Technologies for Detecting Money Laundering

4

At the core of all wire transfer monitoring proposals are one or more computer technologies. Many of these technologies rely upon techniques developed in the field of artificial intelligence (AI). Others involve computer graphics and statistical computing. Wire transfer monitoring proposals generally involve a combination of technologies, institutional structures, and reporting requirements. Four of these combinations are presented as options in chapter 7. However, a limited set of technologies and their relative capabilities form the core of each option.

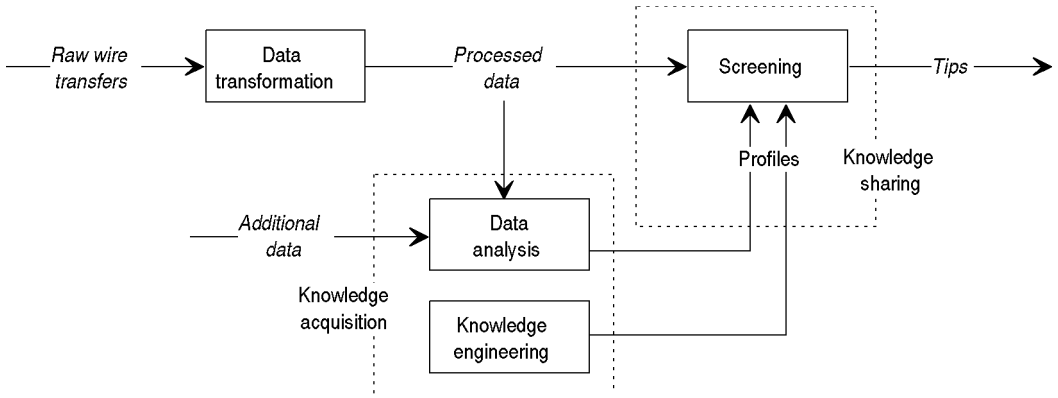
This chapter discusses two topics central to understanding these technical options and the policies surrounding their use. The first section introduces several basic technologies that are employed in one or more options. The second section discusses challenges that must be overcome by all options. These challenges involve characteristics of wire transfer data and money laundering profiles.

BASIC TECHNOLOGIES

There are at least four categories of technologies that may be useful in the analysis of wire transfers. These technologies can be classified by the task they are designed to accomplish:

- *wire transfer screening* to determine where to target further investigations,
- *knowledge acquisition* to construct new profiles for use during screening,
- *knowledge sharing* to disseminate profiles of money laundering activities quickly, reliably, and in a useful form, and

Figure 4-1: How Technologies Relate to Each Other



SOURCE: Office of Technology Assessment, 1995.

- *data transformation* to produce data that can be easily screened and analyzed.

Each category of technology is used in the technical options discussed in chapter 7. Screening is used in all options, knowledge acquisition in some, data transformation in most, and knowledge sharing in some of the options. Figure 4-1 shows the relative roles of these technologies in wire transfer analysis systems.

Wire Transfer Screening

Wire transfer screening is the heart of all options discussed in chapter 7. Technologies for screening wire transfers include knowledge-based systems and link analysis. *Knowledge-based systems* automatically make inferences about wire transfers and other data. Effective use of knowledge-based systems requires effective knowledge acquisition—a way of constructing profiles of money laundering. Effective knowledge acquisition, in turn, requires either human experts who know how to reliably screen wire transfers or a large sample of data that are "labeled" to indicate wire transfers of the sort that should be identified by a working system. *Link analysis* helps identify relationships among individual accounts, people, and organizations. Effective use of link analysis requires a variety of readily available data, some of which provide reliable indicators of money laundering activity.

Some technical options use a knowledge-based system exclusively. Others initially screen all wire transfers using a knowledge-based system and then allow analysts to scrutinize some or all transfers using link analysis. In the latter case, the knowledge-based system can be used to filter transfers—only passing on some transfers to the next stage of analysis—or the knowledge-based system can be used to derive additional data—passing on all transfers along with the additional derived data. The latter use is analogous to one part of the Financial Crimes Enforcement Network (FinCEN) Artificial Intelligence System (FAIS) (see box 4-1).

Banks already use a set of relatively simple systems to screen transactions for illicit conduct. Some of these systems screen currency transactions to identify those which indicate "structuring"—a series of transactions designed to evade current reporting requirements (e.g., five deposits of $3,000 each in a single day). Other systems monitor wire transfers to look for countries or individuals that appear on a list compiled by Treasury's Office of Foreign Assets Control (OFAC). While these systems are quite simple in comparison with the configurations discussed in chapter 7, they are examples of how such systems can be in-

Box 4-1: The FinCEN Artificial Intelligence System

The FinCEN Artificial Intelligence System (FAIS) is currently used to process and analyze all reports received under the Bank Secrecy Act (BSA).[1] Nearly all (more than 90 percent) of these reports are Currency Transaction Reports (CTRs). The Internal Revenue Service (IRS) Detroit Computer Center and the U.S. Customs Service Data Center collect and store BSA reports; FAIS adds value by linking and evaluating these reports.

FAIS uses three basic types of data. BSA reports—referred to as *transactions*—are used directly. Transactions that can be associated with the same person or business are used to create a new data element called a *subject*. Transactions that can be associated with the same bank account are used to create an element called an *account*. The grouping of transactions into subjects and accounts is accomplished by examining information in the transactions (e.g., name, address, social security number). If these items are sufficiently similar, then two transactions are assumed to belong to the same subject.

These three types of data elements—transactions, subjects, and account—are analyzed by another component of FAIS, a knowledge-based system.[2] FinCEN's knowledge-based system is derived from a system originally developed at the U.S. Customs Service for screening CTRs. The knowledge base from the Customs Artificial Intelligence System (CAIS) was re-engineered to function with FinCEN's system, and is continually updated to reflect changes in money laundering methods. The knowledge-based system component of FAIS is used to evaluate the suspiciousness of transactions, subjects, and accounts. Based on indicators that appear directly within transactions, and on additional indicators calculated from those transactions, FAIS assigns a numeric suspiciousness score to each transaction, subject, and account.

On the basis of these scores and several other criteria, FinCEN analysts select subjects and accounts for further investigation. This investigation is accomplished with the link analysis[2] component of FAIS. Link analysis is used to identify networks of financial activities that help to distinguish between legitimate business activities and money laundering.

FAIS uses a variety of commercial hardware and software. The system runs on a 6-processor SparcCenter 2000 server and several SparcStation workstations from Sun Microsystems, Inc. The database component uses an SQL server from Sybase, Inc.; the knowledge-based component uses Nexpert Object from Neuron Data, Inc.; the link analysis component uses NETMAP from ALTA Analytics, Inc. In addition to substantial software development done within these products, some additional parts of FAIS were developed using the language C and using Open Interface from Neuron Data, Inc.

FAIS has been operational since March 1993 and processes approximately 200,000 transactions per week. As of January 1995, 20 million transactions had been entered, linked, and evaluated, resulting in 3 million consolidated subjects and 2.5 million accounts. As of May 1995, the system had generated over 400 investigative support reports corresponding to over $1 billion in potentially laundered funds. FinCEN has received over one hundred feedback forms from outside agencies, as well as internal feedback. Over 90 percent of the feedback indicates either new cases opened or relevance to ongoing investigations.

---

[1] For additional description of FinCEN, see chapter 3.

[2] See main text for an explanation of knowledge-based systems and link analysis.

SOURCES: Ted Senator, Financial Crimes Enforcement Network, personal communications, March 1994 - June 1995. U.S. Treasury, Financial Crimes Enforcement Network, "FinCEN Artificial Intelligence System: Fact Sheet," no date. Ted Senator, Henry Goldberg, Jerry Wooton, Matthew Cottini, A.F. Umar Khan, Christina Klinger, Winston Llamas, Michael Marrone, and Raphael Wong, "The FinCEN Artificial Intelligence System: Identifying Potential Money Laundering from Reports of Large Cash Transactions," *Proceedings of the 7th Conference on Innovative Applications in Artificial Intelligence*, 1995 (forthcoming).

BOX 4-2: Current Monitoring and Compliance Systems

Some banks and wire transfer systems already have systems that examine currency and wire transactions. These systems are substantially less sophisticated than some proposed systems for wire transfer monitoring. However, they help indicate the state of current bank systems and the environment within which new systems would operate.

Currency Transaction Reporting

As noted in box 1-3, Currency Transaction Reports (CTRs) are filed when a customer deposits over $10,000 in cash. However, banks also look for evidence of *structuring*—a series of smaller cash transactions that are intended to evade reporting requirements. Even though these deposits are under the $10,000 threshold, they should be reported because they may indicate money laundering.

Banks and commercial software vendors have developed systems that aggregate multiple currency deposits over specific periods (usually days or weeks). For example, Chase Manhattan Bank, NA, a large money center bank, uses a system that aggregates multiple currency transactions that occur on the same business day. While the activity listed on the system's reports is very low, about 65 percent of Chase Manhattan's Criminal Referral Forms (CRFs) are a direct result of investigating account activity highlighted by the system. Similarly, Atchley Systems, Inc., a commercial software vendor, has developed a system that allows banks to aggregate currency transactions over a fixed specified number of days and report all aggregations that exceed a specific threshold. The system can also flag individual accounts and automatically generate reports on their cash activities each day. The latter component is used when bank managers wish to monitor the cash activity of certain accounts, even though it may not exceed specific thresholds.

Foreign Assets Control

Banks are required to comply with regulations issued by the Treasury Department's Office of Foreign Assets Control (OFAC). The regulations were promulgated under six statutes that prohibit, in various ways, trade with specific countries, including Cuba, North Korea, Libya, Iraq, Yugoslavia, UNITA (Angola), and Iran. In addition, Executive Order 12947 prohibits transactions with terrorists. To assist banks with compliance, OFAC maintains a list of specially designated nationals (SDNs) and blocked persons that contains over 2,500 entries. Each entry is an individual (e.g., Manuel Noriega) or organization (e.g., Hizballah). For individuals, addresses and titles are sometimes given; for organizations, a list of aliases and address information is generally given. In some cases, separate entries are made for alternative spellings or addresses of individuals or organizations. Each entry also contains a designation of what provision resulted in their inclusion in the list.

tegrated with bank operations and of the challenges posed by such integration (see box 4-2).

Knowledge-Based systems

Knowledge-based systems, often called "expert systems," are computer programs that process data in ways that emulate human experts. They differ from conventional algorithms in several ways. First, the knowledge that is embedded within the system is largely separate from the reasoning methods used to operate on that knowledge. Second, they often are designed so that they can display the path of evidence and facts used to reach a particular conclusion—in essence, knowledge-based systems can "explain" the inferences that they have made.

The knowledge embedded in knowledge-based systems often is expressed in terms of rules of the form shown in figure 4-2. Rules can directly connect input data to final conclusions; they can begin with intermediate conclusions of other rules; or they can produce intermediate conclusions that

BOX 4-2: Current Monitoring and Compliance Systems (Cont'd.)

Banks are required by OFAC regulations to block wire transfers going to organizations and individuals on the list. In the past several years, OFAC has imposed millions of dollars in civil penalties involving U.S. banks. Most of the fines were levied because a bank failed to block an illicit transfer that was processed manually (OFAC has not generally penalized banks for failing to block transfers that were processed automatically). Most large U.S. banks have computer systems in place to screen wire transfers. Several dozen banks and vendors have developed systems that automatically screen incoming wire transfers for locations, organization, and persons on the OFAC list. When one or more of these names is found, the transfer is stopped and brought to the attention of a human operator. The presence of such software is "considered favorably" when OFAC investigates a bank that failed to block an illegal transfer.

Fedwire Scanning System

The Federal Reserve Bank has the capability to electronically scan and retrieve records of wire transfers made over its Fedwire system. The system is useful for fulfilling law enforcement requests for Fedwire transfer records, but the capability is extremely limited in comparison to the systems contemplated in this report. In the past several years it has been used only infrequently.

With an appropriate search warrant, law enforcement agencies can request a search of Fedwire records. Each search can specify up to twenty different character strings; each string can represent a distinct item (e.g., name, account number, street address), different permutations of the same item (e.g., multiple spellings of a name), or a combination of the two. Only exact matches are reported.

There are several limitations to the searches. First, searches can cover only records from the past 180 days. Records older than 180 days are transferred to microfiche and must be searched manually. Second, each search can review the records from only one Reserve Bank's Fedwire traffic. If a law enforcement agency is uncertain which Federal Reserve Banks may have processed a desired transfer, it may have to submit multiple requests. Third, searching a single day of Fedwire records takes approximately one hour and searches can only be done during hours when Fedwire is closed and after the end-of-day processing has been completed. Currently, this amounts to only a few hours each night, and this time will be reduced even further when Fedwire expands its hours of operation in 1997.

Despite these limitations, there are reasons that law enforcement agencies might wish to obtain records through Fedwire scanning rather than through records at an individual bank. It may not be known which of the 11,700 banks with access to Fedwire sent or received the transfer. Also, if law enforcement agents believe that bank employees are complicit in money laundering, or that the bank would inform account holders of the records request, then they may wish to obtain the records through Fedwire scanning.

SOURCES: Joyce Goletz, Chase Manhattan, NA, personal communication, April 7, 1995. Jim Atchley, Atchley Systems Inc., personal communication, May 3, 1995. U.S. Department of Treasury, *Foreign Assets Control Regulations for the Financial Community,* April 13, 1995. U.S. Department of Treasury, Office of Foreign Assets Control, *Specially Designated Nationals and Blocked Persons*, April 18, 1995. Louise Roseman, Associate Director, Division of Reserve Bank Operations and Payment Systems, Board of Governors of the Federal Reserve System, personal communication, May 1, 1995. Jo Ann Harris, Assistant Attorney General, Criminal Division, U.S. Department of Justice, Memorandum to All United States Attorneys, January 31, 1994.

will be used by other rules. Knowledge-based systems often employ hundreds or thousands of such rules to emulate expert reasoning within a narrow domain. The collection of rules is referred to as a *knowledge base*.

The knowledge base is the input to an *inference engine,* an algorithm that uses the knowledge base and input data to reach final conclusions that are then provided to the user. The user can query the

Figure 4-2: An Example Rule

IF	Destination bank is foreign; and amount is > $300,00; and originator is not a corporation
THEN	Wire transfer is suspicious

SOURCE: Office of Technology Assessment, 1995.

knowledge-based system to trace its pattern of reasoning.

The knowledge represented in a system's knowledge base can be acquired in one of two ways. Most commonly, knowledge bases are constructed by interviewing one or more experts in an area in ways that are meant to elicit the details of their reasoning processes. Less frequently, knowledge bases are constructed by analyzing a large number of cases where the correct decision is known. Both of these approaches are covered below.

Knowledge-based systems were developed in the 1970s largely as a result of efforts to construct two major systems: the DENDRAL system for elucidating chemical structures[1] and the MYCIN system for diagnosing and recommending treatment for infectious diseases.[2] Knowledge-based systems are now widely applied in many fields, including industry, government, medicine, and science.[3] They have been applied to a wide variety of problem types, including diagnosis, repair, and scheduling.

Link Analysis

Link analysis is a technique to explore associations among a large number of objects of different types. In the case of money laundering, these objects might include people, bank accounts, businesses, wire transfers, and cash deposits. Exploring relationships among these different objects helps indicate networks of activity, both legal and illegal (see figure 4-3).

Link analysis can indicate where to focus investigations. For example, if a person is associated with other persons or businesses that are known to be engaged in criminal conduct, then additional investigation of that individual may be warranted. Similarly, link analysis can help to confirm suspicions. For example, there may be ambiguous evidence of criminal activity for a single individual, but if that person is connected with many other persons and businesses that also appear to be involved in criminal conduct, then the analysis offers some confirmation of the initial suspicion.[4]

Link analysis operates on a set of data records, where each record has several *fields* containing information. These might be records of an individual (with fields of name, address, and phone number), bank account (account number, owner, bank), or business (name, owners' names, board members, address). Link analysis looks for matching fields in each of these records. For example, these matching fields could indicate that two persons live at the same address, deposit into the same bank account, or are involved in the same business.

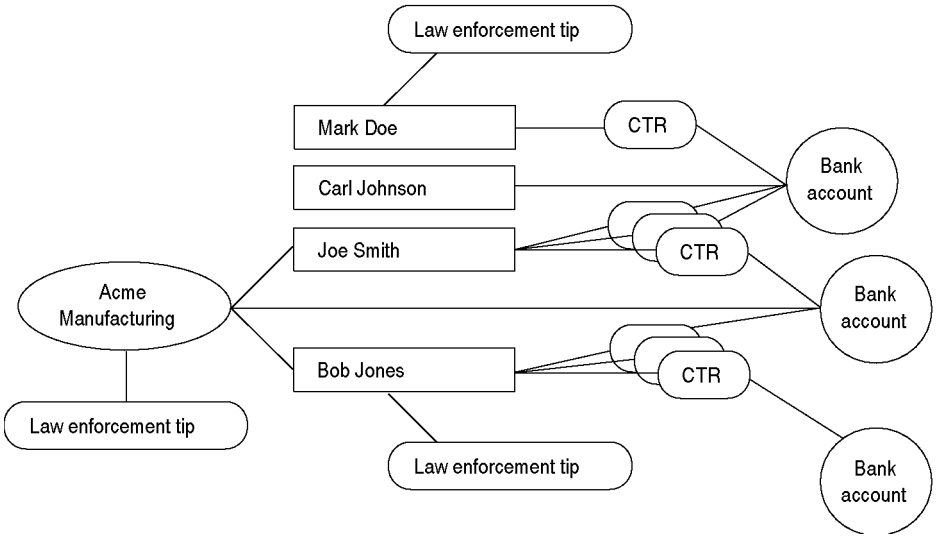
[1] B. G. Buchanan and E. A. Feigenbaum, "DENDRAL and Meta-DENDRAL: Their Applications Dimension," *Journal of Artificial Intelligence,* 11:5-24, 1978.

[2] B. G. Buchanan and Shortliffe, E. H. (eds.), *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Reading, MA: Addison-Wesley, 1984).

[3] Interested readers should consult the proceedings of a conference on AI applications held annually since 1989: *Innovative Applications of Artificial Intelligence* (Menlo Park, CA: AAAI Press; Cambridge, MA: MIT Press).

[4] For additional information, see: Malcolm K. Sparrow, "Network Vulnerabilities and Strategic Intelligence in Law Enforcement," *International Journal of Intelligence and Counterintelligence,* 5(3): 255-274.

Figure 4-3: Example of Link Analysis Results



KEY: CTR=Currency Transaction Report.

SOURCE: Office of Technology Assessment, 1995.

Link analysis is a relatively new technique, although it has quickly gained adherents in law enforcement agencies in the United States and elsewhere.[5] The field has its own journal and a professional society,[6] although these are almost entirely oriented to the use of link analysis in social science, not law enforcement. One early promoter and developer of link analysis in law enforcement is Anacapa Sciences, Inc.[7] Because of the prevalence of this company's training, many law enforcement sources refer to link analysis as "Anacapa charting." Link analyses have been used in many criminal investigations, including serial murders, fraud, and conspiracy cases.

Several commercial software packages can be used to conduct link analyses. One popular commercial package for link analysis is NETMAP from Alta Analytics Corporation.[8] NETMAP is used by both FinCEN and the Australian Transaction and Reports Center (AUSTRAC), as a part of systems developed in-house at both agencies. In addition, NETMAP is used by several state agencies investigating financial crimes by analyzing currency transaction data.[9]

[5] Clive Davidson, "What Your Database Hides Away," *New Scientist*, January 9, 1993, 28-31. Roger H. Davis, "Social Network Analysis: An Aid in Conspiracy Investigations," *FBI Law Enforcement Journal*, December 1981, pp. 11-19.

[6] *Social Networks* and the International Network of Social Network Analysts, respectively.

[7] Anacapa Sciences, Inc., Santa Barbara, California.

[8] Alta Analytics, Dublin, Ohio. NETMAP is a trademark of Alta Analytics.

[9] Besides NETMAP, there are at least three other software packages for link analysis: Criminal Network Analysis (Anacapa Sciences, Inc., Santa Barbara, California); Watson (Harlequin Group, Ltd., Boston, Massachusetts (U.S. Office)); Analyst's Workbench (I2, England). In addition, Syfact (Inter Access Consultancy B.V., Hilversum, The Netherlands) is a specialized package that uses link analysis to search financial data for indicators of money laundering and fraud.

Link analysis is useful for money laundering investigations mostly because it can integrate many different sources of information. The individual records that FinCEN currently receives, and the records that might be available under wire transfer monitoring proposals, provide few indicators of suspiciousness. Link analysis provides a way of combining these different records so that patterns of illegal activities can be discovered. While other methods can supplement it, link analysis may be the only method of analysis that allows these records to be used productively.

Link analysis is a useful way of discovering and displaying links between objects,[10] but it does not automatically construct meaning from those links. That task is left to the analyst. In the case of money laundering, analysts must make judgments about whether a network of links represents a legitimate pattern of personal and business associations, or whether the network represents a criminal organization. Links to database records that show prior criminal activity or suspicious activities (e.g., criminal referrals, suspicious transaction reports, etc.) can aid these judgments.

Link analysis is computationally intensive. Constructing links involves determining whether objects share common data values (e.g., whether a person and a business both share the same address). Consequently, rather than merely examining each record, the analysis must examine each possible pair of records, although some shortcuts can be used to reduce the necessary computation.

Even with these difficulties, however, practical limits on analysis are not unduly restrictive. Using available software and workstations, it is possible to run analyses with tens of thousands of objects. Analyses with hundreds of thousands of objects, however, exceed the capacity of available software and hardware. This indicates that wire transfer data (currently generated at a rate of nearly three million records per day) would have to be segmented or aggregated before it is combined with additional data and analyzed.

Other Techniques

In addition to the relatively sophisticated analysis provided by knowledge-based systems and link analysis, several simpler techniques are useful for screening. For example, FinCEN's FAIS computes statistics based on Currency Transaction Reports (CTRs) and other reports received by the agency. FAIS uses the value of these statistics (e.g., number of CTRs filed in past year, number of suspicious transaction reports filed in past year) to evaluate the suspiciousness of individual subjects and accounts. These statistics are a simple, but relatively powerful, way to evaluate financial records for evidence of money laundering.

Knowledge Acquisition

As previously noted, knowledge-based systems require a *knowledge base*—knowledge about money laundering encoded in ways that the system can use to make inferences. Knowledge bases can be constructed in two ways: by interviewing an expert (often called knowledge engineering) or by analyzing a large number of cases (often called knowledge discovery or data mining).

Knowledge engineering attempts to capture the relevant heuristics, or "rules of thumb," used by experts to reach conclusions in the relevant domain (e.g., wire transfers and money laundering). Knowledge engineering can be difficult, because experts often cannot easily articulate their decisionmaking processes within the narrow language used by knowledge-based systems. In addition, experts sometimes rely on broad "common sense" knowledge in order to draw useful conclusions, making the knowledge engineering task unreasonably large.

[10] The terminology used here ("objects" and "links") is not universal. Some law enforcement agencies refer to "entities" and "relationships"; the mathematical field of graph theory refers to "vertices" and "edges."

Figure 4-4: Example Data Set

Money laundering?	Dollar amount	Foreign beneficiary?	Customer type
No	110,000	Yes	Foreign exchange
No	3.5 million	No	Industrial
No	243,032	Yes	Retail
No	322	No	Individual
No	87,436	No	Bank
...	...	...	...
Yes	574,945	Yes	Retail
...	...	...	...

SOURCE: Office of Technology Assessment, 1995.

Knowledge engineering in the area of wire transfers is only possible if there are people who know how to screen transfers for evidence of money laundering. There are no human experts who scan large numbers of wire transfers and reliably distinguish between legitimate and illegitimate wire transfers. Consequently, knowledge engineering techniques are of little help in building a wire transfer monitoring system. Instead, knowledge discovery techniques must be employed.

Knowledge discovery techniques are diverse and multifaceted, including techniques from statistics and the AI subfield of machine learning. In addition, an emerging set of data visualization techniques are also gaining recognition. Several knowledge discovery techniques have been proposed for use at FinCEN, but none are now used. The boundary between screening and knowledge discovery is not a clear one, and techniques currently in use (e.g., link analysis) can be used to identify new patterns. Some knowledge discovery techniques are only useful when there are a large number of cases where the answer is known—that is, whether the wire transfer (or person, account, etc.) can be labeled as involved with money laundering or not.[11] Other knowledge discovery techniques can be somewhat useful even in the absence of such clear labels.

Machine Learning and Statistical Model Building

Researchers have developed several techniques in the past few decades for automatically finding patterns in large amounts of data. In most cases, the data consist of a large number of *observations*, where each observation represents a single object (e.g., a person, account, or wire transfer) and consists of values for each of several numeric or symbolic variables. A fragment of a fictitious data set is shown in figure 4-4.

Analysis begins by designating one variable (e.g. “Money Laundering?” in figure 4-4) as the

[11] Similarly, link analysis is only effective if some indicators of criminal activity are available. Without evidence that at least some of the objects (e.g., individuals, accounts, businesses) are inherently suspicious, it will be difficult to distinguish between legitimate and illegitimate patterns of activity. If such indicators are not present, or are not present in sufficient number, interpreting link analysis results requires an additional step—determining what patterns of associations indicate money laundering.

variable of interest.[12] The rest of the analysis consists of deriving *models* that attempt to accurately predict this variable by using the remaining variables (e.g., dollar amount, foreign beneficiary, customer type, and others in the example above). Models can be in the form of algebraic equations, logical rules, weighted networks,[13] or any other way of relating the values of one or more variables to the value of another variable.

Models are usually derived by a process of searching through large numbers of possible models. Each possible model may use different sets of variables or combine the same variables in different ways. Models that accurately predict the variable of interest are retained, while less accurate models are discarded. In many cases, it is not feasible to search through all possible models,[14] so techniques often limit the number of models searched by selectively altering the most accurate models that have already been constructed.

Technologies for machine learning and statistical model-building have existed for at least three decades, but they continue to be an active research area.[15] Interest in analysis of large databases has grown tremendously in the past five years, as major corporations have begun to "mine" large databases of customer information. This has spurred interest in massively parallel computing hardware, new algorithms for model construction, and new model forms.

Clustering

Researchers in both statistics and AI have developed methods of looking for closely related groups of objects. Cluster analysis can be used to determine underlying groupings that are not otherwise apparent in the data. For example, cluster analysis of wire transfers could be based on the frequency and dollar amount of each transfer, as well as the type of beneficiary. Such analysis could reveal groups of transfers whose originators are highly similar (e.g., brokerage houses, industrial firms, or money transmitters).

Computational techniques for cluster analysis partition a set of observations into groups based on one or more variables (e.g., frequency and dollar volume). The ultimate goal is to produce groups that differ greatly in terms of one or more variables, but where the individual members of each group differ little in terms of those variables. Figure 4-5 is an example of a graph showing several clusters in terms of two variables.

In financial data, clusters might reveal similar types of accounts, individuals, or organizations. For example, the currency and wire transactions of manufacturing firms might cluster closely together in comparison to other firms. Similarly, insurance companies might resemble each other closely in terms of their financial transactions. These clusterings might allow investigators to identify manufacturing firms whose financial transactions are atypical and examine them more closely to determine whether the corporation is merely a "shell" within which to conceal money laundering.

Visualization

Visualization techniques use color and interactive graphics to allow users to explore the relationships among two or more variables. Rather than automating the construction of useful models like machine learning techniques, visualization tech-

---

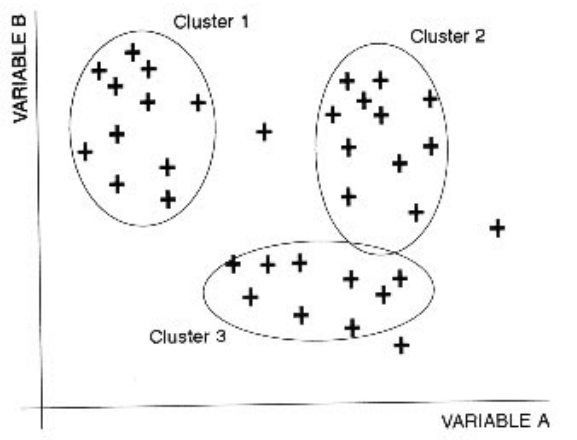
[12] Some techniques do not require designation of a specific variable of interest. An example is cluster analysis, a method that searches for groupings of observations that are all highly similar (see section below). These techniques can help an analyst understand a data set, but they do not directly help construct predictive profiles.

[13] This approach is described in more detail in a later section.

[14] Some methods do search all possible models within a limited range, but they are relatively rare.

[15] Interested readers can consult the proceedings of three groups of workshops and conferences: Machine Learning (held annually since the mid-1980s), Knowledge Discovery in Databases (held periodically since 1989), and Artificial Intelligence and Statistics (held biennially since 1985). Another source is articles in the journal *Machine Learning* (Hingham, MA: Kluwer Academic Publishers).

Figure 4-5: Example of Clustering



SOURCE: Office of Technology Assessment, 1995.

niques give human analysts powerful tools to examine data—allowing analysts to explore and apply their own knowledge to the data analysis problem.[16] In addition, visualization techniques allow analysts to apply their own abilities to recognize patterns in data, a human ability that machines cannot yet duplicate.

Other Techniques

Several other technologies are difficult to classify as either screening or data analysis, but they are potentially relevant to the problem of wire transfer analysis. Case-based reasoning and neural network technologies can be used both to derive profiles from data and to help apply those profiles.

Case-based reasoning techniques rely on the storage and processing of prototypical cases (i.e., observations), rather than deriving an abstract profile based on the values of particular variables. For example, a case-based reasoning approach to profiling wire transfer data might involve selecting records (e.g., wire transfers, CTRs, criminal referrals) that are prototypical of different classes of legitimate traffic, as well as selecting records that are prototypical of different types of illegitimate traffic (e.g., multiple cash deposits under $10,000 in a single day). These prototypical cases would then be compared to new records—helping to determine what type of activity they represent.

Neural network techniques attempt to emulate the information processing of biological networks of neurons, one of the fundamental structures of the brain. Neural networks are a set of interconnected elements called *nodes*. Some nodes are inputs and take on the values of particular variables (e.g., amount of transfer); other nodes are outputs and are used to determine the answer suggested by a network (e.g., whether a wire transfer is suspicious). Many networks also have internal, or "hidden," nodes. Nodes are interconnected and each connection has a weight, indicating the strength of the influence of the value of one node on the value of another.

By adjusting the weights on each connection, neural networks can be made to produce nearly any output based on a given set of inputs. Given a set of data where each observation contains a set of inputs (e.g., amount of transfer, foreign beneficiary, etc.) and a known output (e.g., suspiciousness), the network can be trained to implicitly recognize patterns in the input, if such patterns are present. However, one potential disadvantage of neural networks in the context of wire transfer monitoring is that they can make it difficult or impossible to "explain" why a particular transfer (or person, account, etc.) was identified as suspicious. Neural networks differ from many knowledge-based systems in this regard, because the knowledge represented within the network is not explicit or intelligible. This characteristic would cause difficulties if the results of the network's analysis needed to be explained to law enforcement agents, judges, or juries.

[16] Link analysis can be thought of as a visualization technique. Chris Westphal and Bob Beckman, "Data Visualization for Financial Crimes and Money Laundering Investigations," *Proceedings of the ONDCP/CTAC International Symposium on Tactical and Wide-Area Surveillance*, Chicago, IL, 1993.

Knowledge Sharing

Several of the technical options for wire transfer monitoring require that knowledge-based systems be installed at multiple locations. Some configurations require installation at wire transfer systems (e.g., CHIPS and Fedwire); others require installation at large money center banks, and still others require installation at many or all banks.[17]

Locating knowledge-based systems at several locations poses a unique challenge in terms of updating and maintaining the knowledge base of those systems. Because money laundering techniques can change rapidly, the profiles in knowledge-based systems intended to detect money laundering would have to change as well. Updating multiple screening systems could be done in three ways. First, all banks and wire transfer systems could be required to use a standard software package supplied by regulatory agencies. Such an approach would simplify updating but would also impose regulatory burdens, limit flexibility, and discourage innovation. Second, banks and wire transfer systems could be provided with textual descriptions of new profiles, allowing them to alter their monitoring systems appropriately. This approach would impose little burden on the federal government but would require each bank and/or wire transfer systems to recode their monitoring systems, perhaps causing long delays in the use of the profiles. Finally, banks and wire transfer systems could be provided with the profiles in a way that would facilitate updating multiple, heterogeneous knowledge-based systems.[18]

Some initial research on this latter option, referred to as *knowledge sharing*, has been conducted in the last five years. Much of the research has been conducted under the Knowledge-Sharing Effort, a project sponsored jointly by the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency, the Corporation for National Research Initiatives, and the National Science Foundation.[19] Research on knowledge sharing includes techniques to translate between different languages for encoding knowledge bases, to remove arbitrary differences between such languages, to create a standard protocol for knowledge-based systems to communicate, and to develop generic and reusable knowledge bases.

Although the research is progressing, knowledge sharing techniques are not well-developed and are substantially less mature than many of the other techniques discussed in this chapter. However, wire transfer monitoring poses only relatively small challenges to knowledge sharing. The knowledge bases that are shared are likely to be relatively small. The complexity of the domain is relatively low, given that wire transfers have a small number of fields and that wire transfer screening systems (outside of FinCEN) are likely to employ only small amounts of additional data. Finally, the use of knowledge sharing techniques can easily be phased-in over a period of time, starting with communicating profiles using relatively standard terminology, and perhaps moving toward electronic dissemination of specially formatted knowledge bases.

Data Transformation

Data transformation issues are some of the most troubling and time consuming aspects of analyzing financial records (e.g., CTRs) and experience indicates that wire transfer data are likely to present at least as many problems. For example, determining whether two different transfers originated from the same individual is not easy. Financial re-

[17] This latter possibility could involve an extremely large number of systems. There are approximately 11,500 commercial banks in the United States.

[18] All of these options would disseminate law enforcement profiles of money laundering and would pose a risk of these profiles falling into the hands of money launderers. This concern is discussed briefly later in this chapter.

[19] Robert Neches, Richard Fikes, Tim Finin, Tom Gruber, Ramesh Patil, Ted Senator, and William R. Swartout., "Enabling Technology for Knowledge Sharing," *AI Magazine*, Fall 1991, 12(3): 36-56.

cords do not always contain unambiguous indicators such as a social security number; small variations in format and spelling can defeat simple word matching; addresses are not typically provided and frequently change;[20] money launderers can use multiple, shifting account numbers. As a result, FinCEN and AUSTRAC have explored and implemented various schemes to process textual information to allow matching of names and addresses of institutions and individuals. In addition, AUSTRAC uses some approaches to understanding written text, referred to as *natural language processing*, in order to glean additional information from free text fields of wire transfers.

Other sorts of data transformations involve producing new records from existing ones. For example, FinCEN's FAIS produces new records for individual persons and accounts by aggregating data from CTRs and other reports. Fields in these records are then filled with data calculated using various statistics (e.g., number of CTRs marked as "suspicious," total cash deposits).

Both FinCEN and AUSTRAC use a database that contains both original data records (e.g., CTRs) and records constructed by the system itself (e.g., a record representing an account, constructed by aggregating a number of CTRs). This concept of a database containing both original and constructed records is nearly identical to an AI-based concept referred to as a *blackboard*.[21] A blackboard is a central database where multiple problem-solving agents can share related information about a particular problem over a period of time.[22] In the case of wire transfer analysis, the "agents" may be banks that report wire transfers, conventional computer systems that create aggregated records, knowledge-based systems that enhance or create records, or human analysts who enhance or create records.

A blackboard architecture can allow continuous enhancement and development of knowledge about potential money laundering cases over days, weeks, or months. In theory, the knowledge about those cases can be updated and developed by different analysts whose only communication is through the blackboard. In fact, many law enforcement databases can be thought of as blackboards. Agents enter reports that are used by later investigators without the need for direct communication between them even though they are geographically separated or separated in time.

DETECTING MONEY LAUNDERING

Before examining the applicability of different technologies, it is important to examine the task of detecting money laundering activity in wire transfer data. This section discusses wire transfer data, other types of data that might be combined with it, and characteristics of profiles that might be developed.

Wire Transfer Data

Wire transfers contain three basic categories of information: 1) information on the originator (name, address, account number, bank, routing number); 2) information on the beneficiary (name, account number, bank, routing number); and 3) information about the transfer itself (dollar amount, date, payment instructions, intermediary banks, internal codes).

Analyzing the relatively small amount of data in each transfer presents a surprising array of problems. These problems include the extremely large

---

[20] Addresses and other information will be mandatory under new Treasury Department regulations, although money launderers could evade these requirements by providing false, flawed, or misleading information.

[21] Ted Senator, Henry Goldberg, Jerry Wooton, Matthew Cottini, A.F. Umar Khan, Christina Klinger, Winston Llamas, Michael Marrone, and Raphael Wong, "The FinCEN Artificial Intelligence System: Identifying Potential Money Laundering from Reports of Large Cash Transactions," *Proceedings of the 7th Conference on Innovative Applications in Artificial Intelligence*, 1995 (forthcoming).

[22] A blackboard architecture was first constructed in the HEARSAY speech understanding system. L. Erman, F. Hayes-Roth, V. Lesser, and D. Reddy, "The HEARSAY II Speech Understanding System: Integrating Knowledge To Resolve Uncertainty," *Computing Surveys*, 12(2): 213-253, 1980.

number of transfers, incomplete or faulty data, heterogeneous formats and recordkeeping systems, and other difficulties for supplying cases for data analysis.

Large Volume of Data

U.S. wire transfer systems handle hundreds of thousands of transactions per day. Taken together, CHIPS, SWIFT, and Fedwire handle some 700,000 transactions in the United States each business day. This volume of data dwarfs the Bank Secrecy Act data, some 30,000 reports per day, that are currently received, processed, and analyzed at FinCEN.

Although the number of wire transfers is large when compared to financial reports currently filed with FinCEN, the size of each transfer message is quite small. For example, the current format for a Fedwire transfer is limited to 600 characters. Even the expanded Fedwire format, due to be used in 1997, will use a maximum of 1,700 characters. Wire transfers rarely use all the available characters; wire transfers in both Fedwire and CHIPS average about 300 characters in size.[23] In comparison, CTRs currently collected and analyzed at FinCEN average around 1,000 characters.[24]

The volume of reporting to FinCEN is of particular concern, given past experience with CTR reporting. Until mid-1993, the volume of CTRs far outstripped any ability to analyze and monitor them. Now the FinCEN AI System analyzes every CTR at least once, but banking industry representatives still charge that many CTRs are relatively useless and do little but impose reporting costs on banks. These concerns are behind the recent revision to the CTR reporting requirements designed to reduce the volume of these reports filed by banks. A broad reporting requirement for wire transfers could raise similar objections, but on a far greater scale.

As is the case with CTRs, many wire transfers could be excluded from required reporting by using relatively simple criteria. AUSTRAC uses exclusions to reduce the volume of wire transfer data delivered to the agency, and similar exclusions could be used in the United States.

Clearly, there is some risk to excluding broad categories of wire transfers from reporting requirements. Money launderers could attempt to make their wire transfers fall into the categories excluded from reporting. Reporting exclusions would have to take this risk into account and only exclude categories of transfers that could not easily be used by money launderers. For example, some wire transfers by banks aggregate many smaller transactions. These transfers carry little or no information about the original transactions and could be excluded on the assumption that only regulatory scrutiny could uncover money laundering by banks.

Data Transmission, Processing and Storage

Some of the technology configurations identified by OTA involve the transmission of wire transfer records from banks to FinCEN. Electronic transmission of CTRs by banks to the Internal Revenue Service (IRS) or Customs data centers is increasing, but the addition of wire transfer records could swell the volume of these electronic records by a factor of 10 to 100. The mere transmission of these data would strain current networks, and storage and analysis of these records might be beyond the capacity of current technology. A critical question then becomes whether the number of transfers transmitted might be reduced, either by exempting classes of funds transfers or by requiring banks to commit some preliminary processing of the transfers.

The security of funds transfer information is another issue, both as the information is transmitted and as it is stored. These records, if leaked or stolen, could help competitors identify a company's suppliers and customers, detail its cost structure, or predict its future behavior. Encryp-

[23] Mike Rosenberg, Senior Intelligence Research Specialist, FinCEN, personal communication, February 1995.

[24] Ted Senator, Chief, Systems Development Division, FinCEN, personal communication, February 1995.

tion suggests one manner of ensuring secure transmission, but securing the information at the federal repository is not a simple matter.[25]

There is no centralized database of wire transfers. Depending on the origin and destination of a wire transfer, messages making that transfer may flow over one or more of the three major systems (CHIPS, SWIFT, and Fedwire). Even individual wire transfer systems do not always maintain centralized databases of the transfers traveling through their system. For example, Fedwire data are decentralized in three different locations (although that will shrink to two locations by the end of 1995).

In addition, not all data are kept in a form that is easily accessed. For example, the Federal Reserve (Fedwire) keeps records online for three days, on tape for six months, and on microfiche for seven years. Bank records, although they originate in electronic form, are often stored electronically for only a short time. Some large banks keep long-term records on microfiche and some small banks keep records on paper, although banks are increasingly moving toward electronic storage. In addition, even electronic data are not always easily retrievable. For example, Fedwire data are currently indexed by sending and receiving bank only. Other fields (e.g., recipient account) may be located by using a search program to look for strings of characters, but even a relatively small number of requests (e.g., a few requests for use of the program per day) would be extremely demanding on the current system. See box 4-2 for details on the Fedwire scanning system.

The various recordkeeping and computer systems used to conduct and record wire transfers were not intended for the activities contemplated in monitoring proposals. They were intended to quickly and reliably process a large volume of wire transfers. This mission does not require centralized recordkeeping, long-term electronic storage, or quick retrieval of the sort required for law enforcement purposes. It is certainly possible to construct a system that would allow decentralized storage and retrieval of data.[26] However, it would substantially complicate wire transfer analysis and it would impose substantial new costs on banks and/or wire transfer systems.

Incomplete or Faulty Data

Some wire transfers contain blank fields or relatively useless information. Accurate information in these fields are not required for the transfer of funds, although they would be useful for law enforcement purposes. For example, some foreign banks refuse to reveal the name of the originator of a wire transfer, saying only that the transfer originates from "our good customer." Even where information is required, individuals or organizations wishing to confound analysis could provide false or misleading information.

In addition, wire transfer data sometimes contain errors. In some cases, these errors are mistakes or typographic errors made at the bank level. In other cases, the errors result from operators who use fields in ways that were not originally intended when the format of wire transfers was

[25] For example, see U.S. Congress, Office of Technology Assessment, *Information Security and Privacy in Network Environments,* OTA-TCT-606 (Washington, DC: U.S. Government Printing Office, September 1994).

[26] An example of such a system is NASA's Earth Observing System Data and Information System (EOSDIS). See Office of Technology Assessment, U.S. Congress, *Remotely Sensed Data: Technology, Management, and Markets* (Washington, DC: U.S. Government Printing Office), September, 1994.

created.[27] Occasionally, messages are returned and resent in order to correct errors in the original transmission, and this procedure could complicate simple data analysis schemes that assume each transfer of funds is only associated with a single wire transfer record.

An additional problem is created by variations in individual and business names and addresses. Many of the fields in wire transfers (e.g., originator name, beneficiary name) are entered as free form text. These fields are subject to format differences (e.g., ACME, Inc.; Acme, Incorporated, ACME Corporation; American Consolidated Mining and Engineering, Ltd.) and misspellings. These can make it difficult to identify wire transfers that correspond to the same individual or business. Additional fields, such as address and account number, can be used, but individuals and businesses can operate multiple accounts and use several addresses.[28]

Heterogeneous Data Formats and Data Types

Wire transfers vary greatly in their characteristics. For example, different classes of banks typically make different types of transfers and there are several different wire transfer systems, each with its own format. This produces wide variability in transfer records.

Money laundering analysts emphasize that many different types of entities (e.g., transfers, individuals, accounts, companies) would need to be handled by any comprehensive analysis system. Money laundering profiles developed by law enforcement and regulatory personnel involve relationships among these different entities, rather than the properties or behavior of a single entity.

Fragmentary Records

One approach to detecting money laundering would be to compare the behavior of individuals and companies to general profiles of behavior for types of individuals and companies. For example, the behavior of an individual could be compared to the behavior of others in his or her socioeconomic group. Many fraud detection systems in the credit card, cellular communications, and health care fields rely on this approach (see box 4-3).

However, these fraud detection systems have a distinct advantage—credit and cellular communications companies have relatively complete records of each individual customer, and health insurance companies have relatively complete records of each health care provider. Merely by virtue of doing business with the company, customers and health care providers must supply basic information. In addition, because transaction records are clearly designated as belonging to a particular customer, companies can construct detailed profiles of the customer's typical patterns.

In contrast, FinCEN has only fragmentary records on the individuals and companies that it investigates, and wire transfers offer little improvement in this regard. Social security numbers are not provided on wire transfers, so linking together multiple transactions would require much more effort. Much of the FinCEN AI system is devoted to accurately aggregating Bank Secrecy Act (BSA) data to form records of accounts and individuals by using inexact identifiers such as name and address. Even after this aggregation is accomplished, the resulting records form only a

---

[27] Because of problems with anomolies and errors in wire transfer messages, specific software has been designed to correct errors in some message types. For example, see: Peter Johnson, Joseph Devlin, Stephen Mott, and Jean Jans, "Applying Natural Language Understanding Technology to Automate Financial Message Processing," *Intelligent Information Access,* Proceedings of the BANKAI Workshop, Brussels, Belgium 14-16 October, 1991. Society for Worldwide Interbank Financial Telecommunication S.C. (Editors). Amsterdam: Elsevier Science Publishers, 1992.

[28] As a result of all these problems, according to the American Bankers Association (ABA), some wire messages (such as those associated with bank trust and securities) are often ambiguous enough to confuse trained and experienced human readers. ABA Comments on OTA draft material, received March 24, 1995.

BOX 4-3: Electronic Fraud Detection at the Travelers Insurance Company

Fraud is a substantial problem for insurance companies. The National Health Care Antifraud Association (NHCAA) estimates that 10 percent of all healthcare claims contain some element of fraud. Such fraud is costly to insurance companies, and they have taken steps to detect and investigate potential fraud cases.

The Electronic Fraud Detection (EFD) system assists fraud investigators at The Travelers Insurance Companies in the detection and preinvestigative analysis of health care provider fraud. The system has many similarities to proposed systems to monitor wire transfers, as well as some important differences.

In the past, fraud detection has relied upon manual inspection of claim forms and tips from internal sources, law enforcement agencies, and a telephone hotline. However, increasing use of electronic records has made automated analysis possible and has removed the possibilities of some conventional forms of fraud detection (e.g., examining paper claim forms for signs of alteration, etc.). As a result, The Travelers Insurance Companies undertook the development of EFD, a system to detect fraud using automated analysis.

Two of the challenges faced in development of EFD directly mirror problems in developing a wire transfer system. First, the company had no experts with experience screening large numbers of claim forms. The company had experts in claims processing and experts in investigating fraud, but no individuals with experience in the specific task to be addressed by EFD. Second, current data were insufficient to develop a system. The known cases of fraud were judged to be inadequate for statistical or machine learning approaches. Both problems were cited by the developers as major barriers to developing EFD.

Despite these difficulties, a system was developed that relies upon assembling a detailed statistical profile of each healthcare provider and then comparing that profile to other providers of the same type. Since each provider files a large number of claim forms, statistics can be derived, indicating the number of services of a particular type and the number of services of an unexpected type performed by a given provider. These statistics can then be compared with averages for comparable providers. For example, the statistics of a particular chiropractor can be compared to all chiropractors in the same city.

Potential fraud cases are identified when a provider differs from other providers in ways that are both statistically significant and indicative of fraud. The system uses heuristics or "rules of thumb" that indicate why a particular statistic is indicative of fraud, and what sort of deviations from average are important. For example, some statistics may not be indicative of fraud if they are lower than normal, but only if they are higher than normal.

EFD demonstrates that it is sometimes possible to construct a system where no expert and few data exist. However, there are important differences between health care fraud detection and wire transfer analysis. First, The Travelers has detailed information available on each healthcare provider because providers file a large number of claims each year. Data from wire transfers and CTRs are likely to be fewer and more fragmentary.

Second, based on the NHCAA estimate, 10 percent of all health care claims involve some fraud. In contrast, probably around 0.05 percent of all wire transfers involve money laundering. This poses a much greater challenge, since without high accuracy, an automated monitoring system would produce an unacceptably large number of false positives.

SOURCE: John A. Major and Dan R. Riedinger, "EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud," *International Journal of Intelligent Systems*, 7: 687-703, 1992.

fragmentary record of the individual or account in question.

Difficulties With Supplying Cases of Money Laundering for Data Analysis

It is difficult to label individual transfers, persons, accounts, or businesses as definitely associated with money laundering within the time frame relevant to crime detection. Years often elapse between the time that wire transfer records are generated and the conclusion of a law enforcement investigation of relevant leads or suspects. Even if criminal prosecution records were carefully matched with wire transfers, it is unlikely that concluded cases would identify all, or even most, of the records that were actually involved with money laundering. Law enforcement agencies clearly do not identify or prosecute all money laundering activity and may catch only the incompetent money launderers. Thus, by looking at any set of wire transfers, it is not possible to confidently label each as licit or illicit.

Fraud detection systems for credit cards, telephones, and health care do not suffer from this problem to the same extent. Much fraud is "self-revealing"—clearly detectable after the fact. For example, some cellular telephone fraud schemes involve "cloning" the phone of a customer with no involvement in the fraud scheme, and the customer will usually report the fraudulent toll calls when he or she receives a bill.[29]

While this self-revealing characteristic usually does not allow the fraud to be detected as it is occurring, it does provide investigators with a base of positive cases from which to derive overall patterns of fraud. Unfortunately, money laundering almost never is self-revealing. Investigators can only make inferences based on the schemes that they have caught themselves—leaving open the possibility that many other schemes may go undetected.

If imperfectly labeled data about money laundering are used in knowledge acquisition, the resulting profiles may do little more than confirm known methods. Suppose a set of data is labeled so that each known case of money laundering is used as a positive example and all the remaining cases are used as negative examples. The negative examples almost certainly contain undetected cases of money laundering, perhaps representing as many (or more) cases than are being used as positive examples. If these data are used to derive profiles of money laundering, the profiles will be "trained" to ignore negative examples—even though they may, in truth, involve money laundering. The resulting profiles will faithfully profile known money laundering schemes, rather than detect new ones.[30]

This labeling problem impairs data analysis techniques that might be used to construct profiles directly from data using techniques of statistics, machine learning, and visualization.

Additional Data

Wire transfer data don't exist in a vacuum. There are other types of data that can be used to identify money laundering. In fact, FinCEN currently uses a large number of databases to identify and analyze financial crimes. Table 4-1 details some of the types of information and the specific databases from which it is gathered.

FinCEN information comes from three basic sources: 1) the U.S. Treasury's Financial Database that contains CTRs, Currency and Monetary Instruments Reports, Casino Reports, and Foreign Bank Account Reports; 2) several databases

[29] Not all fraud is self-revealing. For example, some health care schemes involve creating entirely fictitious identities or involve the willing collusion of policyholders. See: Malcolm Sparrow, "The State of the Fraud Control Game; and the Impact of Electronic Claims Processing on Fraud and Fraud Control," Unpublished paper for the 1994 International Symposium on Criminal Justice Information Systems and Technology, 1994.

[30] Malcolm Sparrow, "The State of the Fraud Control Game; and the Impact of Electronic Claims Processing on Fraud and Fraud Control," Unpublished paper for the 1994 International Symposium on Criminal Justice Information Systems and Technology, 1994.

TABLE 4-1: Data Accessible to FinCEN

Category	Type	Selected specific databases
Persons	Name; address; former addresses; phone numbers; social security number; legal filings; criminal referrals; large cash transactions; foreign bank account holdings; travel records	Credit bureaus; news reports; U.S. Postal and commercial change of address; missing children database; phone directories; law enforcement and treasury databases
Businesses	Name; addresses; financial data; names of officers, partners, and agents; legal and regulatory filings	Dun & Bradstreet; Information America
Property	Address, sales information	Courthouse records in 11 states

SOURCE: FinCEN documents, 1995.

of criminal reports including the Drug Enforcement Administrations's Narcotics and Dangerous Drugs Information System, the INTERPOL Case Tracking System, and the United States Postal Inspection Service; and 3) commercial database services from organizations such as Dun & Bradstreet, LEXIS/NEXIS, and credit bureaus.

In addition to these databases of specific information, some useful data may involve general knowledge about money laundering activities. For example, money laundering is generally thought to employ accounts in countries with strong bank secrecy laws.[31] However, information such as this is relevant only in the context of additional information indicating criminal intent. For example, legitimate corporations use offshore bank accounts in countries with strong bank secrecy laws. This activity, in itself, is not a sufficient indicator of money laundering.

Money Laundering Profiles

Another set of challenges for wire transfer monitoring systems involves basic facts about money laundering and the current state of knowledge about it. These include the extremely low incidence of money laundering, the lack of tested profiles, the existence of temporal and spatial profiles, and the dynamic nature of criminal conduct, the similarity of licit and illicit conduct, and the need for multiple levels of analysis.

Extremely Low Incidence

The dollar volume of money laundering appears large (one estimate is $300 billion per year worldwide), but is small compared to the total volume of money moved over wire transfer systems in the United States (at least $2 trillion per business *day*, $500 trillion per year). Assuming that all money laundering moves through U.S. wire transfer systems, that each transaction moves once via a wire transfer, and that money laundering transactions are the same size as other transactions, then laundered money would account for approximately 0.05 percent of all wire transfers in the United States (see box 4-4).

The low incidence of money laundering wire transfers exacerbates the problem of false positive identifications of money laundering by an automated or semiautomated system. Because the false positive rate is likely to be orders of magnitude greater than the 0.05 percent incidence of money laundering, the ratio of false positives to

[31] In 1989, these countries were: Antigua, Austria, Bahamas, Bahrain, Barbados, Belize, Bermuda, British Virgin Islands, Cayman Islands, Costa Rica, Channel Islands, Gibraltar, Grenada, Hong Kong, Isle of Man, Liberia, Liechtenstein, Luxembourg, Monaco, Republic of Nauru, The Netherlands, The Netherlands Antilles, Panama, Singapore, St. Kitts, St. Vincent, Switzerland, and Turks and Caicos Islands. Mike Harrington and Marcus Glenn, "Methods for Analyzing Wire Transfer Data To Detect Financial Crimes," MTR-91-W00057, McLean, VA: MITRE Corporation.

BOX 4-4: What Percentage of Wire Transfers Involve Money Laundering?

It is possible to obtain rough estimates of the percentage of wire transfers that involve money laundering, based on:

- the known volume of money transferred over wire transfer systems in the United States;
- estimates of the total amount of money laundering; and
- assumptions about how money launderers use wire transfers.

Volume of wire transfers: In 1994, Fedwire transferred over $211 trillion and CHIPS transferred over $295 trillion. The volume transferred in and out of the United States through SWIFT messages is not easily estimated, although it is probably of the same order of magnitude as those of Fedwire and CHIPS. However, many SWIFT messages are automatically converted to CHIPS messages, meaning that simply adding the total dollar volumes of the three systems would result in an overestimate. For the purposes of estimation, $500 trillion per year will be used as the total dollars transferred by wire transfers through the United States.

Total amount of money laundering: Estimates of worldwide money laundering are $100 billion to $300 billion annually.

Assumptions and estimates: If it is assumed that all laundered funds move through the United States, that they are transferred only once, and that money laundering transfers are no larger or smaller than other transfers, then the percentage of all wire transfers that move through the United States and involve money laundering is between 0.02 percent (100 / 500,000) and 0.06 percent (300 / 500,000).

The estimate could be substantially lower if it is assumed that not all laundered funds pass through the United States, or that not all laundered funds that pass through the United States are sent via wire transfers. Similarly, it could be substantially higher if the same laundered money is assumed to be sent via wire transfer multiple times (in order to evade simple detection schemes). Taking both of these factors into account, OTA estimates that the total percentage of wire transfers that involve money laundering is probably less than one-tenth of one percent (0.1 percent) and that a reasonable median estimate is one-twentieth of one percent (0.05 percent). Given the uncertainty regarding the total amount of money laundering, and how money launderers use wire transfers, these estimates should be regarded as preliminary and highly uncertain.

SOURCE: Office of Technology Assessment, 1995.

true positives (even if all the true positives are captured by the monitoring system) is apt to be extremely high (see box 4-5).

A high false positive rate would diminish law enforcement's confidence in the system's capabilities. The leads produced by any wire transfer monitoring system must compete for the attention of law enforcement agents. Most law enforcement agencies contacted by OTA noted that they had far too few resources to follow up every possible lead. If most leads provided by a system turn out to be false, law enforcement agents are unlikely to use the output of the system in preference to more reliable information sources.

Lack of Tested Profiles

Building traditional knowledge-based systems involves interviewing an expert about a relatively narrow problem area (e.g., diagnosing bacterial diseases) and constructing a computer-based model of the reasoning process of that expert. Law enforcement agents or analysts do not know how to recognize a wire transfer as money laundering. If wire transfers are examined at all, they are examined in the context of an ongoing investigation, due to limits on law enforcement access to wire transfer data.

BOX 4-5: False Positives

Because most wire transfers are legitimate, an automated wire transfer monitoring system would face a daunting task. If a system merely classified each transfer as "legitimate" or "illegitimate", it would have to pick out a very small number of transfers as illegitimate, while leaving the vast majority of (legitimate) transfers untouched. Any such system will almost certainly make many errors, due to the basic laws of probability.

Assume that a system examines each of 40,000 wire transfers and classifies each as "legitimate" or "illegitimate." Further, assume that the system is reasonably accurate, correctly classifying 95 percent of the transfers (i.e., in only 5 percent of the cases does it classify a transfer as illegitimate when it actually is not, or vice versa). If the incidence of money laundering in wire transfers is 0.05 percent, then only 20 of the 40,000 wire transfers would, in reality, be illegitimate. The system could be expected to correctly classify nearly all of these transfers (19 out of 20, or 95 percent). Of the remaining 39,980 legitimate transfers, most would be correctly classified (37,981 out of 39,980, or 95 percent). However, nearly 2000 of the legitimate transfers (1,999 out of 39980, or 5 percent) would be misclassified. The system would identify them as illegitimate even thought they are not. As a result, the group of transfers identified by the system as illegitimate would consist almost entirely (99 percent) of transfers that are actually legitimate.

Even if the accuracy of the system is nearly perfect, the results are still discouraging. If the system is 99 percent accurate, then all 20 illegitimate transfers would be correctly classified, and 400 legitimate transfers would be misclassified as illegitimate. Therefore, even with a system with remarkable accuracy, nearly all of the transfers identified as illegitimate actually would be legitimate. [1]

[1] The problem of a high false positive rate has been identified in other contexts. In a 1983 study of the use of polygraph testing, OTA concluded that "the mathematical chance of incorrect identification of innocent persons as deceptive (false positives) is highest when the polygraph is used for screening purposes." This is because in screening situations, there is only a very small percentage of the group being screened that might be guilty. U.S. Congress, Office of Technology Assessment, *Scientific Validity of Polygraph Testing: A Research Review and Evaluation*, OTA-TM-H-15, Washington, DC: Government Printing Office, November 1983, p. 5.

SOURCE: Office of Technology Assessment, 1995.

Analysts at FinCEN and law enforcement agencies have little expertise analyzing wire transfers on the scale envisioned by proposals, and until they do, it will be difficult or impossible to construct a traditional knowledge-based system to analyze wire transfers automatically.[32] Another problem stems from the paucity of information contained in a wire transfer. At best, the wire transfer message contains only the names, address, and account numbers of the originator and beneficiary, information about intermediary banks processing the transfer, the amount of the transfer, and optional payment instructions. At worst, the message identifies the banks involved in the particular transfer and the account number of the beneficiary. Such information, unless combined with large amounts of other data, offers few opportunities to identify suspicious transfers.

Even wire transfer systems know surprisingly little about the transfers that flow over their systems. For example, the Federal Reserve Banks only collect information on the total dollar volume and number of transfers processed over Fedwire. The sole exception appears to be a 1987 study of a single day's traffic on CHIPS and Fed-

[32] FinCEN has attempted to arrange a pilot study to examine Fedwire data. However, there is no indication that the legal issues surrounding access to wire transfer data have been overcome.

wire.[33] However, even this study has severe limitations. It sampled only certain categories of the day's traffic and examined only wire transfers from some participating banks.

Patterns in Time and Space

If reliable indicators of money laundering activities are present in financial data, they may necessarily involve multiple transfers over a period of time between geographically dispersed individuals, businesses, and financial institutions. For example, known scenarios of money laundering involve a series of cash deposits into multiple accounts (where each deposit is under the $10,000 reporting threshold), aggregation of the funds into a separate account, and a large wire transfer out of that separate account.[34] Being able to screen for such patterns necessarily involves temporal and spatial concepts.

The need for temporal and spatial screening affects the necessary technical characteristics of a successful monitoring system. First, it emphasizes the importance of examining data from multiple locations and time periods, making localized analysis less likely to be effective—screening at a single bank or for limited time periods may identify relatively few money laundering schemes. Second, the need for temporal and spatial screening implies the need for certain types of databases and analysis tools, making them ill-suited for investigating money laundering. Some tools, particularly those developed for law enforcement (e.g., NETMAP), do allow analysis using temporal and spatial information.

Dynamic and Diverse Forms of Criminal Conduct

There are many ways to launder money. Any system that attempts to identify money laundering will need to evaluate wire transfers against multiple profiles. In addition, money launderers are believed to change their modes of operation frequently. If one method is discovered and used to arrest and convict money launderers, activity will switch to alternative methods.[35]

Law enforcement and intelligence community experts interviewed by OTA stressed that criminal organizations engaged in money laundering are highly adaptable and flexible. For example, in the past two years, law enforcement agencies have seen increased use of nonbank financial institutions (e.g., exchange houses and check cashing services) and increased use of instruments like postal money orders, cashiers checks, and certificates of deposit.[36] In this way, money launderers resemble individuals who engage in ordinary fraud. They are adaptive and devise complex strategies to avoid detection. They often assume their transactions are being monitored and design their schemes so that each transaction fits a profile of legitimate activity.[37]

Similarity of Licit and Illicit Conduct

Many patterns of transactions associated with money laundering differ little from legitimate transactions (see chapter 1). They are recognizable only because of their association with criminal activities. Banking officials emphasize that legitimate wire transfer activities in the U.S. bank-

---

[33] Federal Reserve Bank of New York, *A Study of Large-Dollar Payment Flows Through CHIPS and Fedwire*, December 1987.

[34] This scenario is also consistent with legitimate activity of some small businesses.

[35] One convicted money launderer insists that criminal organizations will know "instantly" when money laundering detection methods are changed, because they have friends in banking, law enforcement, and intelligence communities. Kenneth Rijock, interview at OTA October 6, 1994.

[36] "Current Trends in Money Laundering," Hearing before the Permanent Subcommittee on Investigations, Committee on Government Affairs, U.S. Senate, 102 Congress, Second Session, February 27, 1992.

[37] Malcolm Sparrow, "The State of the Fraud Control Game; and the Impact of Electronic Claims Processing on Fraud and Fraud Control," Unpublished paper for the 1994 International Symposium on Criminal Justice Information Systems and Technology, 1994.

ing system are diverse and wide-ranging, differing in their type, purpose, frequency, origins, destinations, and amounts. Because the ordinary traffic is so heterogeneous, it can be difficult to identify transfers that are "out of the ordinary."

Wire transfer information alone is not enough to determine legality.[38] Money laundering experts told OTA that it is nearly impossible to identify individual wire transfers as suspicious.[39] Most illegitimate uses of wire transfers mirror standard business practices. Officials at the Federal Reserve maintain that all patterns with which they are familiar are also consistent with normal business practices.

Instead, only patterns of transactions (both wire and nonwire) can indicate money laundering. Indeed, even these patterns of transactions can be made to resemble legitimate businesses. However, these data can be combined with other data in order to evaluate the suspiciousness of a pattern of financial transactions. This is one reason why every major effort to search for money laundering in financial data (e.g., those of FinCEN and AUSTRAC) employs link analysis. When data from law enforcement databases are included with financial data, it becomes more feasible to separate licit and illicit activities.

Multiple Levels of Analysis

It is useful to think of wire transfer analysis as consisting of multiple levels.[40] First is the transaction level. Money laundering necessarily involves a set of individual transactions such as currency deposits and withdrawals, wire transfers, and checks.[41] Second is the individual or account level. Multiple transactions are associated with specific individuals and bank accounts.[42] Third is the business or organizational level. An individual business may be a front for money laundering and may involve multiple accounts and multiple individuals. Fourth is the "ring" level which involves multiple businesses, accounts, and individuals in a money laundering scheme of broad scope.

The multiple levels of possible analysis indicate a flaw in analytic approaches that only examine transaction-level data. Schemes that operate at a "ring" or a business level may not be detectable through transaction analysis. Instead, the indicia of these schemes may become apparent only after aggregating data to the individual/account, business, or ring level. Analysis at any single level may miss indicators of activity at other levels. Different levels of analysis may be best done in different places. For example, banks are uniquely equipped to detect money laundering at the transaction and individual/account levels. They have access to customer information and account history which can be brought to bear on evaluating suspiciousness. In contrast, FinCEN is uniquely equipped to detect money laundering at the business and ring level. They have aggregated data and additional information from law enforcement and commercial sources that can be brought to bear.

---

[38] Mike Harrington and Marcus Glenn, "Methods for Analyzing Wire Transfer Data To Detect Financial Crimes," MTR-91-W00057, McLean, VA: MITRE Corporation.

[39] Many people compare the problem of looking for illicit wire transfers to "looking for a needle in a haystack." Ted Senator, Chief of FinCEN's Systems Development Division, notes that the problem is more analogous to "looking for a needle in a stack of other needles." Even if you examine each transfer, it is not obvious which ones are illicit.

[40] This idea is adapted from: Malcolm Sparrow, "The State of the Fraud Control Game; and the Impact of Electronic Claims Processing on Fraud and Fraud Control," Unpublished paper for the 1994 International Symposium on Criminal Justice Information Systems and Technology, 1994.

[41] However, some forms of money laundering involving bulk shipments of currency out of the United States would not involve any transactions that could be captured by monitoring U.S. institutions.

[42] FinCEN's AI System (FAIS) consolidates transactions into precisely these categories: subjects and accounts.

FINDINGS

- Many of the major challenges in constructing an effective wire transfer analysis system are related to data and not technology. In several cases, technologies are available that would be appropriate for wire transfer analysis, but data and expertise do not exist to make those technologies effective.
- There are two basic types of screening technologies: knowledge-based systems and link analysis. Effective use of knowledge-based systems requires either human experts who can accurately screen wire transfers or substantial amounts of data for which the correct analysis is already known. Effective use of link analysis requires a variety of readily available data, some of which provide indicators of money laundering activity.
- In general, there are no experts or data to make the use of knowledge-based systems feasible for detecting money laundering through wire transfer monitoring alone. However, data are available that would make it possible to conduct link analyses on wire transfers.
- The data and expertise necessary to apply link analysis already are assembled at FinCEN (the Financial Crimes Enforcement Network).