CHAPTER

Financial Security against Money Laundering: A Survey

36

Girish Keshav Palshikar and Manoj Apte
Tata Consultancy Services Limited, Pune, MH, India

INFORMATION IN THIS CHAPTER

- What is Money laundering?
- Anti-money laundering efforts
- Estimating the extent of ML
- Data mining techniques for ML detection

Money laundering

Money laundering (ML) is a serious problem for the economies and financial institutions around the world. As a recent example, a global bank paid a fine of $1.9 billion to the US government in a large ML case[1]. As another example, recently an online global currency exchange company was accused by the US government of laundering over $6 billion in seven years through 55 million transactions for millions of customers worldwide[2]. Financial institutions get used by organized criminals and terrorists as vehicles of large-scale money laundering, which presents them with challenges such as complying with regulations, 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), the International Monetary Fund (IMF) and the World Bank [1–4]. In this paper, we begin with an overview of the problem of ML and discuss some commonly used methods of ML and the anti-ML efforts worldwide. After surveying some analytics techniques used to estimate the extent of ML, we survey some data-mining techniques reported in the literature for detection of ML episodes (instances).

[1] http://www.reuters.com/article/2012/12/11/us-hsbc-probe-idUSBRE8BA05M20121211
[2] http://www.nytimes.com/2013/05/29/nyregion/liberty-reserve-operators-accused-of-money-laundering.html?pagewanted=all&_r=0

Emerging Trends in ICT Security. DOI: http://dx.doi.org/10.1016/B978-0-12-411474-6.00036-0

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. 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 that 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*.

Hiding the true source of the dirty money (activities, locations, people, and organizations) is a crucial requirement for ML. The destinations (i.e., the receivers) of the laundered money (i.e., the dirty money brought into the legal economy) are often known legal entities, such as registered businesses and legal citizens, though sometimes these are shell companies or nonexistent persons. The laundered money is often in the form of legal instruments (e.g., cash, bonds) or legal assets (stocks, real estate, jewelry, etc.) held by these legal entities. Sometimes, ML methods attempt to "mix" the dirty money into the legal income of these entities, with the aim that they should appear indistinguishable from each other. In this sense, hiding the true destination of the laundered money is also an important requirement for ML. This is also the case when ML is being done for terrorist financing.

Apart from the social costs arising out of the prevalence of organized crimes, ML also has adverse effects on the legal economy. ML undermines the integrity of the financial institutions: arrival of large sums of laundered money at financial institutions and their sudden disappearance can cause liquidity problems as well as loss of reputation and goodwill. ML may cause inexplicable changes in money demand, supply, increased volatility of international capital flows, and interest and exchange rates, which makes it difficult for countries to control their economic policies. Money launderers are more interested in protecting their proceeds and not necessarily in getting returns from the investment. Thus they invest in activities that are not necessarily economically beneficial; for example, they invest in construction or hotels not because of any demand but because of their short-term interests.

Criminals keep devising complex schemes (or methods) to perform ML, which involve multiple jurisdictions (or countries), dummy (or shell) companies, stolen or fake identities, (mis)use of financial institutions (banks, credit cards, stock market, and insurance), as well as auxiliary businesses such as real estate, shipping, and jewelry. Figure 36.1 shows a broad classification of the steps involved in a typical ML method. The authors of [5] survey emerging ML methods that use virtual reality role-playing games (e.g., Second Life, World of Warcraft) and online multi-player games for ML, because these environments offer opportunities that allow large sums of money to be moved across national borders without restriction and with little risk of detection or tracing.

With the advent of the Internet and online banking and investments, new financial services have made it much easier to conduct international transactions as well as to hide or steal true identities. Since the types and volumes of legal transactions are very large, and often the individual transactions involved in an ML episode appear superficially legal, it is not easy to detect entire ML episodes, particularly because criminals may be using innovative and unknown schemes. Hence automated data mining techniques are needed, which can be coupled with experts' domain knowledge about financial transactions and criminal investigations for effective ML detection.

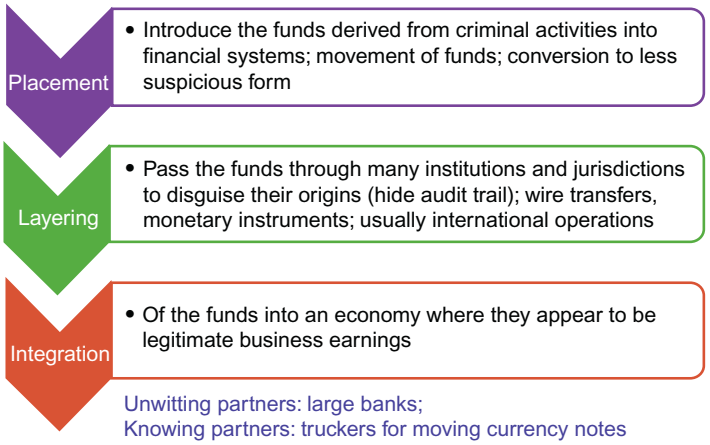

FIGURE 36.1

Broad steps in an ML method.

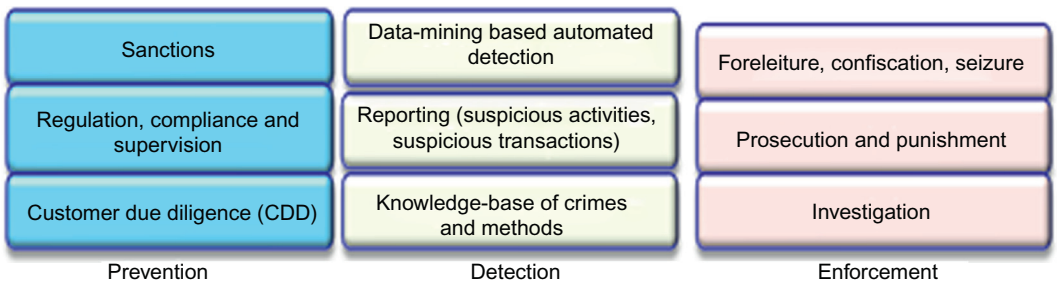

FIGURE 36.2

Pillars of AML efforts.

Anti-money laundering efforts

Realizing the gravity of ML, various nations have started a number of *anti-ML (AML)* activities, along with cooperative international efforts, for the prevention, detection, and control of ML. The main goals of an AML regime are: (i) reduce the illegal drugs trade and other organized crimes (blue/white collar), (ii) protect the integrity of the core financial system, and (iii) control corruption and terrorism-related activities. AML activities mainly consist of passing ML-related legislation, establishing financial crime investigation units, establishing compliance norms for financial institutions, etc. (Figure 36.2) [2]. The UN General Assembly established an action plan against ML in 1996, and in 1998 urged the member states to adopt national money laundering legislation and programs. Many other countries also have similar laws; for example, the Prevention of Money Laundering Act passed in 2003 in India.

Key issues in investigating an ML episode 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, and prosecution. AML also includes preventive actions (audits, employee training) and detection and regulatory compliance steps (e.g., reporting suspicious transactions), etc.

International cooperation in AML activities

International cooperation is crucial in controlling ML, primarily because most large-scale ML methods involve moving money across national borders and making use of international financial institutions. Advent of the Internet-related financial services (e.g., for payments) has allowed criminals to come up with new ML methods. Criminals often take advantage of weak regulations or weak enforcement authorities in some geographies and make extensive use of such "havens" to devise complex ML methods. Hence, international cooperation is crucial for ensuring a smooth exchange of information about ML methods as well as coordination of anti-ML actions; for example, data integration and sharing, knowledge sharing, AML regime compliance, surveillance and monitoring, investigation, prosecution, etc.

In addition to the UN, the World Bank and IMF also take interest in the control of ML and establishment of effective national AML regimes. The Egmont Group (http://www.egmontgroup.org) is an international consortium of the national Financial Investigation Units (FIUs) of over 100 countries, which cooperate in collecting knowledge and exchanging information about individuals, organizations, ML methods, etc. Some global banks have formed the Wolfsberg Group for establishing standards and policies for the financial services industry, with a focus on the *customer due diligence (CDD)*, ML, and terrorist financing [6] (www.wolfsberg-principles.com). The Financial Action Task Force (FATF) is an international consortium of 33 countries (including India) established in 1989 with the aim to "set standards and promote effective implementation of legal, regulatory and operational measures for combating money laundering, terrorist financing and other related threats to the integrity of the international financial system" (www.fatf-gafi.org). FATF periodically issues a set of comprehensive recommendations for coordinating and strengthening the AML regimes in its member countries (1996, 2001, 2003, 2012) [7–10]. The recommendations cover a gamut of AML activities: enforcement (confiscation), prevention and detection (CDD and record-keeping, reporting of suspicious transactions), investigation and compliance (regulation, supervision, dealing with non-compliance with AML standards, transparency), and international cooperation. Some example recommendations are:

- Do not allow anonymous accounts or accounts in obviously fictitious names.
- Identify the customer and verify the customer's identity using reliable, independent source documents, data, or information.
- Take reasonable measures to verify the identity of the beneficial owner.
- Obtain information on the purpose and intended nature of the business relationship.

FATF identifies some ML havens (offshore financial centers ([OFC]) as: Seychelles, Lichtenstein, Nauru, Myanmar, Bahamas, Bahrain, Cayman Island, Mauritius, Luxembourg, the Netherlands, Antilles, and Antigua.

Implementation of an AML regime

Most AML regimes require financial institutions to file various reports such as the Suspicious Activity Report (SAR), suspicious Currency Transaction Report (CTR), and reports on suspicious entities; in the United States, for example, a suspicious CTR involves cash of $10,000 or more, and an SAR involves participation of entities from known ML havens. It is usually the responsibility of banks, stock markets, and other financial institutions to detect, investigate, and prevent ML and to report suspicious cases to appropriate authorities, such as the FinCEN (Financial Criminal Enforcement Network) in United States and the Financial Investigation Unit in India (http://fiuindia.gov.in). Implementing an AML regime and ensuring compliance with the relevant legislations and international commitments are large and complex tasks. It is difficult to sift through the suspicious transaction reports and identify the true instances of ML. Some of the challenges include:

- A high volume of transactions ($> $1000 trillion p.a. worldwide; millions of wire transfer/Internet payments daily);
- Transactions around the world involving millions of people and companies, complex instruments and products (checks, credit cards, investments, etc.), transaction types (wire transfer, cash, etc.), and channels (ATM, Internet);
- Very low transaction volumes for actual ML ($< $1 trillion p.a.); most transactions are "normal";
- Unknown and evolving modus operandi; and
- Partial and incomplete data, knowledge, and investigation skills in any one institution (need for cooperation across multiple agencies).

As a rough estimate, financial institutions spent about $500 million in 2012 on AML products and services. Geiger and Wuensch [11] shed light on the costs and benefits of ML efforts and discuss the possible reasons for failure of AML to fight the predicated crimes and collateral damage caused by AML and call for a thorough review of the current approach followed by the world governing bodies for AML. Much practical advice is available in white-papers from several vendors for implementing an effective AML regime [12–15]. See [16] for a comparative evaluation of several AML products.

Estimating the extent of ML

Estimates of the annual total amounts of money laundered worldwide vary considerably. In one case, it is estimated to be several hundred billion dollars, or alternatively, as high as two to five percent of the global economy. Many techniques have been developed to estimate the extent of ML within a national economy [17–26], which may be defined as the amount of dirty money brought into the legal economy of a particular country; we survey some of them. It is difficult to directly estimate the actual amounts of money being laundered (say, in a year in a particular country). An indirect approach has often been used to infer an estimate of the extent of ML. One approach estimates the extent of the "shadow economy" and then, assuming that some known fraction of the shadow economy is converted into legal economy, we get an estimate of the extent of ML.

So we need to first define what is meant by a shadow economy. Several related definitions are in common use. Feige [18] defines shadow economy as all currently unregistered economic activities that contribute to the officially calculated (or observed) Gross National Product (GNP). Smith [23] defines it as market-based production of goods and services, whether legal or illegal, that escapes detection in the official estimates of GDP. Tanzi [25] discusses the incentives for a country to report inflated or deflated GNP numbers and the need and some limitations of the techniques used for estimating the size of the shadow economy. Frey and Schneider [19] and Schneider and Enste [21] debate on the causes of any increase of the shadow economy and effects of the same on the official economy. They also discuss different approaches for estimating the size of the shadow economy. Direct approaches use surveys and samples based on voluntary replies or tax auditing and other compliance methods to estimate the size of the shadow economy. Indirect approaches use discrepancies between observed and expected values of selected variables to estimate the extent of the shadow economy. Some indirect methods include:

a. The discrepancy between national expenditure and national income statistics
b. The discrepancy between official and actual labor force
c. The transactions approach
d. The currency demand approach
e. The physical input (electricity consumption) method
f. The modeling approach

We discuss some of these approaches here.

Discrepancy between national expenditure and income

As a principle of accounting, the income in the GNP of a country should equal the expenditure in the economy. So if an independent estimate of expenditure is available, then the gap between the expenditure and the officially reported income of the economy can be considered as constituting the shadow economy. This approach can yield good results if an accurate independent estimate of the expenditure is available. However, there is a risk of just finding out the discrepancies of the accounting statistics rather than actually getting the size of the shadow economy using this approach.

The currency demand approach

This is a commonly used econometric method to estimate the size of a shadow economy. This approach assumes that the shadow economy transactions happen in cash payments and so there is a direct relationship between the demand for currency and the size of the shadow economy. An excessive increase in the currency demand is attributed to a rising tax burden and the thriving of the shadow economy. This approach usually assumes that the size of the shadow economy is zero in the base year. Moreover, not all transactions in the shadow economy are actually carried out in cash and may involve barter, gold, and real estate holdings as alternate forms of currency. Another limitation is that this approach attributes all the increased demand for currency to the emergence of the shadow economy, which may not be correct.

Tanzi [24] has used the currency demand approach for estimating the size of the shadow economy. He modeled the currency demand by a regression equation, as follows:

$$\ln\left(\frac{C}{M_2}\right) = \beta_0 + \beta_1 \ln(1 + TW) + \beta_2 \ln\left(\frac{WS}{NI}\right) + \beta_3 \ln\left(\frac{NI}{N}\right) + \varepsilon$$

Here, M_2 is one possible indicator for the money supply. Methods to compute the quantity M_2 vary from country to country and depend on the nature of monetary instruments available. Roughly, M_2 includes the physical money in circulation + demand deposits + other checkable deposits + savings deposits + time deposits + traveler's cheques. It does not include, for example, cash in bank vaults, cash reserves of banks, large time deposits, or money market funds (e.g., investments in stocks, mutual funds, etc.). *TW* is a weighted average tax rate, *NI* is the national income, *WS*/*NI* is the proportion of wages and salaries in the national income, *NI*/*N* is the per capita income, *R* is the interest paid on savings and time deposits (to capture the opportunity cost of holding cash), C/M_2 is the ratio of cash holdings (currency in circulation) to money in current and deposit accounts, and ε is an error term.

The factors on the right are all known to affect the C/M_2 ratio. For example, the sign of β_1 is expected to be positive because, as the tax rates increase, tax evaders shift to tax-evading activities that require currency (rather than traceable monetary instruments such as checks). The sign of β_4 is expected to be negative, because as the economic development increases (as proxied by an increase in per capita income), more activities would use checks over cash, leading to a fall in demand for cash. Tanzi used US economy data from 1929 to 1980 to estimate the regression coefficients and got $\beta_0 = -5.0262$, $\beta_1 = 0.2479$, $\beta_2 = 1.7303$, $\beta_3 = -0.1554$, $\beta_4 = -0.2026$.

These can now be used to estimate the predicted value of the ratio C/M_2, and since M_2 is already known, the predicted value C_1 for *C* is known. Then Tanzi re-estimated the regression coefficients assuming zero tax rates (i.e., assuming $TW = 0$), and used them to get another estimate, C_2, for *C*. The difference between C_1 and C_2 is an estimate of the "illegal money," *IM*. The difference between the money supply indicator M_1 (obtained by excluding savings deposits and time deposits from M_2) and *IM* gives an estimate of the "legal money," *L*. Dividing the GNP by *L* gives an estimate of the income velocity *V* of the legal money. Assuming that the velocity of the illegal money is the same as *V*, an estimate of the size of the underground economy is obtained as $IM \cdot V$. Tanzi got the value of $21.75 billion (in terms of 1972 dollars) as the estimate for the size of the underground economy for the United States in 1980. See [17] for more modern versions of this approach; see [22] for more modern estimates of the shadow economy for various countries.

The electricity consumption approach

Several approaches have been devised to use resource consumption patterns to estimate the extent of the shadow economy. We review here the approach used by Lacko [20]. Lacko assumes that a part of the shadow economy is associated with the non-industrial consumption of electricity and that a part of the household electricity consumption is actually used by the shadow economy. In her model, the electricity consumption is affected by factors such as the GDP, the share of the industrial sector in GDP, the unemployment rate, and the fraction of electricity consumed by the industry. In fact, Lacko uses the year 1989 as the base year and the value of an independent

variable (as on 1994) is really the difference between its values in 1994 and 1989. Lacko's equation is

$$\Delta E = d_1 \Delta G + d_2 \Delta GI + d_3 \Delta EI + d_4 \Delta U + d_5$$

where ΔE is the change in electricity consumption between 1989 and 1994, ΔG is the change of the official GDP between 1989 and 1994, ΔGI is the change of the share of industry in the GDP between 1989 and 1994, ΔEI is the difference between the change of electricity consumption by the industry and the change in total electricity consumption between1989 and 1994, and ΔU is the maximal rate of unemployment between 1989 and 1994. The regression coefficients are estimated using the data for various post-socialist countries. Once an estimate of the electricity used by the shadow economy was obtained for a particular country, it was mapped to the fraction of that country's GDP using a simple baseline method. While this and similar approaches do have merit, they may not be reliable for power-deficient economies of developing countries. It is not clear if the equation is economically justified when using the data of the variables for a particular country over a number of years (rather than over multiple countries in the same year).

Modeling approach

A particular indirect method uses a specific indicator for the shadow economy and defines a model (usually, a regression model) that defines its dependence on various causes. Confirmatory factor analysis (CFA) techniques [27] offer a more general statistical framework, where one can define a model that explicitly identifies the dependencies among multiple indicator and cause variables (which are observed) and multiple factor variables (which are hidden). Structural equation modeling (SEM) is commonly used as the modeling formalism, which allows the modeler to define dependencies among multiple observed indicators, multiple observed causes, and multiple hidden factors (the MIMIC model). Dell'Anno [26] defines shadow economy as the single hidden factor η; six causes, namely, government employment in the labor force (X_1), tax burden (X_2), subsidies (X_3), social benefits paid by government (X_4), self-employment (X_5), and unemployment rate (X_6); and two indicators, namely, real Gross Domestic Product index (Y_1) and labor force participation rate (Y_2). The structural equations relating them are:

$$\begin{aligned} \eta &= \alpha + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + \gamma_6 X_6 + \zeta \\ Y1 &= \delta_1 + \lambda_1 \eta + \varepsilon_1 \\ Y2 &= \delta_2 + \lambda_2 \eta + \varepsilon_2 \end{aligned}$$

The last two equations relate the factors to the observed indicators and hence are called the measurement model. The model identification algorithm (usually based on maximum likelihood) estimates the coefficients and the variances and covariances of the independent variables using the sample data, and using them, estimates the population covariance matrix (which should be reasonably close to the observed sample covariance matrix, if the model is a good fit). Several test statistics, such as the χ^2-test, the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) are used to estimate the goodness of fit of the model to the given data. Once the values of η are estimated, Dell'Anno then computes the extent of the shadow economy as

a percentage of GDP by converting the index of the shadow economy estimated by the first structural equation (17.6 percent of the GDP in 1994 for Portugal).

Data mining techniques for ML detection

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 profiling, clustering, anomaly detection, link analysis, and data visualization have been used for ML detection. Knowledge representation techniques such as expert systems or Bayesian networks can be used to capture and use the domain knowledge of experts; for example, see [28], which uses agent-oriented ontology to capture anti-ML knowledge. Many good surveys are available that review the use of data mining techniques for general fraud detection (not necessarily ML) [29–32]. The authors of [33,34] survey AI/data-mining techniques that can be applied to ML detection.

A common approach for ML detection as used within a financial institute (e.g., a bank) is to first segment the entities (e.g., accounts) into clusters, using a suitable similarity measure and business knowledge. Then a suitable set of summary features (profile) is computed for each entity (based on domain knowledge) using their transaction histories. These profile features are usually nonlinear functions of the transaction data and are designed to be highly representative of the suspiciousness for entities (e.g., based on withdraw/deposit frequencies, transaction amount deviations, transaction volumes and velocities, etc.). Finally, the entities are prioritized on the basis of their profile features and top-k (few) are selected for in-depth investigations.

Senator et al. [35] from FINCEN have created the FINCEN AI system (FAIS) that links and evaluates reports of large cash transactions to identify potential ML and has been in operation at FINCEN since 1993. The objective is to detect previously unknown, potentially high-value entities (transactions, subjects, accounts) for possible investigation. The model supports three belief levels: Reported, Accepted, and Hypothesized. The reported transactions are at the belief level of Reported. These transactions are consolidated in clusters. Summary data like Subject and Account clusters, computed from the sets of reported transactions, represent the next belief level (Accepted). At these levels, certain derived attributes are computed that are necessary for evaluating the data-driven suspiciousness based on information discovered by analysts, including the linkages among the clusters. The highest level of belief (Hypothesized) is used for higher level abstractions like cases and patterns. FAIS has integrated the Alta Analytics NETMAP link-analysis package, which uses the "wagon-wheel" displays. FINCEN uses both wagon-wheel displays and traditional "link-and-edge" charts for analysis. FAIS has attempted use of techniques like Case-based Reasoning (CBR) and data mining (nearest neighbor, decision trees), which were

not very successful due to the lack of many labeled examples. Even unsupervised learning algorithms were found to be not so reliable because of difficulties in deriving appropriate features due to poor data quality and the need for background knowledge. These techniques were found to be useful as knowledge engineering aids. Analysts have used FAIS to generate the suspiciousness score and evaluated the subjects through research and analysis of the data available from all the sources for development of valid leads. These leads are then fully researched and analyzed by the law enforcement agencies. FINCEN uses the feedback from these agencies to make improvements in the system.

Given the fact that only a few entities will be known as having participated in ML, active learning can be used to reduce the need for labeled data. At each step, an active learning method selects one data point for manual labeling by the user and uses it to refine its classification model. Deng et al. [36] use active learning via sequential design for detecting money laundering. For simplicity, we assume that each account in the same cluster has two features, denoted $(x_1, x_2)^{\mathrm{T}}$. Define a convex combination $z = wx_1 + (1 - w)x_2$ to convert the data to the univariate form, where $\mathrm{w} \in [0, 1]$ is the unknown weight. The suspiciousness of an account is defined by the Logit function $F(z \mid \theta) = \mathbf{P}(Y = 1 \mid z, \theta) = \frac{e^{(z-\mu)/\sigma}}{1 + e^{(z-\mu)/\sigma}}$ which has three unknown parameters, $\theta = (\mu,\ \sigma,\ w)^{\mathrm{T}}$, and Y denotes the class label. At the end, given a threshold α (e.g., $\alpha = 0.8$), we identify all accounts z for which $F(z) \geq \alpha$, that is, the threshold hyperplane is $L_\alpha = \{x = (x_1, x_2)^{\mathrm{T}} : \frac{z-\mu}{\sigma} = \log\left(\frac{\alpha}{1-\alpha}\right),\ where\ z = \omega x_1 + (1 - \omega)x_2\}$. Given the current pool of labeled data used so far (starting with one labeled data point), the new value for θ is estimated using a maximum-likelihood technique. Then, since $F(z)$ is known only through a few noisy labeled data points, a stochastic approximation algorithm for finding roots of $F(z)$ [37] is used to find k_0 points closest to the current hyperplane. Among these candidates, the point with the maximum value for the Fisher information matrix is selected and presented to the user for manual labeling.

Observed financial transactions can be summarized as a graph with entities (e.g., accounts) as nodes. Graph mining techniques can be used to identify suspicious money flows across the edges of such a graph. Zhang et al. [38] propose a new Link Discovery based on Correlation Analysis (LDCA) on timeline data to identify communities in the absence of explicit link information. The correlation between two persons is defined through a correlation function between their financial transaction history vectors. If both are part of the same ML episode, they should exhibit similar financial transaction patterns, and thus, one expects a higher correlation value for them. Michalak and Korczak [39] propose a graph mining method for the detection of subgraphs corresponding to suspicious transaction patterns (e.g., a lattice-like sender-intermediaries-receiver pattern). Their method takes into consideration dependencies between individual transfers that may be indicative of illegal activities; see also [40].

Chang et al. [41] have presented a set of coordinated visualization metaphors including heatmap, search by example, keyword graphs, and strings and beads, which are based on identifying specific keywords within the wire transactions. This set of visualizations helps analysts to detect accounts and transactions that exhibit suspicious behaviors. Huang et al. [42] propose a two-step visualization-based solution for fraud detection in stock markets, where first they use 3D treemaps to monitor the real-time stock market performance and to identify a particular stock that produced an unusual trading pattern. Then they perform social network visualization to conduct behavior-driven visual analysis of the suspected pattern, identify the entities involved in the fraud, and further attack plans.

Zdanowicz [43] applies statistical outlier detection techniques to detect ML episodes in import and export trades data, since overvaluing imports or undervaluing exports is a common ML method. Kingdon [44] has developed a set of active agents ("Sentinels"), along with probabilistic methods, to detect unusual events and entities indicative of ML. Wand et al. [45] present two interesting unsupervised techniques to identify suspicious entities. In peer group analysis, an entity (e.g., account) is selected as a target and is compared with all other entities in the database, and a peer group of entities most similar to the target object is identified. The behavior of the peer group is then summarized at each subsequent time point, and the behavior of the target entity is compared with the summary of its peer group. Those target entities exhibiting behavior most different from their peer group summary behavior are flagged. Break point analysis (BPA) slides a window over the sequence of transactions of an account and uses statistical tests to compare a window with earlier ones to detect any sharp changes in the transaction patterns (e.g., frequency, amounts). Zengan [46] developed a cluster-based outlier detection algorithm to identify suspicious ML transaction patterns.

Ju and Zheng [47] have proposed a supervised decision tree algorithm for ML detection, combined with a privacy preserving strategy (Inner Product Protocol) to protect the identity of the account owners, in case they are not identified as part of suspicious ML. Gao and Ye [48] discuss a methodology for AML, in which many steps make use of various data-mining techniques such as outlier detection, link analysis, and community detection. Considering the huge transaction volumes, [49] discusses the use of data-warehousing and OLAP cube technology in AML applications. Phua et al. [50] developed a fraud detection method to predict criminal patterns from skewed data, which uses a single meta-classifier (stacking) to choose the best base classifiers (naïve Bayes, C4.5, and back-propagation), and combines their predictions (bagging) to improve cost savings (stacking-bagging). Given the scarcity of labeled data, generating and using synthetic transactions data that can contain known suspicious patterns is important for validating the ML detection algorithms; see [51].

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 is crucial for the financial security and risk management of financial institutions. In this chapter, we began with an overview of the problem of ML and discussed some commonly used methods of ML and the anti-ML efforts worldwide. After surveying some analytics techniques used to estimate the extent of ML, we surveyed some data-mining techniques that have been reported in the literature for detection of ML episodes (instances). Data-mining and statistical techniques play a crucial role in sifting through enormous volumes of financial transactions data to identify potentially suspect entities. While much has been done, there is a need to continue to develop more effective ML detection techniques, particularly in the face of class imbalance, lack of many labeled examples, evolution of newer ML methods (e.g., those involving identity theft or online games), availability of faster, newer, and more anonymous financial services over the Internet, and the need to capture and use

domain and investigative knowledge. Unlike most other applications, the evaluation of ML detection techniques is hampered by publicly available labeled data about known ML episodes. Nevertheless, a recent spate of high-profile ML detection cases, in the United States and in other countries such as India, is a clear pointer toward the effectiveness of AML regimes. Newer technologies such as big data analytics, text mining, social networks analysis, and anomaly detection should be explored for more effective ML detection.

References

[1] Madinger J. Money Laundering: A guide for criminal investigators. 3rd ed. CRC Press; 2012.
[2] Truman EM, Reuter P. Chasing dirty money: Progress on anti-money laundering. Peterson Institute; 2004.
[3] Turner JE. Money laundering prevention: Deterring, detecting, and resolving financial fraud. Wiley; 2011.
[4] Woods BF. Art and science of money laundering: Inside the commerce of international narcotics trafficking. Colorado: Paladin Press; 1998.
[5] Irwin ASM, Slay J. Detecting money laundering and terrorism financing activity in Second Life and World of Warcraft. Proceedings of the 1st International Cyber Resilience Conference. Edith Cowan University; Perth Western Australia; 2010.
[6] Hinterseer K. The Wolfsberg anti-money laundering principles. Journal of Money Laundering Control 2001;5(1):25–41.
[7] Financial Action Task Force (FATF) on money laundering, the forty recommendations; 2003.
[8] Financial Action Task Force (FATF) on money laundering, report on money laundering typologies 2002–2003; 2003.
[9] Financial Action Task Force (FATF) on money laundering, money laundering using new payment methods; 2010.
[10] Financial Action Task Force (FATF) on money laundering, special recommendations on terrorist financing; 2003.
[11] Geiger H, Wuensch O. The fight against money laundering. Journal of Money Laundering Control 2007;10:1.
[12] Deloitte Financial Advisory Services LLP. Anti-money laundering services: Helping clients implement anti-money laundering detection and compliance programs around the globe; 2008. www.deloitte.com.
[13] Menon R, Kumar S. Understanding the role of technology in anti money laundering-compliance: A strategic model for financial institutions; 2005. www.infosys.com.
[14] Prasanna G. Enterprise-wide anti-money laundering and KYC initiatives. Tata Consultancy Services; 2012. www.tcs.com.
[15] SAS Institute. Reducing the cost of AML compliance; 2011. www.sas.com.
[16] Ray A, Katkov N. Evaluating the enterprise-wide compliance vendors: Solutions for anti-money laundering and anti-fraud; 2012. www.celent.com.
[17] Bhattacharyya DK. On the economic rationale of estimating the hidden economy. Econ J. 1999; 109(456):348–59.
[18] Feige EL. The underground economies: Tax evasion and information distortion. Cambridge: Cambridge University Press; 1989.
[19] Frey BS, Schneider F. Bd 12 Economics Informal and Underground economy. International Encyclopedia of Social and Behavioral Science. Elsevier Science; 2000.

[20] Lacko M. Do power consumption data tell the story? Electricity Intensity and Hidden Economy in Post-Socialist Countries. Budapest Working Papers on the Labour Market 9902. Institute of Economics; Hungarian Academy of Sciences; 1999.
[21] Schneider F, Enste DH. Shadow economies: Size, causes and consequences. J Econ Lit 2000; XXXVIII:77–114.
[22] Schneider F, Buehn A, Montenegro C. New estimates for the shadow economies all over the world. International Economic Journal. Korean International Economic Association 2010;24(4):443–61.
[23] Smith P. Assessing the size of the underground economy: The statistics Canada perspectives. Canadian Economic Observer 1994;7:316–33.
[24] Tanzi V. The underground economy in the United States: Annual estimates, 1930–1980. IMF Staff Papers 1983;30(2):283–305.
[25] Tanzi V. Uses and abuses of estimates of the underground economy. The Economic Journal 1999;109: F338–47.
[26] Dell'Anno R. The shadow economy in Portugal: An analysis with the MIMIC approach. Journal of Applied Economics 2007;X(2):253–77.
[27] Thompson B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. American Psychological Association; 2004.
[28] Wang Y, Wang H, Gao S, Xu D, Ye K. Agent-oriented ontology for monitoring and detecting money laundering process. In InfoScale '07: Proceedings of the 2nd International Conference on Scalable Information Systems, Article 81. Institute for Computer Sciences. Social-Informatics and Telecommunications Engineering (ICST). Brussels, Belgium; 2007.
[29] Bolton R, Hand D. Statistical Fraud detection: A review. Statistical Science. 2002;17(3):235–55.
[30] Fawcett T, Foster T, Provost F. Adaptive fraud detection. Data Mining and Knowledge Discovery 1997;1:291–316.
[31] Phua C, Lee V, Smith K, Gayler R. A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review; 2005.
[32] Yue D, Wu X, Wang Y, Li Y, Chu C-H. A review of data mining-based financial fraud detection research. Int Conference on Wireless Communications. Networking and Mobile Computing; 2007. p. 5519–22.
[33] US Congress Office of Technology. Information technologies for the control of money laundering. OTA-ITC-630; 1995.
[34] Watkins RC, Reynolds KM, Demara R, Georgiopoulos M, Gonzalez A, Eaglin R. Tracking dirty proceeds: Exploring data mining technologies as tools to investigate money laundering. Police Practice and Research 2003;4(2):163–78.
[35] Senator TE, Goldberg HG, Wooton J, Cottini MA, Umar Khan AF, Klinger CD, et al. The financial crimes enforcement network AI system (FAIS): Identifying potential money laundering from reports of large cash transactions. AI Magazine. 1995 Winter;16(4):21–39.
[36] Deng X, Joseph VR, Sudjianto A, Jeff Wu CF. Active learning via sequential design with applications to detection of money laundering. J Am Stat Assoc 2009;104(487):969–81.
[37] Kushner HJ, Yin GG. Stochastic approximation and recursive algorithms and applications. Springer-Verlag; 2003.
[38] Zhang Z, Salerno JJ, Yu PS. Applying data mining in investigating money laundering crimes. SIGKDD'03 2003;747–72.
[39] Michalak K, Korczak J. Graph mining approach to suspicious transaction detection. Federated Conference on Computer Science and Information Systems 2011;69–75.
[40] Moll L. Anti money laundering under real world conditions-Finding relevant patterns [MS thesis]. Department of Informatics: University of Zurich; 2009.

[41] Chang R, Lee A, Ghoniem M, Kosara R, Ribarsky W, Yang J, et al. Scalable and interactive visual analysis of financial wire transactions for fraud detection. Information Visualization 2008;7(1):63–76.
[42] Huang ML, Liang J, Nguyen QV. A visualization approach for frauds detection in financial market. In: Proc 13th Int Conference on Information Visualisation; 2009. p. 197–202.
[43] Zdanowicz JS. Detecting money laundering and terrorist financing via data mining. Communications of the ACM 2004;47(5):53–5.
[44] Kingdon J. AI fights money laundering. IEEE Intell. Syst 2004;19(3):87–9.
[45] Bolton RJ, Hand DJ. Unsupervised profiling methods for fraud detection. Proc Credit Scoring and Credit Control VII 2007;5–7.
[46] Zengan G. Application of cluster based local outlier factor algorithm in anti money laundering. In MASS '09: Proc Int Conference on Management and Service Science; 2009. p. 1–4.
[47] Ju C, Zheng L. Research on suspicious financial transactions recognition based on privacy preserving of classification algorithm. In ETCS '09: 1st Int Workshop on Education Technology and Computer Science 2009;2:525–8.
[48] Gao Z, Ye M. A framework for data mining-based anti-money laundering research. Journal of Money Laundering Control 2007;10(2):170–9.
[49] Korczak J, Marchelski W, Oleszkiewicz B. A new technological approach to money laundering discovery using analytical SQL server. In AITM 2008: Advanced Information Technologies for Management. In: Korczak J, Dudycz H, Dyczkowski M, editors. Research Papers, 35. Wroclaw University of Economics; 2008. p. 80–104.
[50] Phua C, Alahakoon D, Lee V. Minority report in fraud detection classification of skewed data. SIGKDD Explorations 2004;6(1):50–9.
[51] Barse E, Kvarnström H, Jonsson E. Synthesizing, test data for fraud detection systems. In: Proceedings of the 19th Annual Computer Security Applications Conference; 2003. p. 384–95.