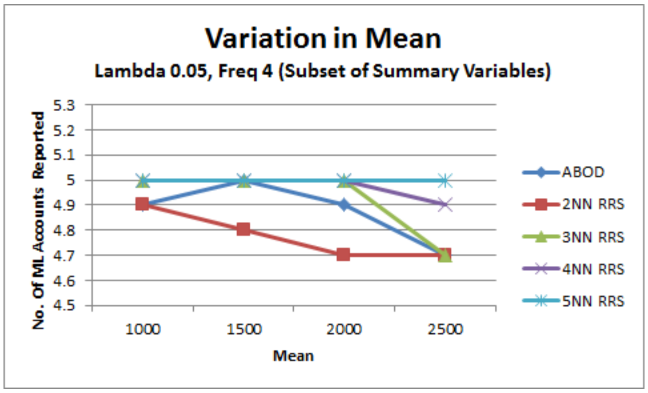

Figure 13 No. of ML Accounts Reported for Variation in Mean(Muling) for Muling for Subset of Summary Variables

For change in mean of normal transactions of Mule accounts from 1000 to 2500, we see that RRS captures the Mule accounts as outliers but FastVOA fails to do so when all Summary variables are used. However when the subset of attributes (chosen manually) is used, we see that FastVOA starts capturing the Mule accounts effectively.

8.4.2.2 *Experiment 2*

We kept *λ (Muling)* = 0.05, *μ(Muling)* = 1000 and changed ψ from 4 to 10.

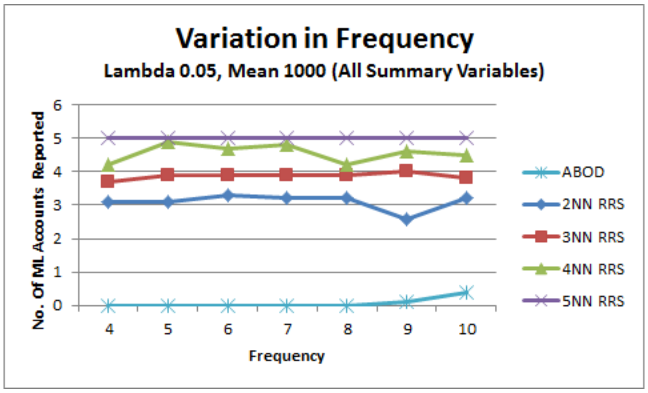

Figure 14 No. of ML Accounts Reported for Variation in Frequency for Muling for All Summary Variables

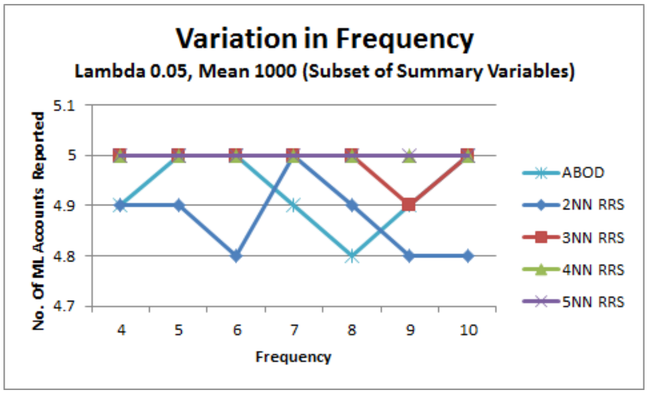

Figure 15 No. of ML Accounts Reported for Variation in Frequency for Muling for Subset of Summary Variables

Here when the Frequency of Muling is increased from 4 to 10 we see that RRS captures the Mule Accounts but FastVOA captures them only when the Subset of Summary variables are used.

8.4.2.3 *Experiment 3*

We kept ψ = 4, *μ(Muling)* = 1000 and changed *λ (Muling)* from 0.06 to 0.15.

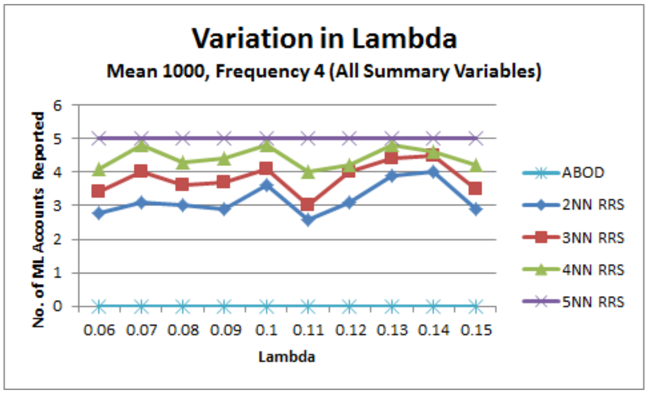

Figure 16 No. of ML Accounts Reported for Variation in Lambda (Muling) for Muling for All Summary Variables

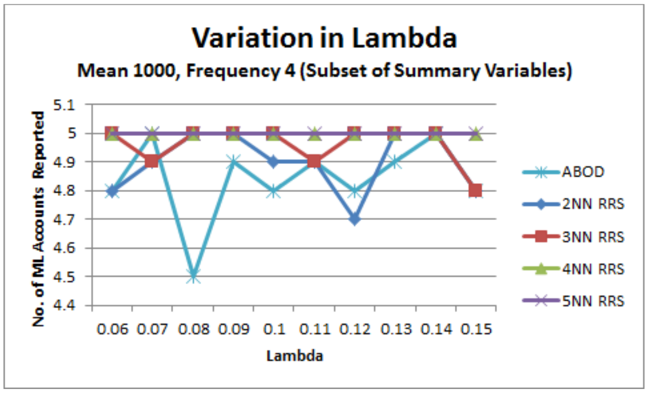

Figure 17 No. of ML Accounts Reported for Variation in Lambda (Muling) for Muling for Subset of Summary Variables

When we change the transaction arrival rate for Mule accounts from 0.06 to 0.15, RRS captures the Mule accounts correctly. However, like above experiments for Muling, Fast VOA captures the Mule accounts only when Subset of Summary Variables are used.

9. Future Work

We have completed the analysis for two ML patterns and would proceed with generation of data for more ML patterns which will include

- Shell Corporations: In this scheme accounts of some corporations which really do not exist are used (Shell Corporations). Numerous cash deposits are made to each of these accounts which are then transferred via Wire Transfer to some other account.
- Overseas Transactions : In this scheme a large number of transactions are made to and from overseas accounts

We would also experiment with more anomaly detection algorithms which include working on temporal data to find outliers based on peer group behavior [13], integrating community matching and outlier detection for mining evolutionary community outliers [14].

10. Conclusion

Financial institutions get used by organized criminals and terrorists as vehicles of large-scale money laundering, which presents these institutions with challenges of regulatory compliance, maintaining financial security, preserving goodwill and reputation and avoiding operational risks like liquidity crunch and lawsuits. Hence prevention, detection and control of ML are crucial for the financial security and risk management of financial institutions. In this paper, we began with an overview of the problem of ML, discussed some commonly used methods of ML, some existing AML Products and a real life case study of a premier banking institution which was fined for not implementing enough AML controls.

While there are mandatory AML regulations which are to be adhered to by the banks the regulations themselves focus more of knowing the end users (KYC) and banks face limitations in implementing the same. This creates a need for independent analysis from the dealings of entities in the bank or financial institution and detects suspicious activity.

The Money Laundering Detection Program which identifies the specific money laundering patterns, develop effective unsupervised learning algorithms on a combined dataset of banking transactions, other financial data like annual report, funds flow etc., Public information database. The results would represent a set of suspicious entities and incidents which need to be investigated in detail.

Experimentation was conducted on two of the anomaly detection algorithms RRS Outlier Detection and FastVOA. We observed that the RRS algorithm gives accurate results specifically for 5NN RRS. Other instances i.e., 4NN, 3NN and 2NN RRS give fairly good results. FastVOA performs fairly well in Regular Cash Withdrawals pattern and when a Subset of Summary variables is considered for Muling Pattern. However FastVOA fails to capture the Mule accounts as outliers when all the summary variables are used.

11. REFERENCES

[1] Madinger, J, *Money Laundering: A Guide for Criminal Investigators*, 3/e, CRC Press, 2012.

[2] Reuter, P. and Truman, E. M., *Chasing Dirty Money: Progress on Anti-Money Laundering*, Peterson Institute, 2004.

[3] Turner, J. E., *Money Laundering Prevention: Deterring, Detecting, and Resolving Financial Fraud*, Wiley, 2011.

[4] Woods B.F., *Art and Science of Money Laundering: Inside the Commerce of International Narcotics Trafficking*, Paladin Press Colorado, 1998.

[5] Palshikar, G. K. and Apte, M. *Financial Security against Money Laundering: A Survey*. In Akhgar, Babak and Arbania, Hamid (Eds.) *Emerging Trends in ICT Security*, Elsevier, 2013

[6] http://www.hsgac.senate.gov/subcommittees/investigations/media/hsbc-exposed-us-finacial-system-to-money-laundering-drug-terrorist-financing-risks

[7] Stewart, D., et al., *SAS® Money Laundering Detection*, SAS, 2009.

[8] Handa, G. et al., *Best Practices for Anti Money Laundering (AML): System Selection and Implementation*, Oracle Financial Services, 2010.

[9] Nisal, O. and Ram, V., *A case for Enterprise-wide Solutions for AML and Anti-Fraud*, Tata Consultancy Services, 2009.

[10] Ramaswamy S,et al., *Efficient algorithms for mining outliers from large datasets*, Proc.SIGMOD2000, pp. 162-172, ACM Press, 2000

[11] Kriegel, H.-P., Schubert, M., and Zimek, A., *Angle-based outlier detection in high-dimensional data*. Proceedings KDD'08, pages 444-452, 2008.

[12] Pham, N., Pagh, R., *A Near-linear Time Approximation Algorithm for Angle-based Outlier Detection in High-dimensional Data*, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[13] Bolton, R. J., and Hand, D. J., *Unsupervised Profiling Methods for Fraud Detection*, Credit Scoring and Credit Control VII (2001): 235-255.

[14] Gupta, M., Gao, J., Sun, Y., & Han, J. *Integrating community matching and outlier detection for mining evolutionary community outliers*, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 859-867). ACM, August 2012.