Chapter 7

Fraud Analytics

Introduction

Fraudulent activities account for billions of dollars lost in the insurance, banking, healthcare, retail, transportation, manufacturing, and communications industries each year. Likewise, fraudulent activity riddles our federal and local governments; virtually every industry is vulnerable to fraud.

The U.S. General Accountability Office estimates[1] that $1 out of every $7 spent on Medicare is lost to fraud and abuse.[2] Depending on the reference, each year Medicare loses up to $20 billion dollars[3] to fraudulent or unnecessary claims. The insurance industry (e.g., covering property/casualty, medical, life, and automobile) estimates that about 25 percent of each premium dollar is spent on covering fraudulent or inflated claims, putting the yearly costs at an estimated $30 billion nationally. In a more highly publicized type of fraud, the identity theft epidemic affects approximately 10 million people and it is estimated that over $50 billion is lost to identify theft each year.[4] To put these numbers into perspective, consider that only 82 of the 183 countries ranked by the World Bank in 2006[5] had a gross domestic product (GDP) over $20 billion. In other words, the losses from fraudulent activity in the U.S. insurance market alone exceed the GDP for more than half of the world's countries.

These numbers are staggering, especially considering that they are largely paid for by the consumer. More effective methods must be deployed to minimize these losses. Industry experts estimate that for each dollar spent on combating fraud, $5 to $15 is saved, depending on the industry being served. This return-on-investment is cumulative because it minimizes future losses for the same fraudulent activities.

Flexibility remains a critical aspect for quickly responding to changing fraud patterns. It is crucial to dynamically expose new patterns of fraud without having to reprogram, retrain, or reinvent the underlying systems. Most important is to expose the fraud before it impacts the operations or business foundations. Keep in mind that before patterns are classified they first have to be discovered. Discovering insurance fraud

[1] Stephen Barrett, "Insurance Fraud and Abuse: A Very Serious Problem," February 15, 2005, http://www.quackwatch.com/02ConsumerProtection/insfraud.html.

[2] Charging for services not performed, double billing, unbundling claims, miscoding and upcoding procedures.

[3] In 2006, Medicare benefit payments totaled $374 billion (13 percent of the $2.65 trillion in federal spending). *Medicare: A Primer* (San Francisco: Henry J. Kaiser Family Foundation, March 2007).

[4] Mary Monahan, "2007 Identify Fraud Survey Report," *Javelin Strategy and Research*. February, 2007.

[5] http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf.

is not really any different from exposing money launderers, terrorists, smugglers, embezzlers, or entities involved in elusive behaviors.

The data associated with workers' compensation, property and casualty, personal injury, and other types of insurance-related matters can be viewed in its most basic form—as interrelated objects. Generally there will be a subject (policy holder, claimant, injured party, lawyer, doctor, etc.), addresses, phone numbers, accounts (policies), and, of course, the claims themselves. How the objects are related is based on the nature of the claims submitted, and behaviors can be exposed through repeated claim submissions. It is this repeated behavior, connecting the different objects, that provides the patterns of interest. Figure 7.1 shows an example of a basic network derived from insurance claim data.

The focus is on finding anomalies in the construction of these networks where the frequency of connections and the commonality among the entities show patterns of interest. It could be something as simple as two people sharing the same phone number to something more complex, such as network of physicians and lawyers in a conspiracy, with a ring of perpetrators to inflate the damages and losses incurred. It might include a corrupt body shop providing kickbacks, padding the estimates, or not even performing the repairs. Figure 7.2 depicts and abstraction of a collusive network of entities that emerges across multiple accident claims.

There have been a multitude of new technologies introduced into the antifraud marketplace over the past several years, including link analysis and other systems for detecting nonobvious relationships and associations. Perhaps even more important are the refined analytical methodologies that help to interpret the complex networks and patterns presented by these technologies. Better understanding of the data will inevitably lead to better pattern detection, and ultimately, lower fraud incidence. Once a pattern has been exposed, it is up to the affected company to act on that knowledge by changing business processes to flag related or similar

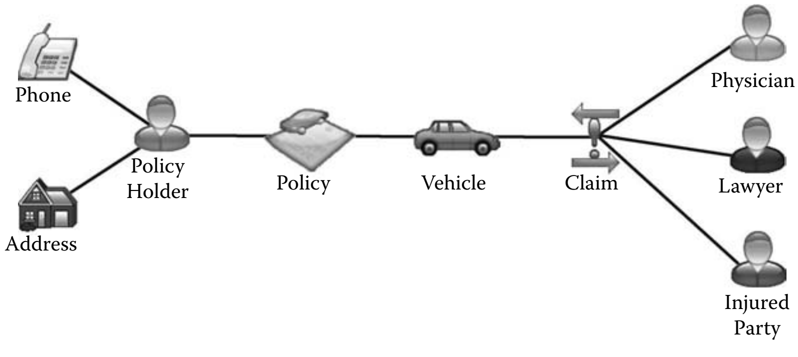


Figure 7.1 Sample insurance analytical model.

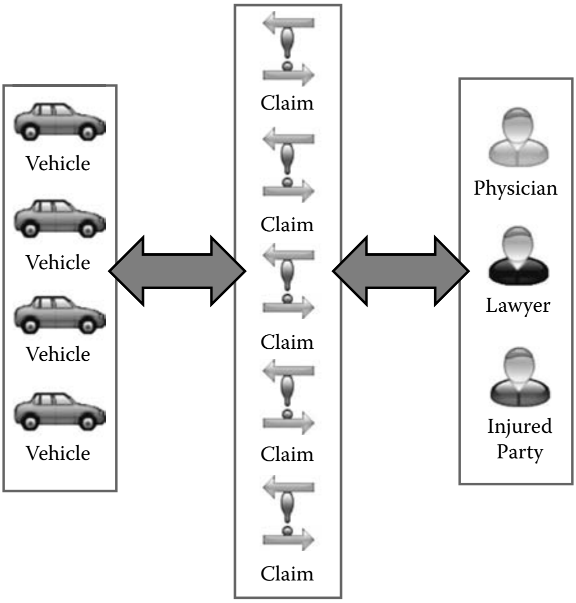


Figure 7.2 A collusive network of entities.

occurrences of the pattern. Remember that there are always exceptions to the rule, and there are exceptions to the exceptions.

Warranty Fraud Anecdotes

Warranty fraud comes in all flavors and covers a wide number of industries ranging from computers to home appliances. It comes from a mixture of consumers wanting to cover or minimize their repair costs to the authorized service representatives blatantly submitting false warranty claims. In fact, one of the largest technology providers, Hewlett Packard (HP), spends approximately $1.8 billion a year on warranty claims[6] and has determined that 6 to 8 percent is fraudulent.[7] HP estimates that the loss of $140 million to warranty fraud would be equivalent to the profit generated on the sale of an additional 15 million printers.

In HP's case, the fraud is committed in a variety of ways, including swapping new units for refurbished models or simply manufacturing false repair orders and submitting fabricated claims. In one particularly

[6] Top 100 Warranty Providers, *Warranty Week,* January 10, 2007, http://www.warrantyweek.com/archive/ww20080110.html.

[7] http://www.warrantyweek.com/archive/ww20050419.html.

shameless scenario, HP's warranty process and systems manager[8] said, "Companies sent staff into computer retail storefronts in search of floor models from which they could copy down the serial numbers. Worse, each seemed to share the serial numbers they gathered with the other company. Over a span of 12 months, these scammers cost HP an estimated $2 million."

Perhaps the biggest and most costly warranty repairs stem from the automotive industry. In the first nine months of 2007, Ford Motor Co. spent over $2.8 billion (2.5 percent of product sales) and General Motors almost $3.4 billion (2.6 percent) in warranty claims. Based on industry estimates that an average of 6 percent of revenue is lost to fraudulent activity, the amount of warranty fraud losses for just these two companies would be $168 million and $204 million, respectively. Any improvements in fraud detection help impact these numbers in a positive fashion and can result in significant savings to the manufacturers.

Automobile Warranties

This discussion describes a scenario where the auditing department of a foreign automobile manufacturer was concerned that they had fraudulent warranty claims being submitted by their affiliated dealerships. They were not sure where it was or what it looked like, only that they knew it was there. As with all car manufacturers, they pay dealerships to perform maintenance warranty repairs (e.g., three years, 36,000 miles, bumper-to-bumper) on its cars to fix vehicle problems and satisfy customer complaints. The charges that are incurred by the manufacturer reflect the costs for parts, labor, sublets (outsourced work), and miscellaneous expenses (see sidebar on dealership charges). With more than 1,200 dealerships in North America, the amount paid out annually by this particular manufacturer for warranty repairs exceeded $350 million at the time the analysis was performed. Because dealership mechanics are paid based on the amount of work they generate (see sidebar), there is a potential for some repair orders to be padded with extra costs for work that has not been performed or are considered inappropriate charges. Even small percentages of fraud result in significant losses when scaled to this type of industry.

This particular manufacturer had established a progressive warranty audit team that was chartered with identifying unallowable charges and patterns of noncompliant claims. The team recognized

[8] William Fung, PC & Appliance Warranty Fraud Panel*, 1st Annual Warranty Chain Management Conference*, San Francisco, March 3, 2005.

Dealership Charges

In many repair shops, including car dealerships, mechanics are typically paid[9] on a flat rate (also called a book rate) that represents the average time required for performing a specific repair. The flat rates are published by the manufacturers as a way to ensure consistency and reasonableness in the time that is allowed to be charged for completing the repair. Say the flat rate for completing an oil change is set at 15 minutes; this is the total amount of labor time the mechanic will be paid regardless of whether it takes 10 minutes or an hour to finish the job. Examples of published manufacturer flat time rates include: 3.2 hours[10] to replace a water pump on a 1989 Chevy G20 Van with a 6.2 liter diesel, 2.6 hours[11] to replace the oil pan gasket on a 1993 Ford Ranger 3.0 v6 with an automatic transmission and two-wheel drive, and 1.5 hours[12] to replace the latch assembly on a roll-up door for a Trainmobile trailer.

The more jobs a mechanic can get done in a day, the more he or she will be paid. In fact, it is possible for decent and experienced mechanics to log more than eight hours of book time (e.g., 12 or even 15 hours) in an actual eight-hour day. This can lead to a nice, plump paycheck for the mechanic, and there is some debate as to whether or not the use of flat rates leads to rushed jobs that are not properly completed, inevitably requiring the same job to be redone multiple times. This situation is further compounded because many service advisers are paid by the number of labor hours they sell, and of course the dealerships are also paid an overhead cost plus additional fees to process the warranty claims. Everyone makes a cut on the warranty repair. As a side note, some manufacturers use the concept of a "warranty time" to further reduce the flat-rate time required to complete a repair.

[9] http://wiki.answers.com/Q/How_much_money_does_an_auto_mechanic_earn.

[10] http://autorepair.about.com/library/faqs/bl983e.htm.

[11] http://www.autoqna.com/Maintenance-Repairs/1024-2-autoqna-3.html.

[12] http://www.trailmobile.com/site/files/638/54471/213436/285691/TMFlatRateMaintenance.pdf.

the need for using advanced technologies and analytical techniques to help with processing the almost 2 million warranty repairs that are performed each year by their authorized dealerships. They wanted to become proactive in their audits so that they could effectively seek out and discover the fraudulent claims. They realized it would require an analytical system that could support a number of parameters ranging from dealership regions and repair types to car models and/or mechanic training, where virtually any one of the hundreds of variables contained within their datasets was fair game. In this example, a number of different data sources were identified for analysis including warranty repair orders, customer complaints, technician training, and customer satisfaction survey data.

The patterns exposed can be generally broken down into four different types of categories,[13] which are outlined below:

- **Routine and Known:** These types of warranty activities occur based on known risks and probabilities. The exceptions are "flagged" as being outside of standard parameters. For example, the manufacturer will allow the dealerships to use only certain parts and if a part is submitted that is not found in the standard parts list, the system will reject the entry. This represents standard business operations and procedures.
- **Routine and Unknown:** The nature of these warranty repairs is based on taking advantage of situations where existing systems have limited detection capabilities. The size and scope of these activities are left to be uncovered through alternative methods. For example, a dealership performs routine transmission repairs, but does not employ technicians trained at the required levels.
- **Nonroutine and Known:** The warranty repairs slotted into this category are based on discontinuous patterns. The circumstances occur based on unfamiliar sequences of activities. For example, the warranty repairs for a particular dealership are above average because the time of year for that geographic region or zone has fewer COD clients (i.e., nonwarranty-covered repairs), requiring the technicians to make up their extra pay through increased warranty work.
- **Nonroutine and Unknown:** This is the most damaging situation directly affecting the manufacturer. The models used here help identify unknown patterns and practices, detect covert/unexplained practices, and have the capability of exposing organized activity. The types of behavior that occur in this category are yet to be discovered and are of most interest and value to the manufacturer.

One of the most notable patterns found in this manufacturer's warranty database was based on an initial query looking at claims involving vehicles with less than 100 miles on the odometer. Generally, this is a somewhat unrealistic mileage for performing warranty work. Unless the defect renders the car unusable, such as a broken starter motor, or makes it annoying or uncomfortable (e.g., squeaking brakes or a broken A/C unit), most people won't report the issue until they bring the car in

[13] Christopher Westphal and Teresa Blaxton, *Data Mining Solutions: Methods and Tools for Solving Real World Problems* (New York: John Wiley & Sons, 1998), 62.

for its first oil change around 3,000–5,000 miles. Therefore, the resultant set of claims basically reflected new vehicles that were most likely still located on the dealership lot and had not yet been sold. Additionally, the data extraction used for the analysis was set to remove any part replacement codes—showing labor-only repairs (e.g., like soft-tissue injuries in insurance fraud claims) exclusively. These claims are based entirely on a mechanic's time under the hood where no parts were replaced (and, therefore, not traceable) and no work was outsourced to any external or third-party entity (e.g., radio repairs).

The results of the query were subsequently presented using visual-clustering techniques. When the "repair type" was used for grouping the claims, it became explicit that there was a dominant group (i.e., a specific type of repair) represented in this data. Surprisingly, the most prevalent group in this set was *cigarette lighter* repairs and each claim had a single hour of labor time charged that, depending on the dealership rates, was between $45 and $65. After reviewing the pattern with the audit group, it was obviously clear—when people test-drive cars, the cigarette lighter is often removed (i.e., stolen), and because it is both a functional as well as cosmetic component in the car, the dealerships were charging an hour of labor to recover the cost of replacing the part. The general flow of this pattern is depicted in Figure 7.3.

Needless to say, this was not a circumstance that the manufacturer was required to cover under warranty. In fact, the manufacturer could go back, up to three years, to deny any paid claims. For the time period reviewed, the number of dealership franchises that existed, and the number of claims made of this type, this was a multimillion-dollar fraud. Nowadays, to help deal with this situation cigarette lighters are packed in a welcome kit that comes with the vehicle—often encased in a plastic binding that is unpackaged once the car is prepped after the sale is made. Generally, cigarette lighters have been repurposed into power ports to

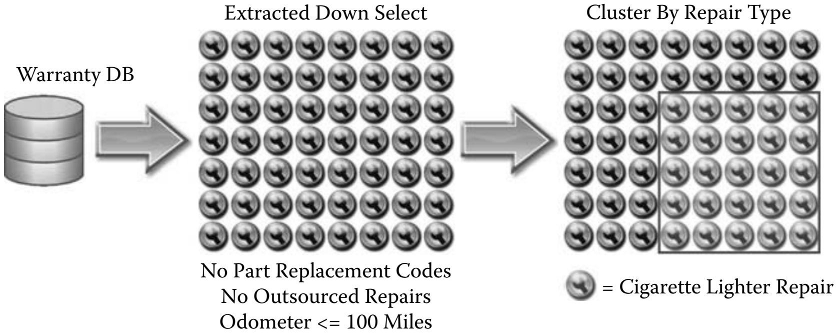


Figure 7.3 Cigarette warranty claim repairs.

run modern electronic devices, such as radar detectors and GPS satellite navigation systems. For some manufacturers, it has become an option that costs extra.

Other types of warranty patterns were also identified in the dataset. One particular make and model distributed by the manufacturer had a faulty paint job, such that the paint flaked off certain areas of the car (e.g., hood, roof, and trunk), for a period of approximately two years. This was a situation known by the warranty team, but was immediately spotted within the data. What came to light was the range of costs associated with this type of repair. Generally, the claims were between $400 and $800 to repaint the affected areas. However, unknown to the auditors was that there were quite a number of claims exceeding several thousands of dollars. The only justification for those claim amounts would be an entirely new paint job and/or body work.

Other patterns included repair "itises" (where certain types of claims are always done together even though the problems are unrelated), duplicate repair submissions, having the same problem fixed multiple times, and even tracking customer complaints and issues before they reach a boiling point (i.e., customer will never buy another car from this manufacturer again). Below is a simple list of other repair patterns that were reviewed along with their colorful name definitions in brackets.

1. Look for commonality among dealerships based on customer name, VIN, address, telephone number. [Merry-Go-Round] (Similar to the HP example discussed previously.)
2. The warranty repairs for a dealership exceeded the zone average.
   a. Dealership charges excessive warranty repairs to lot cars based on problems indicated during a test drive. [Bait & Switch]
   b. Service adviser is observed authorizing additional time expenses exceeding preset time allotments for repairs. [Padding]
   c. Technicians spend a majority of time working on nontraceable repairs (e.g., rough engine) to inflate warranty charges. [Soft Repair]
3. Vehicle requires multiple "remove and replace" repairs. [R&R]
4. Customer invoice indicates repair X and the manufacturer invoice shows repair X, repair Y, and repair Z. [Creative Repair]
5. Car is brought in for a specific problem and the dealership identifies several additional "nonsafety" repairs. [Ambitious Technician]
6. Repair technician has been trained to a certain level (A–E) and repair order indicates problem skill is above certification level of technician. [Brainiac]
7. Dealership charges the customer for a repair and then submits it to the manufacturer as a warranty claim. [Double-dipping]

There are also reports of dealerships initiating warranty claims, forging signatures,[14] and rolling back the odometers to qualify for warranty work.[15]

Traditional methods limited manufacturers to only a few detailed audits each quarter. With the help of more automated systems, they have the capability to review and audit hundreds of dealerships by focusing on the anomalies that present themselves within the warranty data. Through development of the system, the manufacturer also has the ability to effectively identify and detect unallowable warranty repair orders. This allows them to deny and directly charge the costs back to the dealerships. Those mechanics, service advisers, or dealerships that have an excessive or recurring denial of claims can often be traced back to a misinterpretation of established manufacturer policies and procedures. The audit department, through close association with the training department, can recommend that corrective processes be initiated in these cases.

Hurricane Katrina

Pre-9/11 the big concern was counter-*narcotics,* and post-9/11 it is counter-*terrorism* (ironically, both are nine-letter words). After the attacks there was massive spending to ensure that public services personnel, including police, fire, and emergency workers, could communicate with one another; there were large quantities of anthrax antidotes, biohazard suits, and gas masks stockpiled throughout the country; and huge numbers of personnel including Transportation Security Administration (TSA) inspectors, air marshals, and special agents were hired to help battle this new and emerging threat. Although critical, it represents a significant amount of additional investment to deal with situations after they have occurred. Is the United States, or the world, a safer place? By all accounts it is, but that is largely based on how one defines "safer."

Regional emergencies, such as Hurricane Katrina,[16] certainly showed that there is still a lot of room for improvement with respect to how data is used to better manage situations, services, and operations. The GAO has reported[17] that there has been at least $1.4 billion in fraud,

[14] http://scholar.lib.vt.edu/VA-news/VA-Pilot/issues/1997/vp970807/08070482.htm.

[15] http://www.usdoj.gov/usao/eousa/foia_reading_room/usam/title4/civ00159.htm.

[16] "Waste, Fraud and Abuse in Hurricane Katrina Contract," United States House of Representatives, Committee on Government Reform—Minority Staff, Special Investigations Division, August 2006, http://en.wikipedia.org/wiki/Hurricane_Katrina.

[17] Gregory Kutz and John Ryan, "Hurricanes Katrina and Rita Disaster Relief: Improper and Potentially Fraudulent Individual Assistance Payments Estimated to be between $600 Million and $1.4 Billion," GAO-06-844T, June 14, 2006, http://www.gao.gov/new.items/d06844t.pdf.

misrepresentation, and theft from the coffers that were put in place to aid the victims of Katrina, and there are billions[18] more in overcharges and mismanagement by contractors providing relief and recovery services, often awarded without a competitive bidding process. Certainly the aid provided was desperately needed by many people and used for the appropriate purposes, but without any type of oversight or controls in place to properly manage this process, the government doled out much more money than was required for this crisis.

After the hurricane hit, the levees burst, and the damage was done, the Federal Emergency Management Agency (FEMA) offered $2,000 in disaster assistance through the use of prepaid debit cards to those people in need to help cover immediate food, shelter, clothing, and basic living necessities. To be eligible, recipients had to have a primary residence in an area damaged by the hurricane. Additional funds, including disaster unemployment assistance, were also made available from the Louisiana Department of Labor (LDOL) as well as other state and local agencies. Other benefits[19] included housing assistance (e.g., manufactured housing and mortgage and rental assistance), individual and family grants, and a number of funding avenues to help offset the massive losses encountered during this crisis. In total, over 2.5 million applications requesting disaster assistance were received by FEMA.

Unfortunately, the management of the crisis was not handled well by FEMA officials and proper oversight and controls were largely lacking from their assistance programs. As such, there are numerous reports[20] of people receiving duplicate aid payments from FEMA, and other questionable expenditures, such as purchases of alcohol, prostitutes, tattoos, weapons, and the paying of gambling debts, traffic fines, and adult club fees. There was even one report[21] of a person giving the address of a cemetery for their claim information. The true nature of the abuses will never be fully realized because a significant amount of the aid provided through the distribution of the debit cards was converted to cash and, is therefore, untraceable.

In one particularly shameless case of fraud,[22] an individual who lived in an area located a short distance from downtown Atlanta, Georgia, and some 450 miles from New Orleans submitted more than 50 fraudulent

[18] http://oversight.house.gov/Documents/20060824110705-30132.pdf.

[19] http://katrinalegalrelief.org/index.php?title=FEMA_Benefits.

[20] Gregory Kutz, "Expedited Assistance for Victims of Hurricanes Katrina and Rita: FEMA's Control Weakness Exposed the Government to Significant Fraud and Abuse," GAO-66-403T, February 1, 2006, http://www.gao.gov/new.items/d06403t.pdf.

[21] Frank Bass, "FEMA Wants More than $300 Million in Hurricane Aid Returned," *Associated Press*, February 6, 2007.

[22] http://www.usdoj.gov/katrina/Katrina_Fraud/pr/press_releases/2006/jan/01-20-06whittakerindicted.pdf.

applications for disaster unemployment assistance. Basically, he fabricated a number of names, all sharing the same date of birth, using one of two common last names and false Social Security numbers (SSNs) that were very similar to each other (e.g., one or two of the middle digits of the SSN were changed in an incremental numbering fashion). He claimed that these people lost their jobs as a result of the hurricane and then received dozens of debit cards that were all mailed to the same post office box in Georgia.

He was eventually caught because automated methods within the LDOL computer systems flagged that multiple claims/payments were being made to the same address. What is particularly interesting about this case is that on September 16 (2005) he filed a claim using his own name, and then on September 27 filed a new claim using a completely different name. Presumably, after receiving payment without additional follow-up, questions, or any type of red flag being raised by FEMA or the LDOL, he decided to take advantage of the circumstances. His next set of claims came on October 27, when seven claims were filed under different first names using a common Latin surname, each with a slightly different SSN and all with the same date of birth and same address. The next day, on October 28, he filed another 32 claims using a similar configuration of the same surname (different first names), date of birth, and address. Finally, a few days later on November 1, an additional 10 claims were filed using a different last name (also a very common Latin name) with the same date of birth and address as the claims filed the previous week. An abstraction of this data is presented in Figure 7.4.

To the credit of the government personnel involved, this case was sent to the U.S. Attorney's Office on November 7 for prosecution and the last 10 payments, for the November 1 claims, were stopped before payment was disbursed. In fact, all the debit cards were electronically zeroed out once it was determined there was a potential fraud. Once the fraud pattern was detected, it was acted upon and shut down in fairly short order helping to minimize the damage and losses of money to fraud.

The LDOL normally has a number of safeguards in place to help minimize the risk of fraud, but based on the mitigating circumstances, many of these safeguards were removed for the first 12 weeks to help expedite the claim processing. Under normal operating conditions the standard process requires all recipients to call in each week to request money and answer specific questions about their accounts; however the phone systems were down because of the widespread damage inflicted by the hurricane and the proper follow-ups could not be conducted. It was decided that debit cards would be used to fund the employment benefits because the postal system was also devastated by the hurricane and

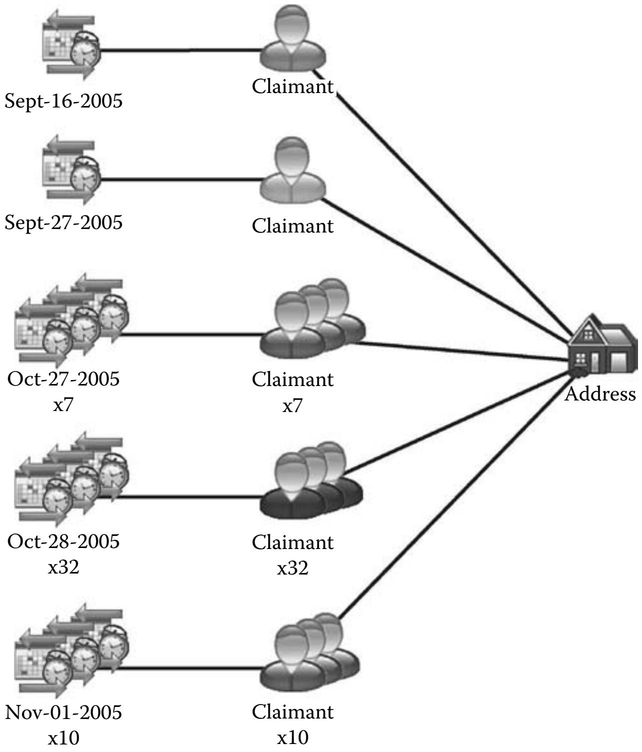


Figure 7.4 Fraudulent Katrina benefit claims.

not operational in many areas, further compounding the situation and resulting in more delays or potential nonpayment to many needy persons. Therefore, electronic funds passed via a debit card were the most logical alternative to ensuring people received their benefits in a timely manner. The individual in question eventually pleaded guilty,[23] was sentenced to 27 months of prison, and ordered to pay restitution.

Other patterns of fraud were discovered when the investigators saw an influx of applications coming from a specific street, apartment complex, or area not affected by the hurricane. Typically, someone would file a fraudulent claim, receive payment, and brag about it to their neighbors and friends, prompting others to start filing fraudulent claims to receive debit cards. Many of these cases were exposed by the tip line established for people to report suspected wrongdoing and fraud. Once the applications got pulled up for review, the investigators quickly saw the pattern and were able to deal with the situation accordingly.

[23] http://www.usdoj.gov/katrina/Katrina_Fraud/pr/press_releases/2007/jul/07-05-07 whittaker.pdf.

Disturbingly, there were also instances of government employees from FEMA arrested for soliciting bribes as public officials. Several FEMA individuals[24] running base camps located in Louisiana inflated the head counts for the meal services being run from their facilities in return for kickbacks from the contractor supplying the meals. Still other officials, from both the federal and state governments including police organizations, were charged with theft of property, filing false claims, and even overcharging for labor and vehicle use. One large-scale incident[25,26,27] involved a call center operated by the Red Cross located in Bakersfield, California. The volunteers staffing the call centers filed fraudulent claims for themselves, family, and friends due to the minimal amount of data required to issue a claim number to collect the funds. One person went to the same Western Union office on three different occasions to collect payments, which aroused the suspicion of store employees who reported the incident to the Red Cross. This led the investigators to follow the thread, which ultimately led to more than 80 prosecutions within the Eastern District of California resulting from this scheme. There are many references[28,29] in the open-source reports to a wide variety of fraud associated with this disaster.

On September 8, 2005, within two weeks of the hurricane landfall, the Hurricane Katrina Fraud Task Force[30,31] was set up by the U.S. Attorney General with the expectation to address the frauds and abuses associated with the aftermath of disasters. Within the first six months[32] of the establishment of the task force, there were more than 200 people charged with fraud-related crimes. After a full year,[33] there were over 6,000 fraud-related tips and more than 400 people charged with fraud crimes from hurricanes Katrina, Rita, and Wilma. Unfortunately, trying to re-collect the money once it is distributed is a much harder task than being more diligent when processing the aid requests in the first place. Of course, due to the severity of the situation, some of these cases were unavoidable.

It must also be pointed out that not all of the fraud reported in the press and news is due to incompetence or the inability of the government

[24] http://www.usdoj.gov/katrina/Katrina_Fraud/pr/press_releases/2006/jan/1-27-06USAOEDLA.pdf.

[25] Kareen Wynter, "Dozens Indited in Alleged Katrina Scam; Red Cross Workers Accused of Filing False Claims," *CNN,* December 29, 2005. http://www.cnn.com/2005/LAW/12/28/katrina.fraud/index.html.

[26] http://sacramento.fbi.gov/dojpressrel/pressrel06/katrina_fraud070306.htm.

[27] http://www.usdoj.gov/katrina/Katrina_Fraud/pr/press_releases/2006/mar/03-17-06eightindicted.pdf.

[28] http://www.publicintegrity.org/katrina/filter.aspx?cat=14.

[29] http://www.usdoj .gov/katrina/Katrina_Fraud/pr/press_releases/.

[30] http://www.usdoj.gov/katrina/Katrina_Fraud/.

[31] http://www.usdoj.gov/opa/pr/2005/September/05_ag_462.htm.

[32] http://0225.0145.01.040/katrina/Katrina_Fraud/docs/katrinarerportfeb2006.pdf.

[33] http://www. usdoj.gov/katrina/Katrina_Fraud/docs/09-12-06AGprogressrpt.pdf.

to detect these schemes. There are many restrictions, due to privacy laws, that make it hard to deal with these situations. It becomes a balancing act between exposing criminal behavior and protecting an individual's privacy. The Computer Matching and Privacy Protection Act of 1988 (5. U.S.C. 552a), an extension of the Privacy Protection Act of 1974,[34] defines the regulations for record keeping, disclosures, and sharing of data. These laws put a number of limitations and restrictions on what different agencies can do with respect to their use of data sources. Civil liberty rights groups have long espoused their concerns regarding the potential abuses involved with collecting and combining data from multiple sources and were instrumental in the downfall of the Total Information Awareness (TIA[35,36]) program sponsored by the Defense Advanced Research Projects Agency (DARPA) back in 2003.

Before the hurricane, there were no Memorandums of Understanding (MOUs) in place between FEMA and the Social Security Administration (SSA) and, therefore, many reported identities could not be verified. Even after the establishment of the Task Force, access to the National Emergency Management Information System (NEMIS) operated by FEMA (used to enter and manage all information regarding disaster assistance from registered applicants) remains tightly controlled. An official MOU was executed between the Task Force and FEMA with access granted only to approved staff members and only for use checking a specific allegation or fraud under the premise of law enforcement protocols. This ensures that there are no fishing expeditions or witch hunts being conducted by the government.

To be fair, it should be recognized that the $1.4 billion estimate made by the Government Accountability Office (GAO) is an extrapolation from a sampling of claims and includes losses from both fraud and mismanagement. This number includes applications that were filed where the information did not properly or adequately support the claim being made and technically should not have been paid. This occurred, for example, in about 2,300 applications where a post office box was listed as the physical address of the property damaged. In a high percentage of these claims, investigators were able to confirm, through the postal databases, that the victims actually had real property and residences located within the affected areas damaged by the hurricane. This type of filing happened so frequently because the physical property no longer existed and the applicants mistakenly or inadvertently put the contact address into the wrong

[34] http://www.usdoj.gov/oip/04_7_1.html.

[35] http://en.wikipedia.org/wiki/[nformation_Awareness_Office.

[36] http://www.epic.org/privacy/profling/tia/.

part of the application. Therefore, the value for the entire lot of claims was deemed unacceptable and included into the baseline losses reported by the GAO, which skews the total number.

Corporate Frauds

In the commercial world, there are innumerable ways in which to conduct internal frauds against a corporation, including improper billing practices, padding expense reports, filing duplicate invoices, submitting fictitious receipts, tampering with checks, or voiding cash entries—the list is virtually endless. Frauds can be perpetrated throughout the corporate hierarchy, from top management officials involved in complicated investment scams all the way down to the mailroom clerk stealing from the petty cash drawer. Generally, the amount of loss incurred by the business community in the United States is estimated[37]at approximately 5 percent, which translates to over $650 billion in fraud losses for 2006, and this number is expected to continue to rise.

Fraud is basically a theft against the organization and is performed in a concealed or stealthy manner so as to avoid detection. Fraud has many different names, including embezzlement, bribery, kickbacks, forgery, falsification, and conflicts of interest, to name a few. One particular conflict of interest comes in the form of procurement fraud, where purchasing agents earmark contracts for a favored or preferred vendor without requiring competitive bids. This situation can also manifest itself in a pattern of employees also acting as vendors of the corporation—where they might have inside knowledge regarding the budgets, specifications, or competition bidding for the work.

Employees as Vendors

Some very basic and fundamental checks can be performed on company data sources to check addresses or phone numbers from an employee master file against the vendor master file to expose any potential commonalities. For example, the fax number listed for a company turns out to be the same as the number listed for the emergency contact information of an employee. An illustration of this is shown in Figure 7.5. Although simple, it does occasionally expose some questionable

[37] 2006 Report to the Nation on Occupational Fraud and Abuse. *Association of Certified Fraud Examiner, Inc.,* http://www.acfe.com/documents/2006-rttn.pdf.

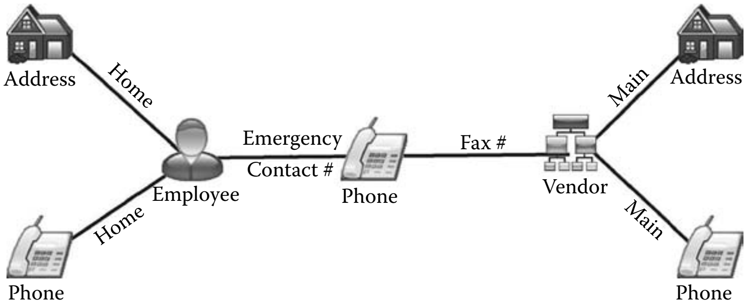


Figure 7.5 Employee linked directly to a vendor.

relationships within the operations of a business, especially some of the larger entities.

In other cases, there could be less obvious connections that require the incorporation of additional sources of data. Many times corporations are run or influenced by a cadre of people including owners/founders, senior management, and board directors. Often, these people also have similar roles in other corporations. Therefore, knowing the chain of command often helps in understanding how decisions can be influenced. For example, the diagram[38] in Figure 7.6 depicts a large, well-known U.S. retailer at the center of the network and all of its directors as its immediate linkages, shown as male or female person icons. A number of these board members are also affiliated with other large companies addressing a wide spectrum of business offerings, including equipment manufacturing, computer sales, cloths retail, news media, insurance, investment and financing, communications, restaurant, banking, and others. As the diagram shows, several are also in common to multiple companies, showing how a select few can have large impact across a number of different industries. Of course, each of these businesses can be further expanded thereby extending the network of influence even more.

A second example of this is shown in Figure 7.7. This information[39] is more directly based on corporate ownerships and which companies own other companies. Each company will have a president or a CEO that will act as the official figurehead and there will be a multitude of senior executive vice presidents and others not explicitly listed. Some individuals can prove to be quite active and represent multiple companies or subsidiaries at the same time, as is shown in Figure 7.8. These just represent other avenues where improper procurement practices could be encountered due to the indirect nature of these kinds of relationships.

[38] Generated from *http://www.theyrule.net* using data circa 2004.

[39] Can be derived from sources like Dun & Bradstreet.

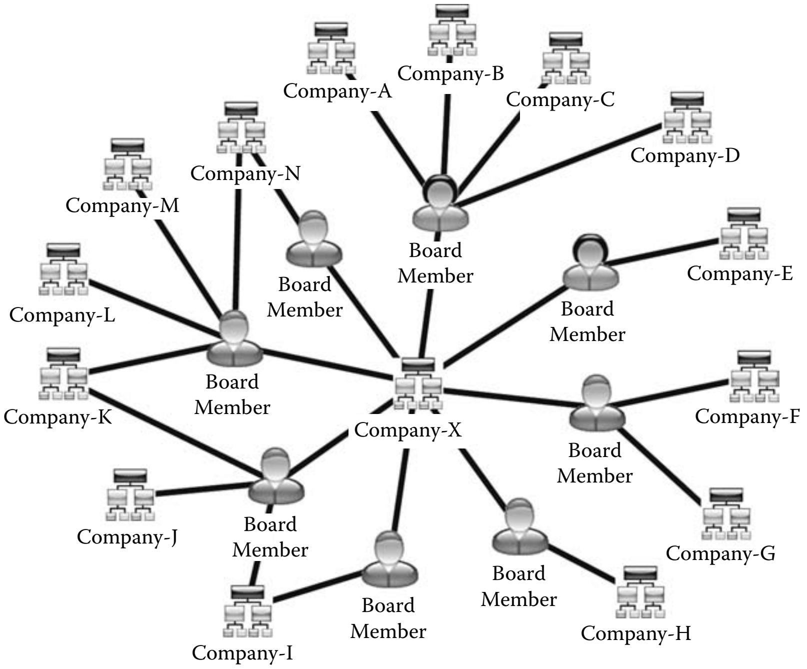


Figure 7.6 Corporate board member intrarelationships.

Vendors as Vendors

As was introduced in the previous section, companies can operate in an official capacity and subsume, control, purchase, own, and influence other companies. There are also more cozy and comfortable relationships that are forged where a company (e.g., vendor) is indirectly associated with other vendors, potentially doing similar work. There is concern about these types of situations because a vendor might act as a front company, submitting unreasonably high quotes for a job only to make another vendor look more favorable, yet both companies are owned or controlled by the same entity. Figure 7.9 shows an example of two companies using the same phone numbers for both their main call-in lines and fax numbers.

The same circumstances also exist when the organizations share a common address. Figure 7.10 provides a depiction of this type of network. There may be legitimate reasons for such activity, including, for instance, when the facilities represent a large, shared warehouse space and a distributor handles the related processing originating from the same place. This is somewhat of a stretch and, in reality, the CFO of the agency should investigate why such conditions exists.

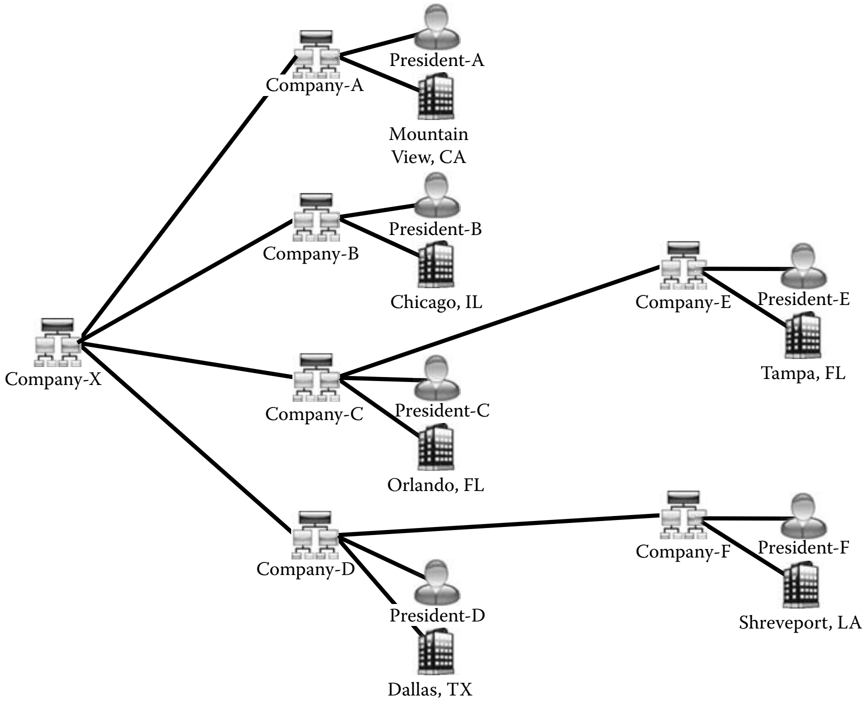


Figure 7.7 Corporate ownership networks.

Expanding on this concept, the use of common phones and addresses is one way to help identify inconsistencies in the underlying data and potential areas of fraud against the corporation. The diagram shown in Figure 7.11 presents an address with three related vendors, all with similar names. The number below the vendor name represents the vendor ID code assigned in the accounting system. To the system, there are three entirely separate vendors capable of doing business with this corporation. Most likely, this situation exists because the procurement staff did not take the time to see if an existing entity was already present in the data, or their search was unable to return an exact match. Not knowing the duplications that are present in the data can complicate accounting matters because there is never a true accountability of how much money has been spent with the vendor overall. One or more of the vendors could also be a front for special pet projects or kickbacks. Regardless, it represents circumstances that should be investigated.

A variation of this is shown in Figure 7.12 where the vendor ID appears in sequential order. What is interesting to note in this case is that the middle entry is the only valid value. One would have to question why this sequence was entered into the system, why the last entry made

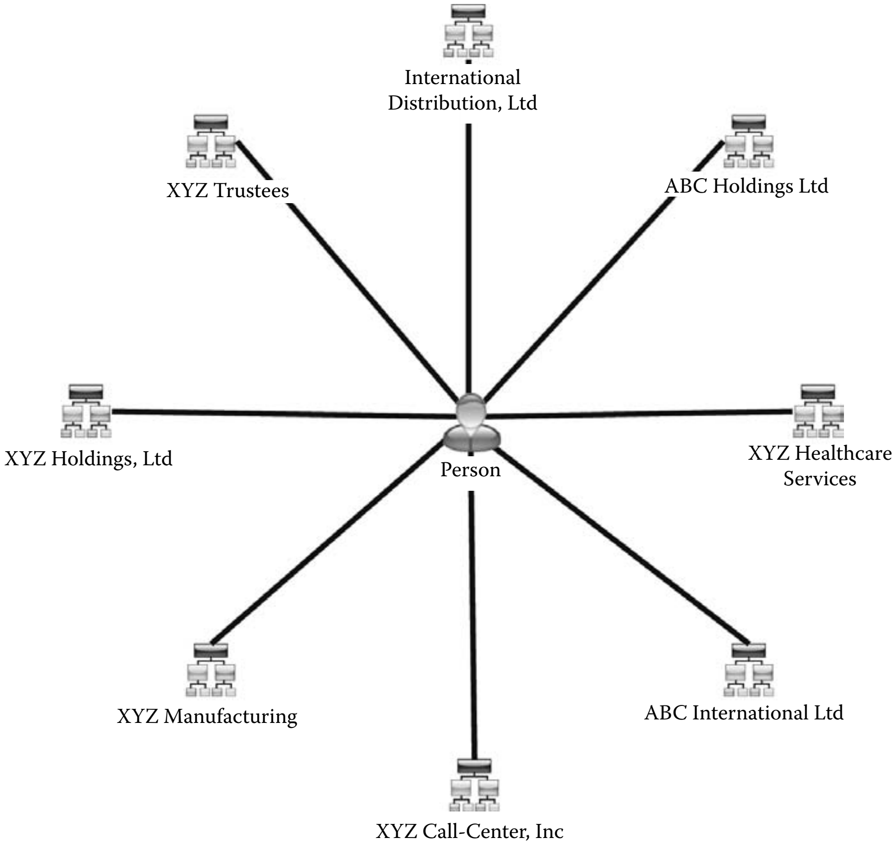


Figure 7.8 Individual with large corporate influences.

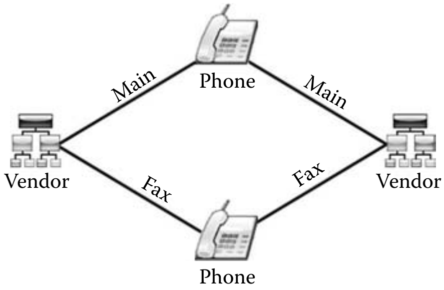


Figure 7.9 Common use of same corporate phone numbers.

is invalid, and if the two vendors are actually invalidated or if the system treats them as active vendor ID codes. Again, this situation needs to be brought to the attention of management and audits run against the system to determine if there are any improper procurements or

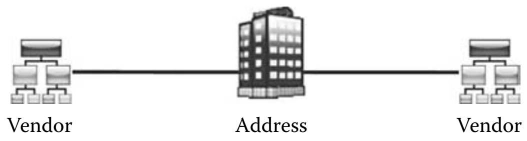


Figure 7.10 Common use of same corporate address.

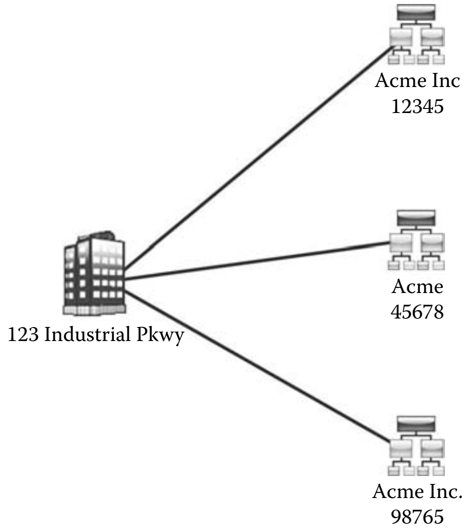


Figure 7.11 Multiple vendor ID codes assigned to the same vendor.

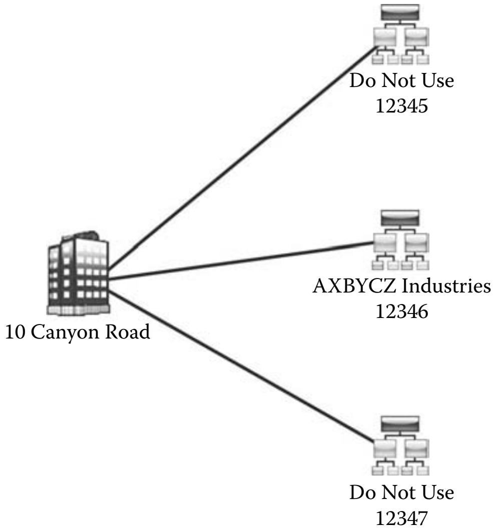


Figure 7.12 Improper vendor ID codes assigned to a single vendor.

invoices associated with these codes. Ideally, they should be deleted from the system so this condition does not represent a risk to the corporation.

To further exemplify how inconsistencies impact the overall reliability and integrity of procurement systems, very large networks can be produced from just the different variations present in the data. They do not always reflect any type of fraud, but often poor controls in the accounting and procurement systems that, if allowed to persist, make it harder to differentiate legitimate activity from fraud. A sample of the vendor information contained in an invoicing system for a property management firm is shown in Figure 7.13, depicting a popular company that offers bath and body gifts, fragrances, and skin care products. Even if the spelling of the company name is identical, each object represents a different client ID (e.g., vendor ID) showing a total of 11 different entries for one company, plus a number of variations on the payment address. Clearly, some better internal controls need to be employed by the management firm because this degree of repetition could easily lead to duplicate invoicing and other improper posts or ledger errors.

Corporate Expenses

This next round of examples is based on some fundamental processes common to all businesses large and small, from around the globe—namely, expense reimbursements. They are one of the necessary tribulations associated with doing business, especially when travel is involved. There are many ways in which to "embellish" an expense report, which is just another form of stealing from a company through padding costs, fabricating expenses, and bait-and-switch expenditures. Often it can help supplement the salary of a disgruntled employee and will tend to repeat itself over and over. It is also a nice loophole for earning money as a form of nontaxable income.

In one scenario,[40] large computer hardware manufacturing company was concerned that there were employees embezzling money through various loopholes in their expense-reporting systems. The security office had excellent physical measures to keep their equipment from being stolen from their facilities and warehouses, but they had little control of or insight into how the employees were expensing their travel costs. The company's main interest was in determining whether there were any

[40] Updated from the previous work described in: Westphal, Christopher and Blaxton, Teresa, *Data Mining Solutions: Methods and Tools for Solving Real World Problems*, (New York: John Wiley & Sons, 1998) 53.

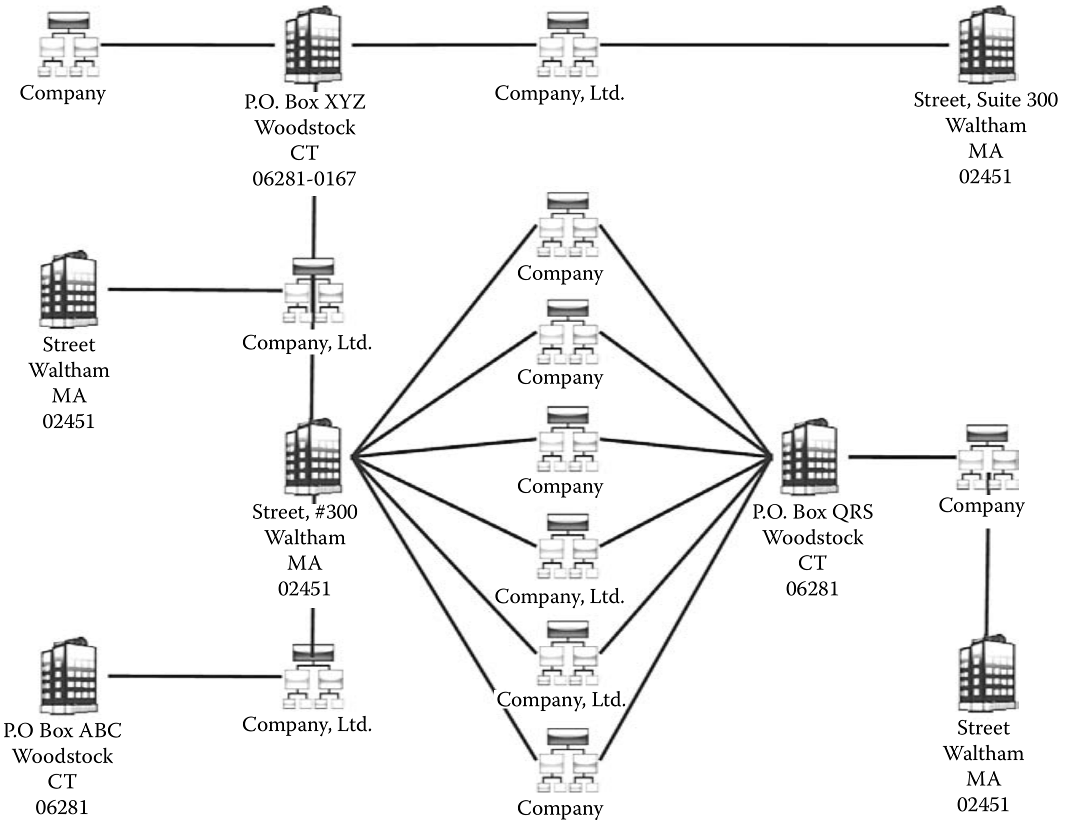


Figure 7.13 Large variations on a single company profile.

patterns of theft that could be detected without too much room for ambiguity with respect to the nature of the activity.

In this particular application the patterns were derived from two electronic data sources. The first was an online expense-reporting system developed for internal use, which contained employee reports of business expenses, out-of-pocket charges, and travel reimbursements. The second data source contained actual charges incurred on the company-owned credit cards (American Express) issued to company employees. Each data source was mined individually to detect intradomain patterns indicating personal use of the credit cards for purchases in liquor stores, home furnishings, and women's lingerie shops, and for questionable business expense reimbursements, such as large phone call charges, hotel services charged back to a room, or cash advances.

At the time, the company had strict policies on air travel bookings. They required that all air-related travel be handled through their appointed travel agency. The travel agency could find better travel rates, manage flight changes, and properly address all the related travel logistics. Naturally, one of the first queries made into the system was for all credit card purchases involving an airline carrier, as shown in Figure 7.14. Surprisingly, there were a significant number of airline ticket references that were identified.

In continuing with defining the pattern, the next logical step was to extract all travel expense reports with an associated airline ticket purchase. Most of the expense reports flagged looked fairly typical and included airfare, meals, lodging, and local transportation, such as cab fare or rental cars. A depiction of this is shown in Figure 7.15. At this point it was easy to spot those employees that traveled frequently, those who had high-dollar reports, and those who were more compliant with submitting their reports and properly breaking down each expense.

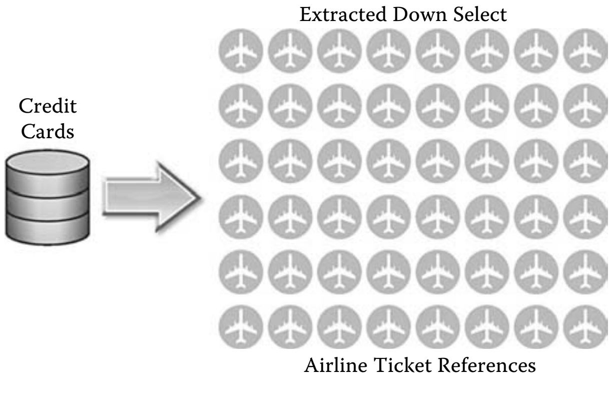


Figure 7.14 Credit card ticket references.

The online expense report and credit card data were correlated showing all reimbursements for air travel during similar time periods, as shown in Figure 7.16. In most cases there was a one-to-one correspondence of credit card charges and expense report reimbursements, exemplified by EMP #1, indicating that employees submitting airplane ticket charges to the online expense reimbursement system were automatically compensated for their charges. All that was required was a copy of the receipt for the ticket.

However, when the color/style of the credit card transaction was changed to reflect whether the charge was a credit or a debit, a whole new pattern emerged. For several particular employees (EMP #2 and EMP #3), it was apparent that they were buying full-fare, fully refundable airplane tickets on the corporate credit card (upward-facing airplane), submitting

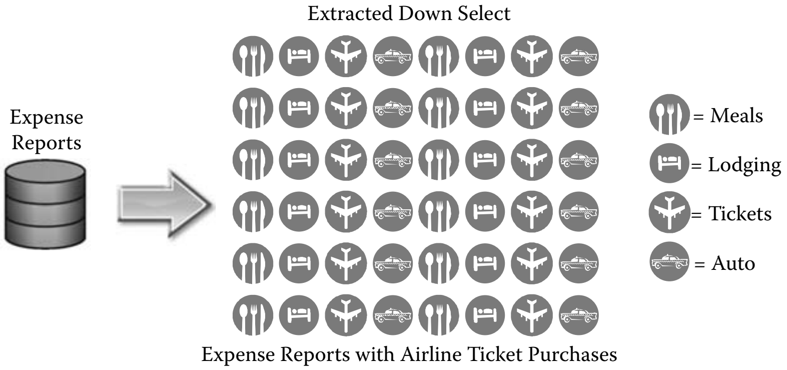


Figure 7.15 Expense reports with airline ticket purchases.

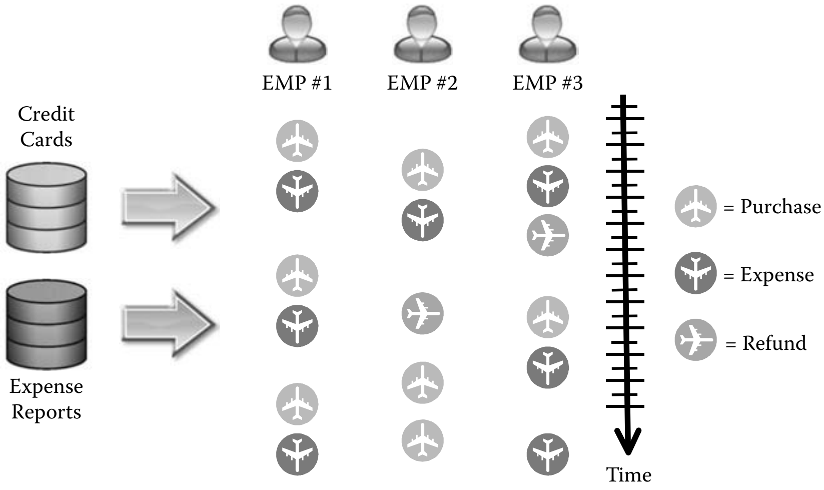


Figure 7.16 Combined data sources showing air travel expenses.

for a reimbursement through the online expense system (downward-facing airplane), and then returning the unused tickets to the airlines for a credit back on their charge accounts (sideways-facing airplane). The net result was that they were pocketing the cost of the ticket at the expense of the company. Some of these fares were quite expensive, especially a full-fare, round trip from the West Coast to the East Coast.

Another pattern that was observed occasionally showed a ticket expense without any correlating credit card debt, also shown by EMP #3 as two downward-facing airplane icons in a row, which implies the ticket was probably purchased on a personal credit card. This tends to be out-of-the norm because most employees don't want to float the cost of corporate travel expenses on their personal accounts. There could certainly be special circumstances causing this type of event to occur; however, it should be evaluated and reviewed by the internal accounting staff to ensure it is legitimate.

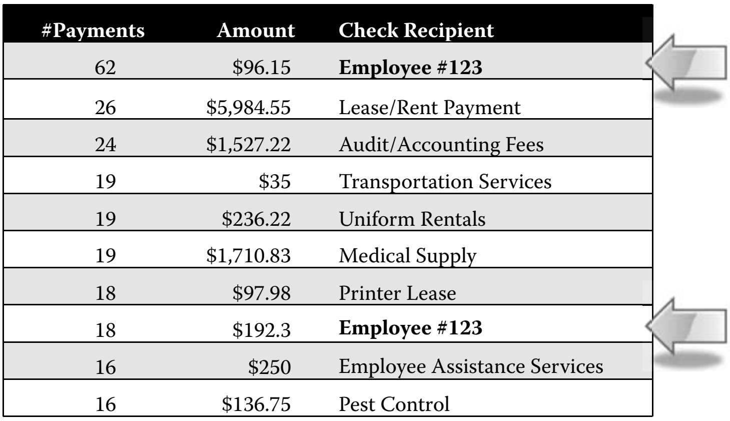
Duplicate Payments

There are many ways to detect fraud and some approaches are quite simple. Basic list sorting, accumulating values, and other combinations of data can help expose situations that should be reviewed and justified by auditors. Using traditional online analytical processing (OLAP) approaches is one of the quickest ways to get a breakdown of various data fields, such as payments, amounts, dates, and other content where multiple (repeating) instances of the values is considered questionable behavior. OLAP-like approaches are often used for understanding transactional behaviors, such as payment and invoicing frauds.

The results from this analysis are based on the payments made from a medical company over a period of one year. Figure 7.17 shows the top 10 payments, in terms of frequency ("# Payments"), in the system queried on the vendor master file. Each row in the results table presents the total number of payments made to a specific vendor for a specific dollar amount.

Upon closer inspection, there are a few items that stand out as "questionable" and require further evaluation. What is immediately revealed is that the top entry, Employee #123, has 64 checks[41] issued to her for the exact amount of $96.15. The nature of her business is unclear; however, basic breakdowns for monthly or weekly reimbursements do not correlate to any type of known payment frequency (e.g., monthly parking, Internet fees, meals, mileage reimbursement, lease or rental costs, etc.). In fact, when the checks are presented using a date grid shown in Figure 7.18, there are several items that appear problematic in terms of when some of these checks were issued.

[41] The Company records payments made by check to employees as vendor payments.



#Payments	Amount	Check Recipient
62	$96.15	**Employee #123**
26	$5,984.55	Lease/Rent Payment
24	$1,527.22	Audit/Accounting Fees
19	$35	Transportation Services
19	$236.22	Uniform Rentals
19	$1,710.83	Medical Supply
18	$97.98	Printer Lease
18	$192.3	**Employee #123**
16	$250	Employee Assistance Services
16	$136.75	Pest Control

Figure 7.17 Table showing payment by frequency.

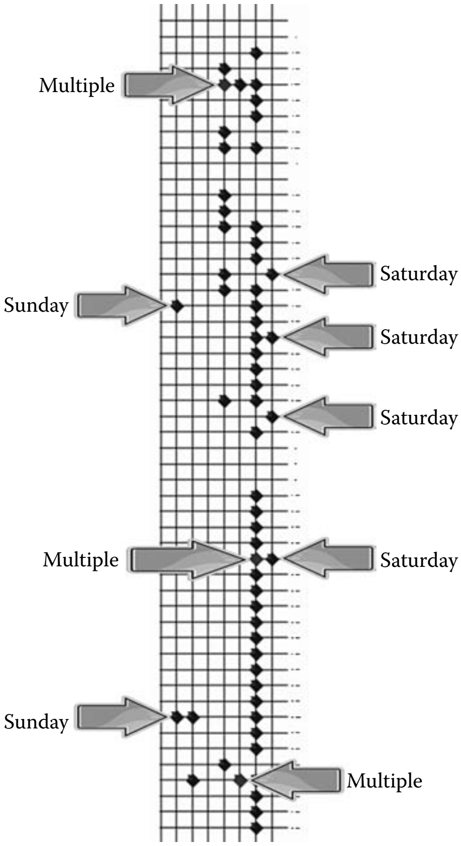


Figure 7.18 Date grid for the 64 check payments.

Quite clearly, the scheduled check-cutting day for this company appears to be Friday; however, there are also a number of Saturday and Sunday check issue dates, which raises flags as to why anyone from the accounting department was working over the weekend to issue checks. It would not be inconceivable to work over certain weekends to get a backlog of work cleared out; however, it represents a highly unusual situation that should be reviewed. Furthermore, there are at least three instances of multiple checks being cut on the same day, and as can be seen in the diagram, there are weeks when two and even three checks are issued to Employee #123, which begs the question of why these were not rolled up into a single check.

Looking further down the table presented in Figure 7.17, there is a second entry for Employee #123, showing 18 payments of $192.30, which is exactly double the $96.15 payment amount ($96.15 × 2 = $192.30). This can be considered an additional 36 payments of $96.15, which conceptually brings our total up to exactly 100 payments of $96.30. These payments are shown in Figure 7.19. Furthermore, there are no other payment amounts for this vendor in the data, only these specific amounts.

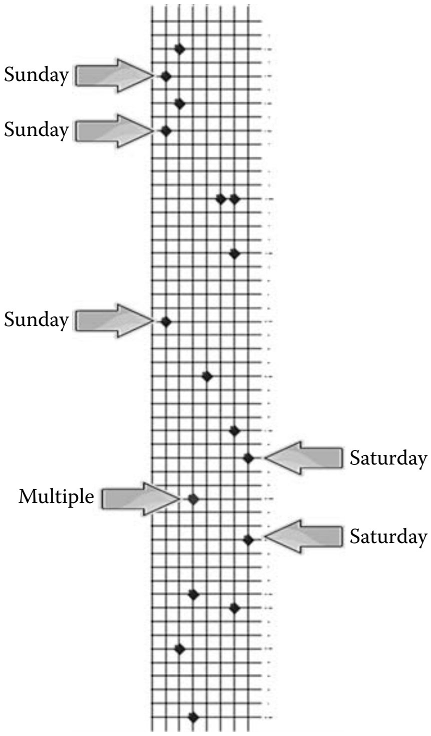


Figure 7.19 Date grid for the 18 check payments.

Why did the company not roll many of these into bundled payments? There is a lot of extra overhead and resources going into processing those checks each and every week. Figure 7.20 presents a final image with the two date grids combined where the $192.30 checks are shown marked with an arrow. This situation raises a lot of questions; however, it does not necessarily mean that any wrongdoing has occurred. As with all patterns the truth must be established and the internal auditors need to review the expense statements to see if they represent appropriate and legitimate cost expenditures for the company. Until this final step is performed, the pattern remains high value but unconfirmed.

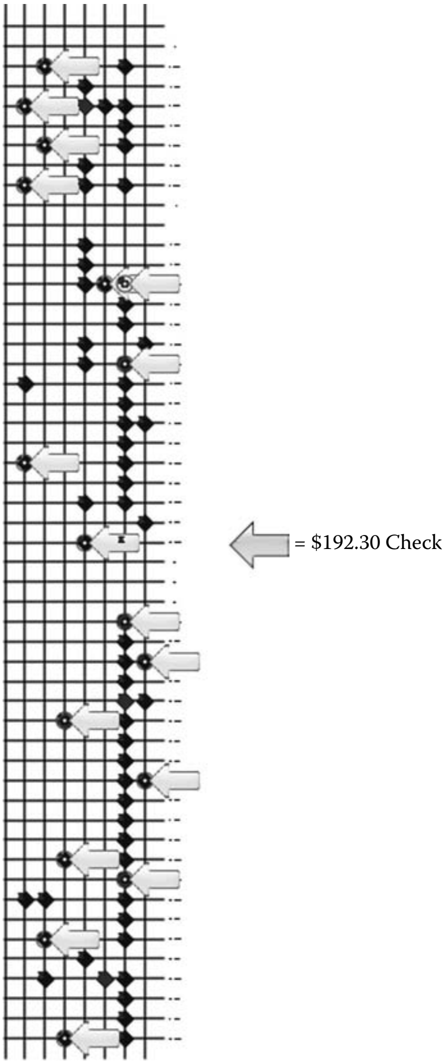
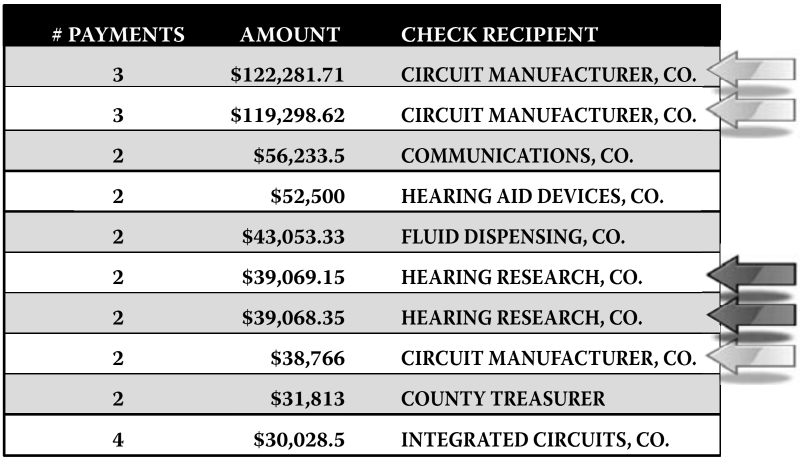


Figure 7.20 Date grid for the combined check payments.



# PAYMENTS	AMOUNT	CHECK RECIPIENT
3	$122,281.71	CIRCUIT MANUFACTURER, CO.
3	$119,298.62	CIRCUIT MANUFACTURER, CO.
2	$56,233.5	COMMUNICATIONS, CO.
2	$52,500	HEARING AID DEVICES, CO.
2	$43,053.33	FLUID DISPENSING, CO.
2	$39,069.15	HEARING RESEARCH, CO.
2	$39,068.35	HEARING RESEARCH, CO.
2	$38,766	CIRCUIT MANUFACTURER, CO.
2	$31,813	COUNTY TREASURER
4	$30,028.5	INTEGRATED CIRCUITS, CO.

Figure 7.21 Table showing payment by amount.

In expanding on these concepts, the results of the previous database query are now re-sorted using the Amount column, shown in Figure 7.21, to expose the highest payment amounts with multiple checks. The light arrows show a particular vendor receiving some of the largest duplicate payment amounts recorded. The concern here is to determine if the payments are part of a financing plan (e.g., equipment, construction) or if the payments potentially represent duplicate payments for the same invoices. It is somewhat unusual for a vendor to receive three payments for $122,281.71 and another three payments for $119,298.62. A check on the actual details shows that these checks were all cut and paid within three weeks of one another. The dark arrows show the payments for a different company where a similar type of pattern seems to exist, except with only two payments for each. Each of these vendors also has quite a number of additional payments for various amounts; however, these particular entries appear “questionable” and need to be further investigated.

When performing these types of reviews, there are also checks made to see if any of the amounts tend to be more “rounded” (whole numbers, no cents) and “clustered” around the same range. In this dataset, there are numerous payments (not shown) clustered around the $30,000 range, which might be an attempt to circumvent the signature levels for purchase authorizations by unbundling the costs into multiple payments. Also, there were several dozen payments made to specific vendors where the invoice amounts did not match the payment posted such that they were off by a considerable amount; in some cases, up to several thousands of dollars. The companies in question are large organizations (e.g., overnight delivery, travel agencies, temporary staffing) and therefore no fraud is probable, but rather there is more likely a flaw in the accounting software someplace. Additional review of these payment scenarios was initiated to determine the nature of these payments.

Human Resources

The cliché "good help is hard to find" applies throughout all levels of business. Periodically reviewing the indirect relationships among employees can help to spot trouble areas that may lead to future problems, especially when an employee is terminated. The indirect relationships can be established through e-mail networks, interoffice phone calls, and personal residences. In this next example, all of the employees of the organization were tracked in a human resources (HR) database and given a status indicating whether they were "active" (A), "terminated" (T), or "leave" (L), as presented in Figure 7.22.

There are a total of 191 employees represented in the database with 163 active employees, 127 terminated staff, and 1 out on leave (maternity leave). As it turns out, there are also six people listed in the database with both an "active" and "terminated" status code, which is logically impossible and indicates some type of data inconsistency in the HR database. For reference, these people are considered terminated. At this point, the data is expanded to show the home addresses associated with each employee, shown in Figure 7.23.

As expected, the majority of these networks represent a one-to-one relationship between an employee and their home address. The eight networks highlighted in the upper-left corner of this diagram contain larger numbers of entities, indicating there is some type of shared asset—either an employee with multiple addresses or an address with multiple employees, with the latter being more common. A closer look at these eight networks is displayed in Figure 7.24.

The first network (#1) shows that there are three employees, one active and two terminated, with the same last name living at the same address. From the information provided, the relationships between these persons are not known, insofar as whether they are siblings, cousins, or some combination of parents, spouses, and children. Regardless, this situation should concern HR representatives because depending on what the reasons[42] were for the terminations, it could directly impact the working attitude and ability of the active employee.

[42] If they were terminated for cause for stealing, tardiness, or incompetence, it would not reflect well on the reamaining employee. If they left to return to school, it would not be a material reason to be concerned.

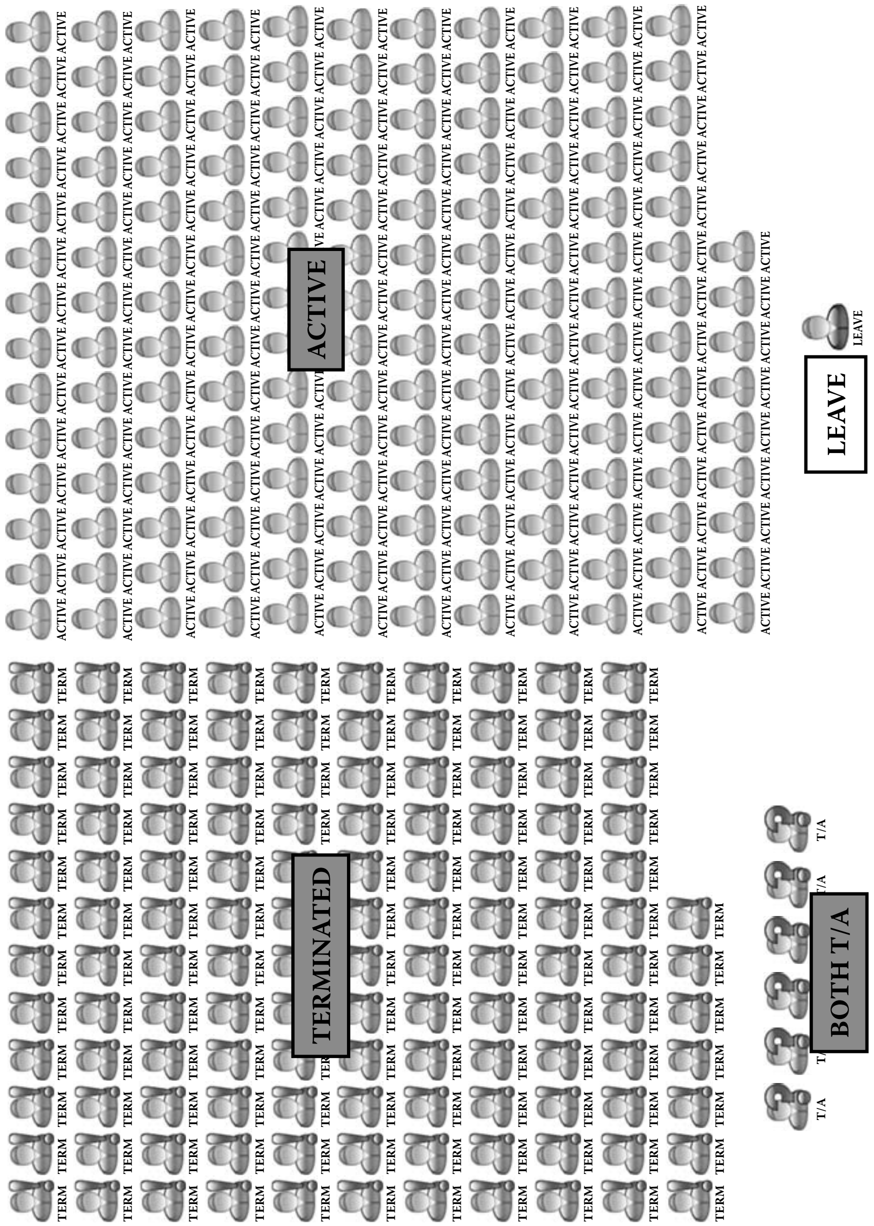


Figure 7.22 Employment status based on HR database.

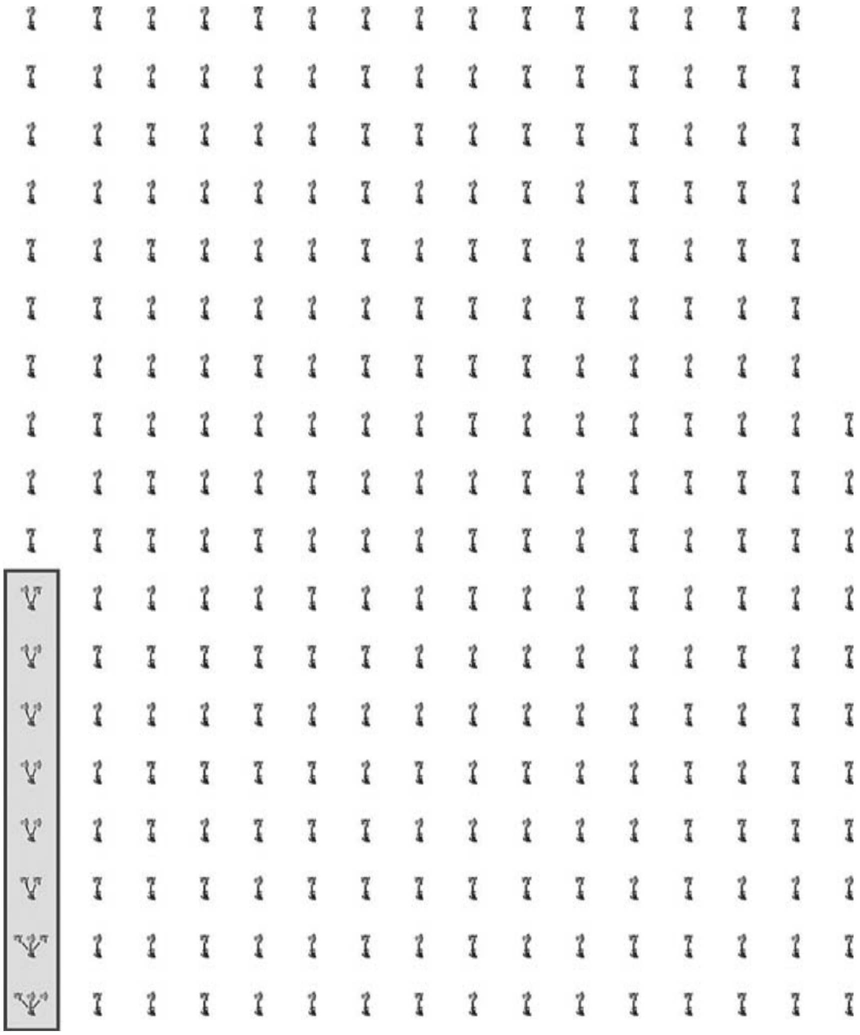


Figure 7.23 Network of employees' home addresses.

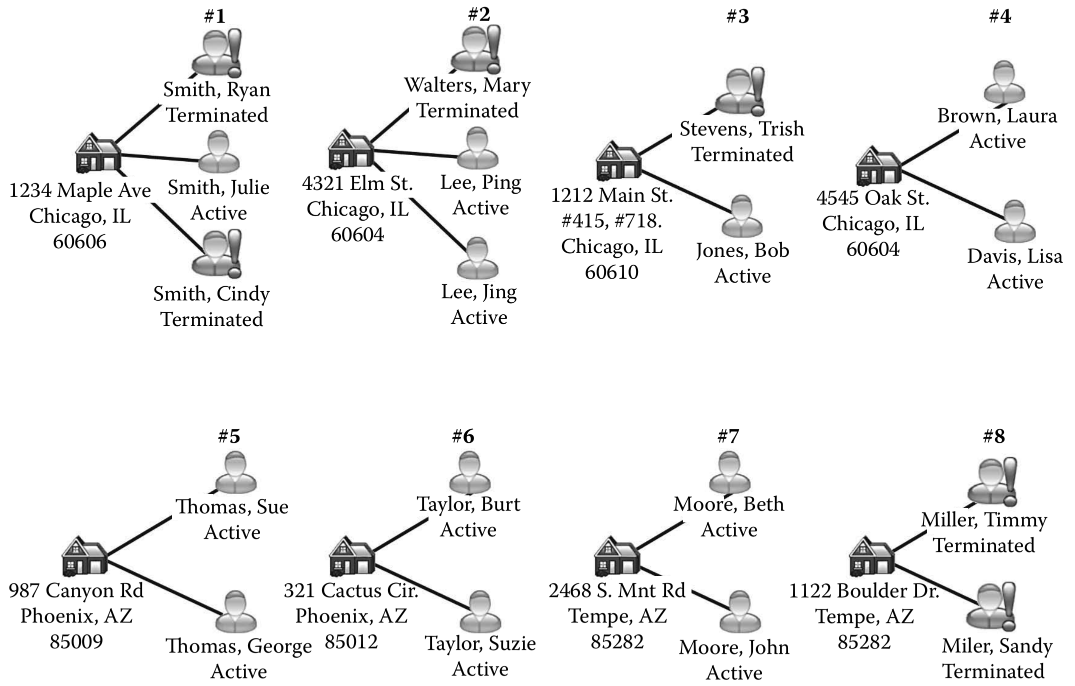


Figure 7.24 Many-to-one connections between addresses and employees.

Network #2 shows a terminated employee living with two active employees who share the same last name—perhaps husband and wife or brother and sister. Because the couple can be influenced negatively by the terminated employee, there should be some type of follow-up or review to see if there are any problems or questionable behaviors.

Network #3 is of less concern because the address depicted is really an apartment building and the apartment numbers show that these employees are living on different floors. Thus, even though this is a close match, it is not a direct relationship. The employees may have known each other and perhaps even carpooled into the office together, but the influence of the terminated employee is of less concern than if they had been living together. Networks #4, #5, #6, and #7 are virtually identical and are not of any importance at this time because both employees are active. Finally, the last network, #8, does not set off any alarms because both employees are terminated and no longer with the company.

Reviewing these types of networks can help gain insight into the relationships among employees and how changes, such as termination, can affect and impact other employees. Although this company had less than 200 employees, you can imagine the delicate networks and cross-relationships that exist in larger organizations. This is also where social network analysis (SNA) can come into play to help better understand advice, trusts, and influence networks within an organization. Although a detailed discussion of SNA is outside the scope of this book, there are a number of government agencies that use SNA approaches to help understand, prioritize, and apply confidence to the networks that exist within their data sources.

Gift Card Fraud

What do you get someone for a birthday, graduation, or holiday gift? Many people are now turning to giving gift cards because they allow the recipient to get exactly what they want. Many retail stores and restaurants offer their own brand of cards and virtually all credit card companies also offer gift cards. The cards are convenient to use and can be refunded or replaced if they are ever lost or stolen. The National Retail Federation estimated that consumers spent $26.3 billion on gift cards during the 2007 holiday season. As with any financial instrument that has a monetary value or worth, it is subject to various types of fraud[43] and scams.

[43] http://www.snopes.com/fraud/sales/giftcard.asp.

This next example provides a snapshot of a scenario that was analyzed when stolen credit cards were used to purchase gift cards from a national hardware retailer. Typically, a criminal will use a stolen credit card to purchase in-store gift cards, which in turn, can be easily resold for cash. Usually stolen credit cards are reported within a short amount of time and are shut down very quickly. Purchasing gift cards gives criminals more time to act because it can take days or weeks for authorities to track down the individual cards and deactivate them. According to one source,[44] "the group used the increasingly common tactic of using the bogus credit cards to purchase gift cards and then cashing them at Wal-Mart and Sam's Club stores. The group usually purchased $400 gift cards because when the gift cards were valued at $500 or more, they were required to go to customer service and show identification."

Depending on the type of gift card being compromised, there are different approaches and amounts that are used to commit the fraud. Overall the process involved is fairly straightforward. The template shown in Figure 7.25 depicts a stolen credit card on the left that is used to make a purchase (e.g., the transaction in the middle) at a specific store location, on a specific date, for a gift card of a specific amount. If there are multiple purchases made, then multiple transaction objects will appear in the network, which can be used to view the temporal behaviors of the perpetrators.

In the next example, the stolen credit card was used at the same store location to purchase five gift cards, worth $500 each, on the same day. The charges initially went through; however once the card was reported as stolen to the merchant, the gift cards were automatically shut down and voided from future use. Figure 7.26 provides a representation of this particular network. In this case the quick reaction of the store manager limited the losses incurred by the merchant. Obviously, these types of conditions (e.g., patterns) can easily be encoded into a series of rules or alerts or be used as the inputs to a predictive analysis system.

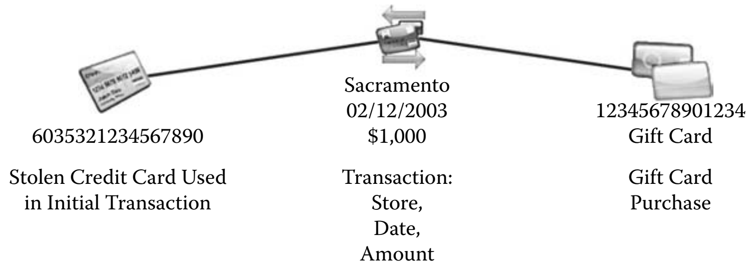


Figure 7.25 Gift card purchase using a stolen credit card.

[44] http://www.bestsecuritytips.com/news+article.storyid+205.htm.

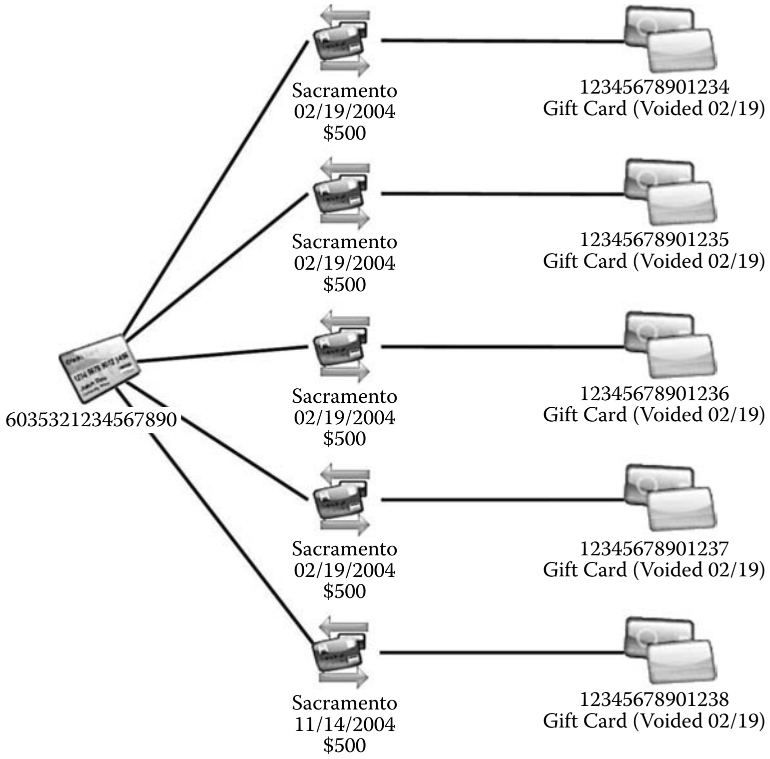


Figure 7.26 Multiple gift cards purchased at same merchant location.

The next example, shown in Figure 7.27, is virtually identical to the previous example, except that the stolen credit card was used at different stores throughout a region. In this case not all the stores identified the gift cards purchased with the stolen credit card and therefore not all gift cards were voided. Only after a broader analysis was done by the merchant was the card shown to connect purchases among the different stores. As a potential safeguard the merchant could consider enacting additional audit rules pertaining to the scope and scale of the gift cards purchased by a single credit card.

Based on the analysis performed for this merchant, the transactions associated with purchasing gift cards were arranged according to their amount, ranging from $25 up to $5,000. The distribution, shown in Figure 7.28, depicts both a circular and linear placement of the transactions, where those cards having the same face value appear as clusters in each format and the single instances represent unique dollar amounts. There are a few items to note in this figure, including that most cards purchased were based on “rounded” amounts such as $500, $1,000, and $2,000. There is also an anomaly where three gift cards were purchased at the same store for $1,072.85, which is a very specific and unusual amount.

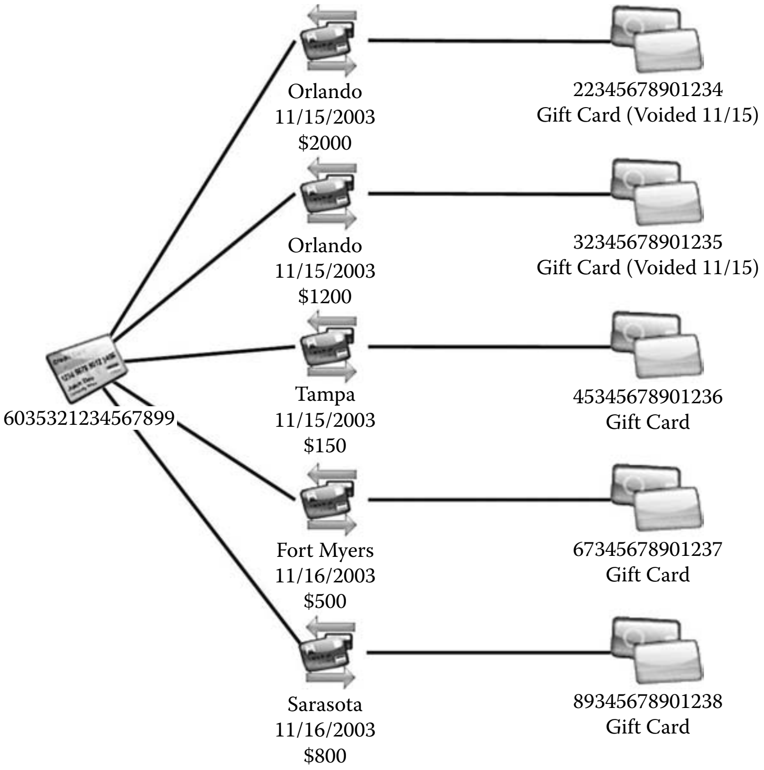


Figure 7.27 Multiple gift cards purchased at different merchant locations.

Taking a broader look at the gift card purchases shows that there are a variety of different network sizes and shapes, as presented in Figure 7.29. In this diagram, representing only a subset of the entire data, the transaction object is removed and the credit card is linked directly to the gift card. The direct relationship implies the stolen credit card was used to purchase the gift card. Most of the network structures are fairly common, where one or more gift cards were purchased.

Upon closer examination the two networks in the lower left show something slightly different. One of the gift cards purchased with the stolen credit card was subsequently used to purchase additional gift cards. Essentially they appear to be layering the transactions, making them harder to track. When reviewing larger samples of this merchant's dataset, the gift-card-to-gift-card purchases never went more than one level deep and tended to be only for one or two other gift cards.

Selecting one of these networks and expanding it to show all of its related transactions reveals a much more complicated network, as shown in Figure 7.30. In this diagram purchases made by the credit card (shown

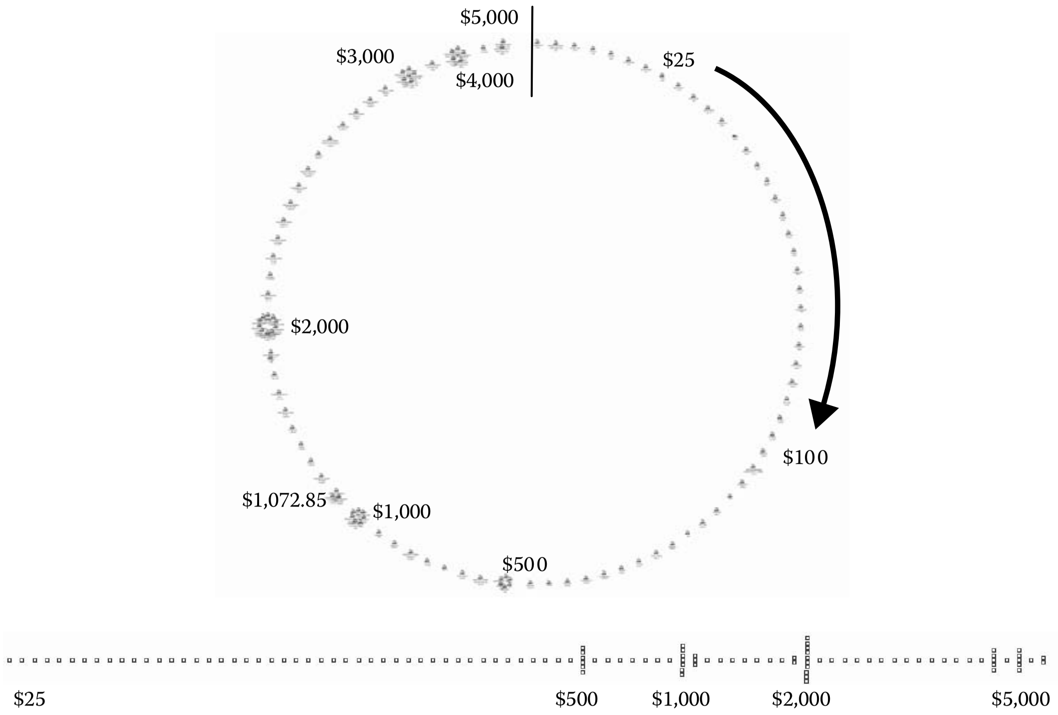


Figure 7.28 Distribution of gift card value.

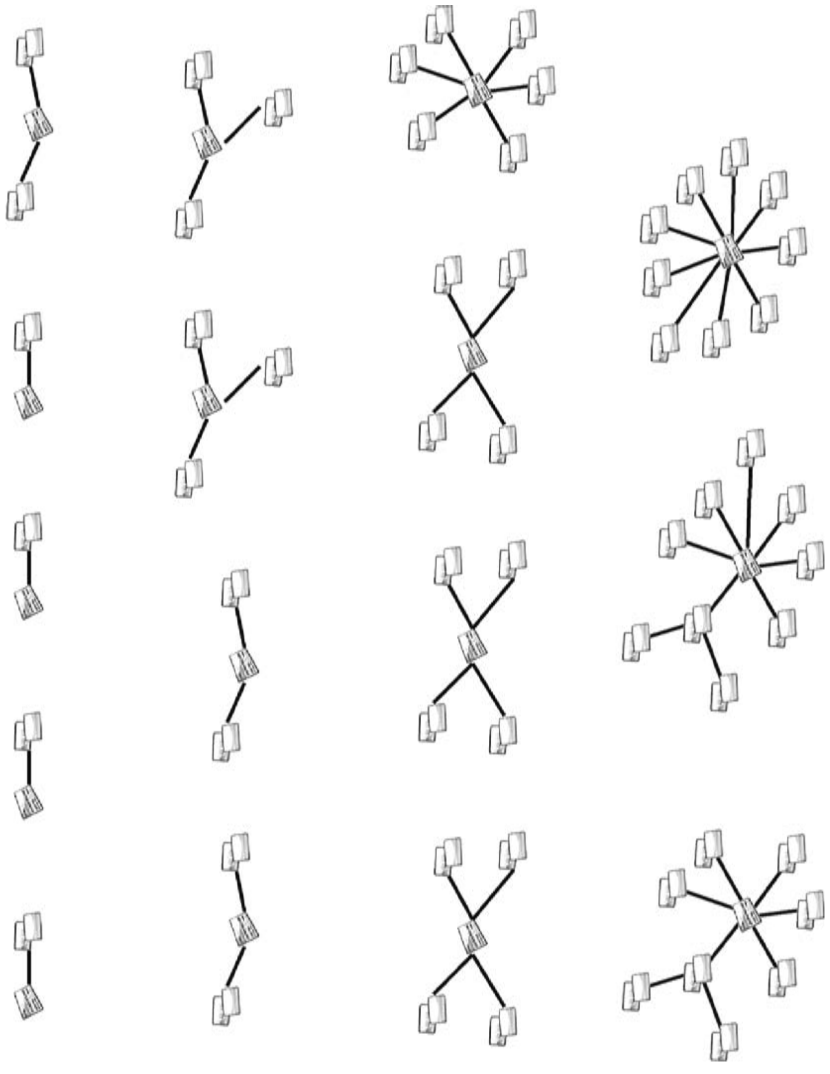


Figure 7.29 An overview of the network structure.

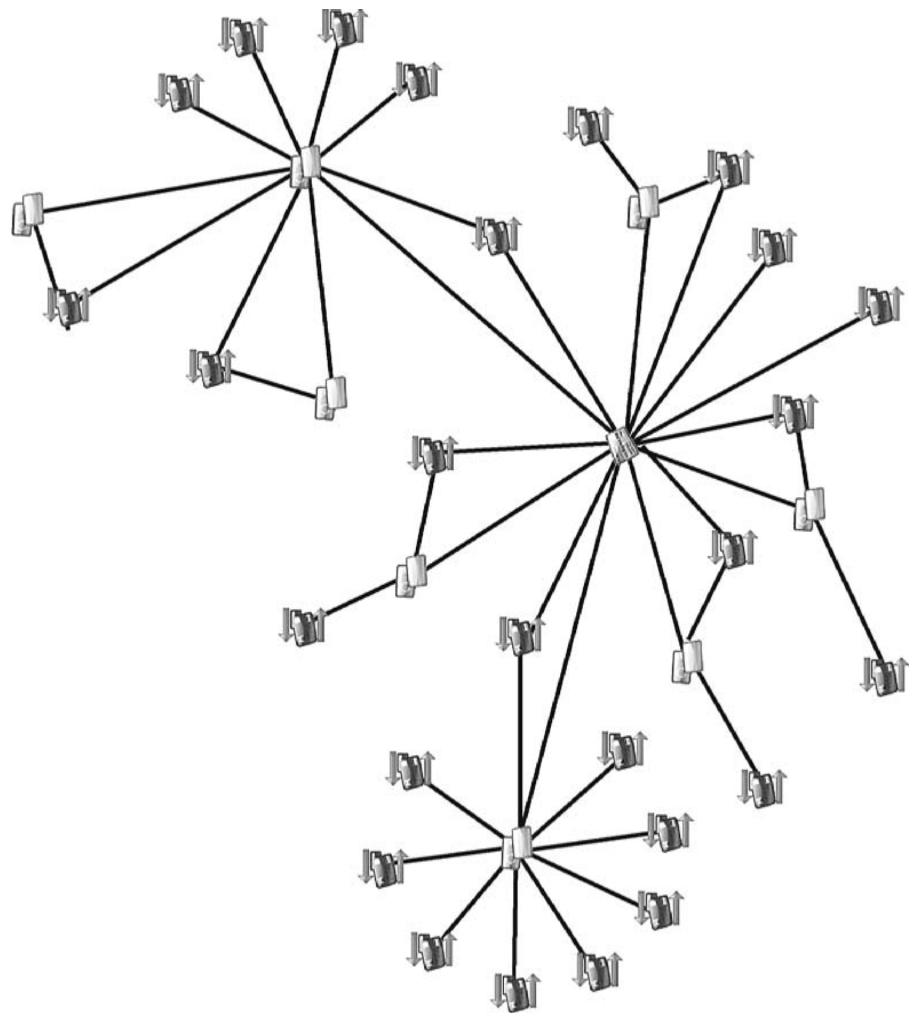


Figure 7.30 Full transaction details stemming from a single credit card.

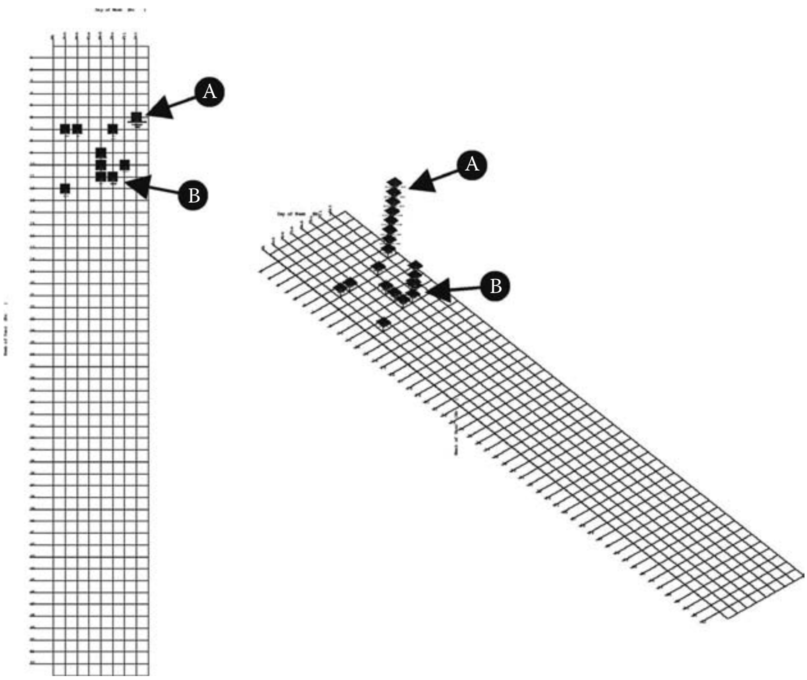


Figure 7.31 Date grid for card transactions.

near the center) or any of the gift cards are displayed as a transactional object and it becomes clear that the gift cards are being drawn down in value by additional purchases of merchandise; where some have only a few purchases, others have many purchases.

Taking advantage of the date information contained in the transactional objects, a date grid is generated to show the exact behavior in this situation (presented in Figure 7.31). These transactions all take place over a one-month period, in the March time frame. The activity starts on a Saturday with a large number of purchases, indicated by the (A) arrow, that continue into Sunday and Monday. There is then a three-week period where transactions occur on Wednesdays and, just before finishing, the perpetrators executed multiple transaction on another day, Thursday, indicated by the (B) arrow.

Additional Examples

There are too many different industries and fraud scenarios to cover in this section; however, understanding the parameters, boundaries, and relationships goes a long way in uncovering patterns and trends. Generally, the types of patterns found in one domain can typically be abstracted and used in other domains because the structure of the patterns is often similar in terms of connections, frequencies, and sequences of activities. The following quick examples simply show the concept of employing visualizations to better understand novel patterns.

Pharmaceutical

The medical and healthcare industries are prime candidates for data analysis and visualization tools. There are a myriad of practical applications of such tools in these industries. The ability to access and analyze data related to illness, injury or disease occurrence, frequency, and prognosis allows for accurate tracking and cause resolution of outbreaks. Pattern analysis can also contribute to the medical community's ability to prepare for and respond to uncommon illness, injury, or disease occurrences. Additionally, thorough data analysis can help uncover fraud, both by and through a medical practitioner. Data analysis exposes such fraudulent situations as unbundling, upcoding, pharmacy fraud, and use of ghost patients. Figure 7.32 shows an example of how prescription utilization and pharmacy compliance can be reviewed using visualization techniques.[45]

Notice that in the area above the patient icon there are clusters of similar prescriptions that are filled multiple times for Zofran,[46] sodium chloride, Dexamethasone,[47] and Kytril.[48] The label of the claim shows us the medication, pharmacy, date filled, cost, and number of days supplied. This individual is most likely being treated for cancer and related symptoms. This approach can help spot pricing anomalies, issues with refilling prescriptions before the prescribed supplies are used up, and other types of anomalies that may cause concern.

Phishing/Click Fraud

Another example comes from cyberspace, where fraud and deception are commonplace in a number of online resources. Everyone is familiar

[45] Several prescription drug names were referenced in the dataset and several represent registered trademarks including the following. Duragesic® (Ortho-McNeil) and Dilaudin® (Abbot Laboratories).

[46] Zofran® (GlaxoSmithKline) is an ondansetron that is used to prevent nausea and vomiting assoicated with chemotherapy and radiation.

[47] Dexamethasone (Decadron® Merck & Co.) is a class of drugs also referred to as steroids typically used to help reduce swelling.

[48] Kytril® (Roche Pharmaceuticals) is another medication used to control nausea and vomiting from chemotheray and radiation treatments.

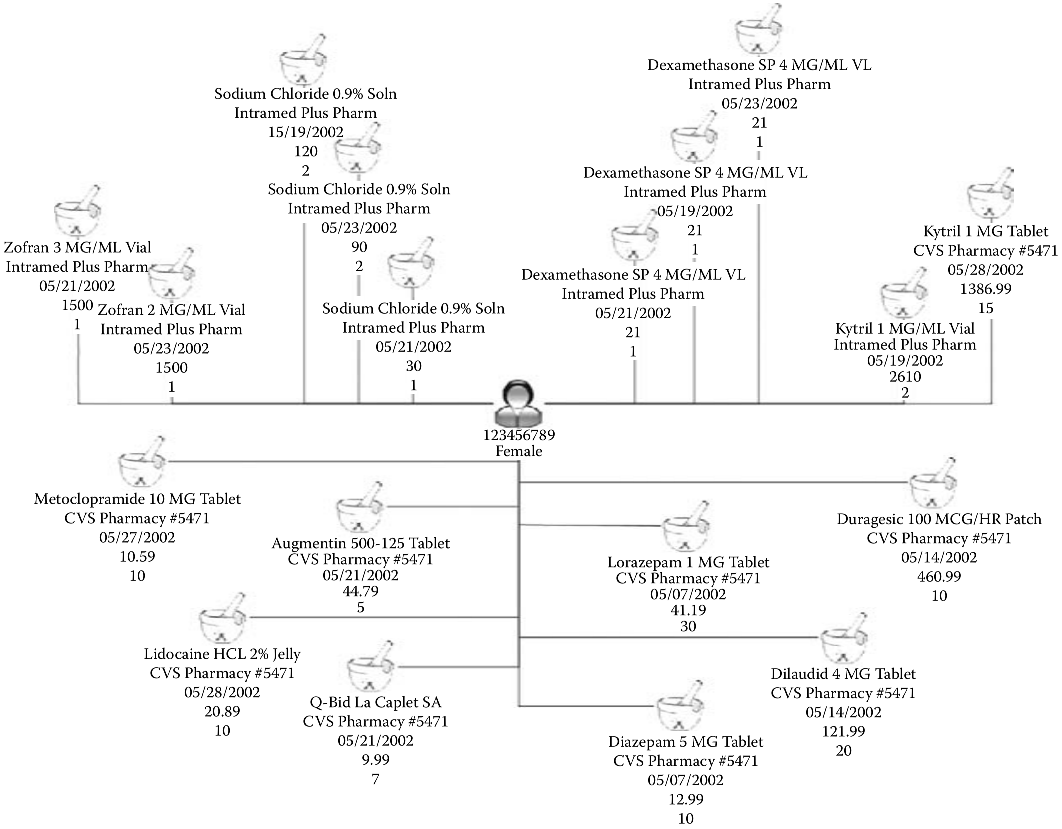


Figure 7.32 Prescription claims filed for patient.

with the spam e-mails they receive for every type of male enhancement pill, insider stock pick, and winning lottery scheme. A lot of spam also comes in the form of phishing where a legitimate-looking e-mail from a bank, a retailer, or some other industry tries to acquire personal information under false pretenses stating that one's account will be suspended or closed. Webopedia[49] defines *phishing* as:

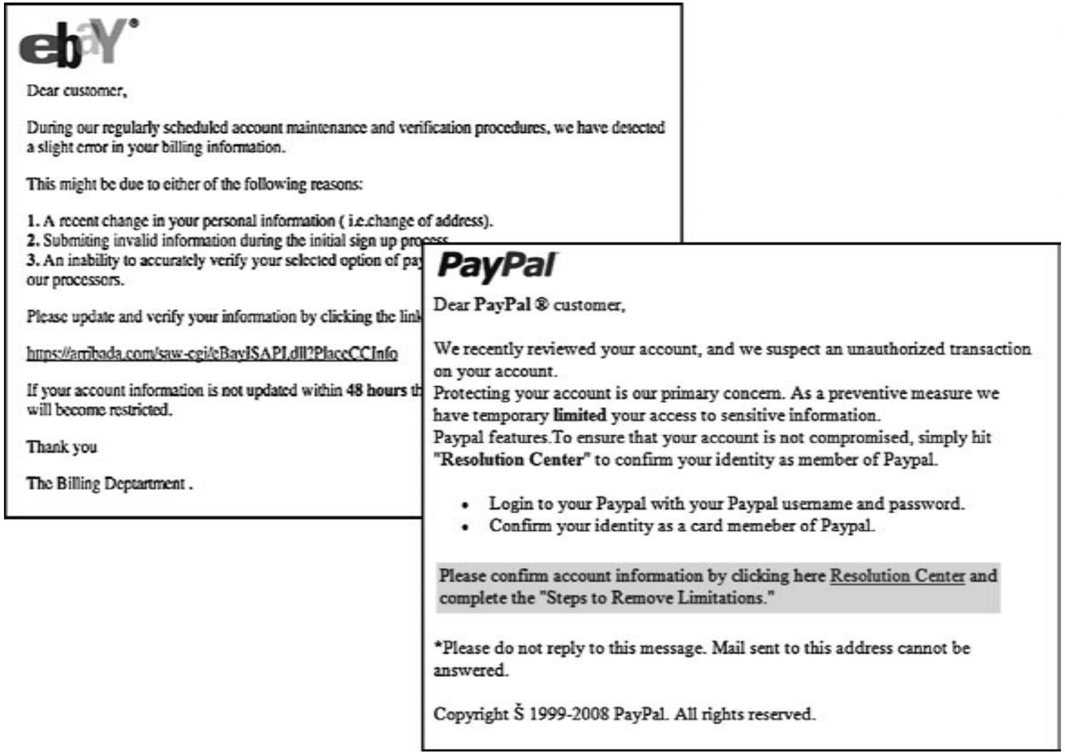
> *(fish´ing) (n.) The act of sending an e-mail to a user falsely claiming to be an established legitimate enterprise in an attempt to scam the user into surrendering private information that will be used for identity theft. The e-mail directs the user to visit a Web site where they are asked to update personal information, such as passwords and credit card, social security, and bank account numbers, that the legitimate organization already has. The Web site, however, is bogus and set up only to steal the user's information.*

The most common phishing attacks have come from sites posing as eBay or PayPal. Figure 7.33 shows an example of a false eBay and PayPal phishing e-mail. Notice that the embedded URL for the eBay message does not actually point back to eBay, but rather some farce site with a registrant in France. Many times the e-mail will mask the URL so it appears as a legitimate address until it is selected.

Using the log files of a Web server and some domain name service (DNS) lookup utilities, some insight can be gained to develop an approach to detecting the originating IP addresses from which phishing attacks are launched. The goal is to cause as much disruption as possible to perpetrators of phishing attacks through a process that may also assist with the prosecution of the offenders. One approach to tracking phishing attempts is to reference each URL to a traced host name and IP address and isolate the host IP or host name for each URL. This can then be related to show a Host, Page, and Visitor as shown in Figure 7.34.

As all the data is pulled into visualization, a larger-scale network starts to form, as shown in Figure 7.35. From here, Visitor commonalities can be exposed as well as more active Pages and Hosts. Interpretation of the results can vary depending on the tasks at hand. These approaches are not too dissimilar to click-fraud techniques where the behavior, origin, and frequency of the clicks can be interpreted and classified into questionable activities.

[49] http://www.webopedia.com/DidYouKnow/Internet/2005/phishing.asp.



Dear customer,

During our regularly scheduled account maintenance and verification procedures, we have detected a slight error in your billing information.

This might be due to either of the following reasons:

1. A recent change in your personal information (i.e.change of address).
2. Submiting invalid information during the initial sign up process
3. An inability to accurately verify your selected option of pa
our processors.

Please update and verify your information by clicking the link

https://arribada.com/saw-cgi/eBayISAPI.dll?PlaceCCInfo

If your account information is not updated within **48 hours** th
will become restricted.

Thank you

The Billing Department .

Dear **PayPal** ® customer,

We recently reviewed your account, and we suspect an unauthorized transaction on your account.
Protecting your account is our primary concern. As a preventive measure we have temporary **limited** your access to sensitive information.
Paypal features.To ensure that your account is not compromised, simply hit "**Resolution Center**" to confirm your identity as member of Paypal.

- Login to your Paypal with your Paypal username and password.
- Confirm your identity as a card memeber of Paypal.

Please confirm account information by clicking here Resolution Center and complete the "Steps to Remove Limitations."

*Please do not reply to this message. Mail sent to this address cannot be answered.

Copyright Š 1999-2008 PayPal. All rights reserved.

Figure 7.33 Sample phishing email.

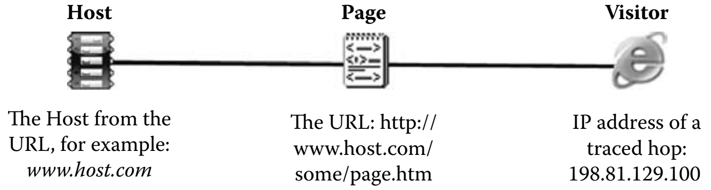

Host
Page
Visitor
The Host from the URL, for example: *www.host.com*
The URL: http://www.host.com/some/page.htm
IP address of a traced hop: 198.81.129.100

Figure 7.34 Sample phishing representation.

Tax Evasion

Finally, tax evasion is a type of fraud against the government and its taxpayers where individuals and corporation try to structure their earnings and losses in a way as to maximize their savings. Unfortunately, many resort to blatant misrepresentations, undervalued reporting, and other fabricated values and figures to justify their tax returns. Every year in the United States, over 230 million tax returns are filed[50] with the IRS. For corporate returns, it is important that the IRS identifies abusive schemes and illegal offshore tax shelters. Many of the companies of concern fall into one of three categories, a 1065 (partnership income), a 1041(estates and trusts), or an 1120S (S corporations). The following provides more detail for each:

1. Form 1065—Partnership Income[51]
   a. Form 1065 is an information return used to report the income, deductions, gains, losses, etc. from the operations of a partnership. A partnership does not pay tax on its income but "passes through" any profits or losses to its partners. Partners must include partnership items on their tax returns. A partnership is the relationship between two or more persons who join to carry on a trade or business, with each person contributing money, property, labor, or skill and each expecting to share in the profits and losses of the business whether or not a formal partnership agreement is made.
2. Form 1041—Estates and Trusts[52]
   a. The fiduciary of a domestic decedent's estate, trust, or bankruptcy estate uses Form 1041 to report:
      i. The income, deductions, gains, losses, etc. of the estate or trust.
      ii. The income that is either accumulated or held for future distribution or distributed currently to the beneficiaries.

[50] http://www.irs.gov/pub/irs-soi/12proj.pdf.
[51] http://www.irs.gov/pub/irs-pdf/f1065.pdf.
[52] http://www.irs.gov/pub/irs-pdf/f1041.pdf.

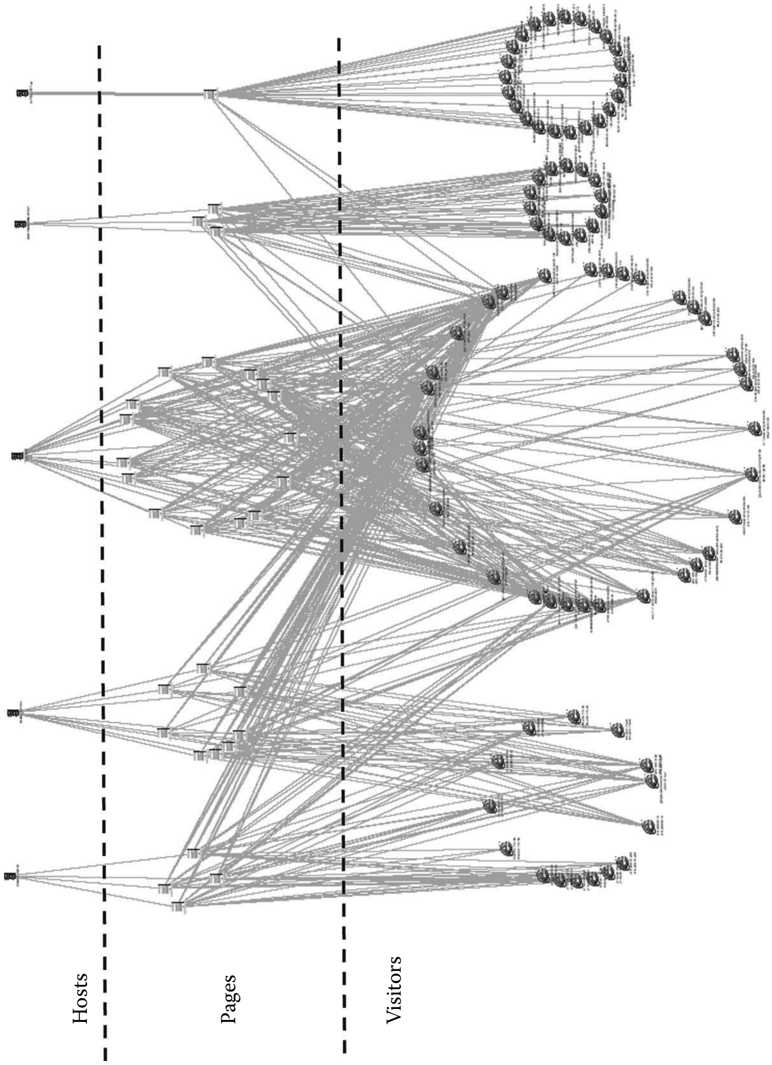


Figure 7.35 Larger phishing network/click-fraud network.

  - iii. Any income tax liability of the estate or trust.
  - iv. Employment taxes on wages paid to household employees.
- b. Abusive Trust Arrangements—Certain trust arrangements purport to reduce or eliminate federal taxes in ways that are not permitted under the law.

3. Form 1120S Corporation[53]
   - a. Form 1120S is used to report the income, gains, losses, deductions, credits, etc., of a domestic corporation or other entity for any tax year covered by an election to be an S corporation. Generally, an S corporation is exempt from federal income tax other than tax on certain capital gains and passive income. On their tax returns, the S corporation's shareholders include their share of the corporation's separately stated items of income, deduction, loss, and credit, and their share of nonseparately stated income or loss.

The type of information collected on each form varies somewhat, but each collects standard information, such as name, address, Employment Identification Number (EIN), year of filing, and backup information regarding deductions, income, payments, dividends, etc. Naturally, companies are owned by other companies or individuals and the income and the tax liabilities can pass through to these other taxpayers. Therefore, networks of connections among these companies exist and the monies can be tracked to see where the profits are skimmed off and the losses pass through to the owners—resulting in a nice write-off. The shareholders of these companies receive a Schedule K-1 form, which defines the specific income, deductions, credits, and other items.

Figure 7.36 shows an example of a fairly simple and well-bounded network where an 1120S-declared company makes three distributions to its core owners, shown as SSN icons. Schedule K-1 of Form 1120S is used to report each shareholder's prorated share of net income or loss from an S corporation. In this case, the corporate ownership can be understood because two of the distributions are the same at $25,000 payout while the third is doubled at $50,000 (ordinary business income in this case). The arrow heads indicate the flow of the money.

While this network is fairly basic, they can get more complex and much more seasoned analysts are required to fully understand the dynamics and interactions among the companies. Large, intertwining corporate networks, such as Enron, are almost impossible to fully understand because of all the layering, numbers, and sheer volume of filings. In Figure 7.37 a slightly

[53] http://www.irs.gov/pub/irs-pdf/f1120s.pdf.

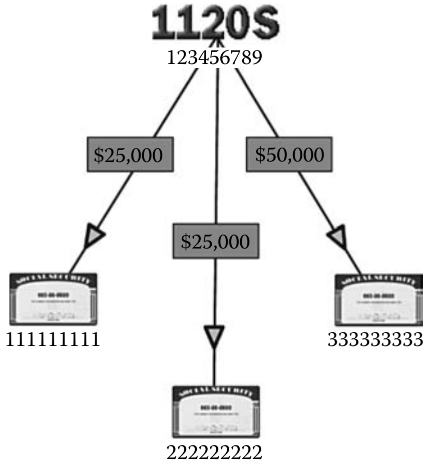


Figure 7.36 Example of 1120S K1 distributions.

more complex distribution network is depicted where the two 1065 companies located at the top of the diagram have generated substantial earnings reported as ordinary corporate income. The total, $62 million ($28 million + $34 million), flows into the middle 1065 company where somehow it is converted into two identical losses of $31 million, each of which flows down into the individual owners, but also manages to lose an additional $38 million, which passes into a flow-through entity (FTE) for distribution to the same two shareholders. All- in-all, a $64 million profit was somehow turned into a $100 million loss by this particular enterprise or collection of partnerships.

The schemes identified and encountered in these datasets are virtually endless, and require astute analysts who fully understand the tax codes and know why certain combinations of values, and how the entities relate together, are important. Remember, there are no right answers and no wrong answers, only situations that appear questionable and require further evaluation and review before passing a financial decision. In this case, the result would be to open a tax case against the perpetrators of the scheme/fraud; other times, it might be to try and identify more details of a criminal enterprise. Ultimately it is up to the investigating agency to determine how to respond to what has been identified.

Medicare Claim Fraud

Some approaches to identifying fraud and other questionable patterns contained in transactional data sets use the entity uniqueness as a ratio of the related transactions. Depending on the circumstances, the interest can focus on high ratios or low ratios. For example, when reviewing claims

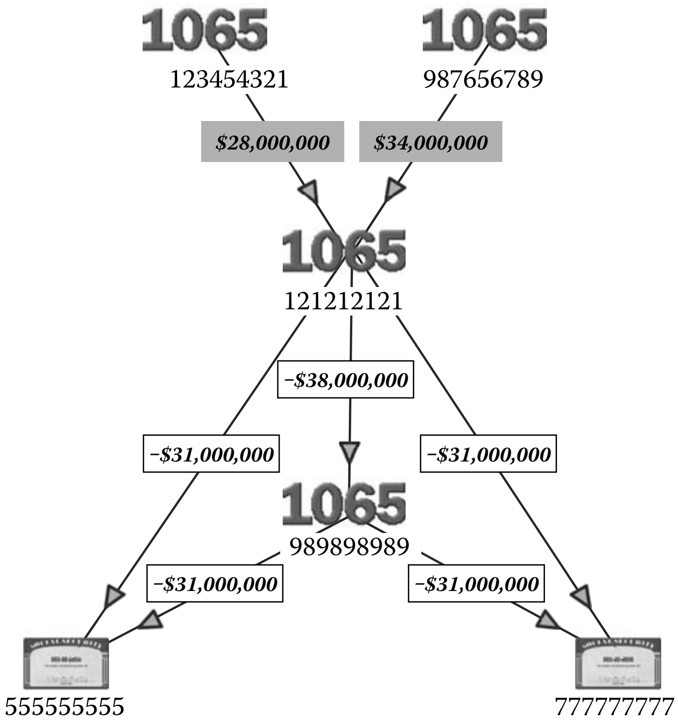
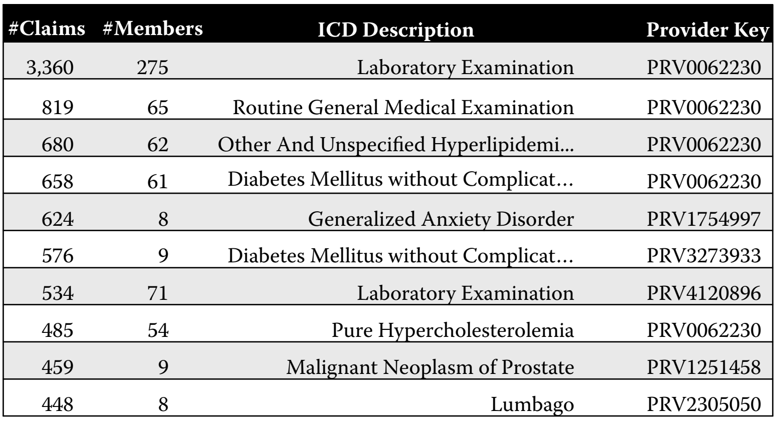


Figure 7.37 Example of 1065 K1 distributions.

submitted by medical practitioners to insurance companies, Medicare, or other governmental services, there is an interest in looking at high-ratio filers as a factor of the number of claims submitted versus the number of patients served. The justification for this review is that prescribing more procedures per patient helps drive up the overall cost of the services performed. Therefore, instead of pushing more patients (high volume) through the system, the goal is to increase the unit cost of a smaller number of patients. There are reported incidents where the number of claims per patient has exceeded hundreds.

The following shows a sample of medical data focusing on the ratio of the number of claims provided for each participating member based on the International Classification of Diseases (ICD) description (see boxed text) submitted by the provider. The table in Figure 7.38 shows four columns corresponding to each of the variables used in the query (claims, members, ICD, and provider). For each unique combination of the provider number and the ICD code, a count was performed to summarize the total number of claims and the total number of unique members who received that particular service from the provider. The table is sorted by the number of claims made, from highest to lowest. This data sample shows that the top provider submitted 3,360 claims for 275 different members (e.g., the insured party) for basic "laboratory examination," which represents a modest ratio of 12:1.



#Claims	#Members	ICD Description	Provider Key
3,360	275	Laboratory Examination	PRV0062230
819	65	Routine General Medical Examination	PRV0062230
680	62	Other And Unspecified Hyperlipidemi...	PRV0062230
658	61	Diabetes Mellitus without Complicat...	PRV0062230
624	8	Generalized Anxiety Disorder	PRV1754997
576	9	Diabetes Mellitus without Complicat...	PRV3273933
534	71	Laboratory Examination	PRV4120896
485	54	Pure Hypercholesterolemia	PRV0062230
459	9	Malignant Neoplasm of Prostate	PRV1251458
448	8	Lumbago	PRV2305050

Figure 7.38 Top claim and member counts for ICD and provider.

ICD Codes

The International Statistical Classification of Diseases and Related Health Problems, commonly referred to as ICD (International Classification of Diseases), are international standard codes (up to six characters) used to classify diseases, symptoms, and other health problems. Classification of diseases, more specifically, causes of death, originally started back in the 18th century[54] and steadily evolved over the next 150 years until a Frenchman named Jacques Bertillon (1851–1922) was credited for establishing one of the first international standards for uniformly classifying the causes of death. Generally, these classifications became more refined, improved, and consistent as more governments and health organizations adopted their use and started to standardize their reporting needs. Eventually, around 1945, it was decided that the causes of morbidity and mortality were closely related to the classification of sickness and injury and the reporting codes were updated to include diagnostic terms as well. This was also when the United Nations was formed and discussions were had about creating a World Health Organization (WHO), which was chartered in 1948 and given the responsibility for overseeing, revising, and supporting the list, which became known as the ICD. Over the years, the list has been refined and updated; the United States is currently using the ICD-9 standard, published by the WHO in 1977, which became the standard for reporting Medicare- and Medicaid-related services. Most other countries have since adopted the ICD-10 standard, which was completed in 1992. ICD-11 is currently under development and is expected to be implemented by 2013.

[54] History of the Development of the ICD, (http://www.who.int/classifications/icd/en/).

Rows 2, 3, and 4 also have similar ratios (12:1, 10:1, and 10:1, respectively). However, row 5, with 624 claims filed on eight members for ICD code "Generalized Anxiety Disorder" (ICD #30002), has a 78:1 ratio, warranting further investigation into why there are so many claims being filed for this group of patients by a single provider. The specific information is pulled from the underlying database and presented using a date grid as shown in Figure 7.39. The diagram shows a separate grid for each patient (e.g., unique member #) and is arranged to show the day of the week on the *x*-axis and the week of the year on the *y*-axis.

There appears to be one overwhelming pattern for all eight patients—they receive treatment literally every week from this provider and their treatment schedule is very regular. For example, patient P1 tends to prefer Friday and Saturday visits, while patient P2 prefers Sunday and Monday sessions, P6 has visits on Saturday and Sunday, and patient P7 is treated on Wednesdays and Thursdays. One slightly misleading fact about this display is that each grid shows treatments over a four-year period and the apparent "double" visits per week are actually across multiple years. This is clarified in Figure 7.40 where only claims pertaining to patient P1 are presented and the grids are grouped according to the year.

Interpreting this diagram is fairly straightforward. In late 2002, the patient (e.g., member) started seeing this particular provider for Friday appointments. This continued throughout all of 2003, except for a few weeks where service was not rendered (i.e., a claim was not submitted). In 2004, the regular day of therapy changed to Saturday and remained consistent until the end of the year. The services were apparently then discontinued, with only an occasional visit or two in 2005. Generally, one could argue that the nature of these claims, their frequency, and their temporal pattern would be consistent with behavioral health claims made by a psychiatric doctor providing psychotherapy sessions (e.g., 45 to 50 minutes) for his or her patients.

A final check can be done on the temporal patterns associated with the cumulative number of claims made by the provider. Figure 7.41 presents the same data presented in Figure 7.39; however, this diagram is grouped according to the year the service was rendered. The majority of the weeks show that the provider worked five of seven days, with an occasional six-day workweek. In 2003, the nonworking days are shown as Monday and Tuesday and, in 2004, they are Tuesday and Wednesday. With little variation around the summer months, there appears to be no time taken off for a holiday or vacation. Thus, one might ask if all of these claims truly represent services rendered.

Returning to the original result set presented in Figure 7.38, the next item on the list shows 576 claims for nine members receiving

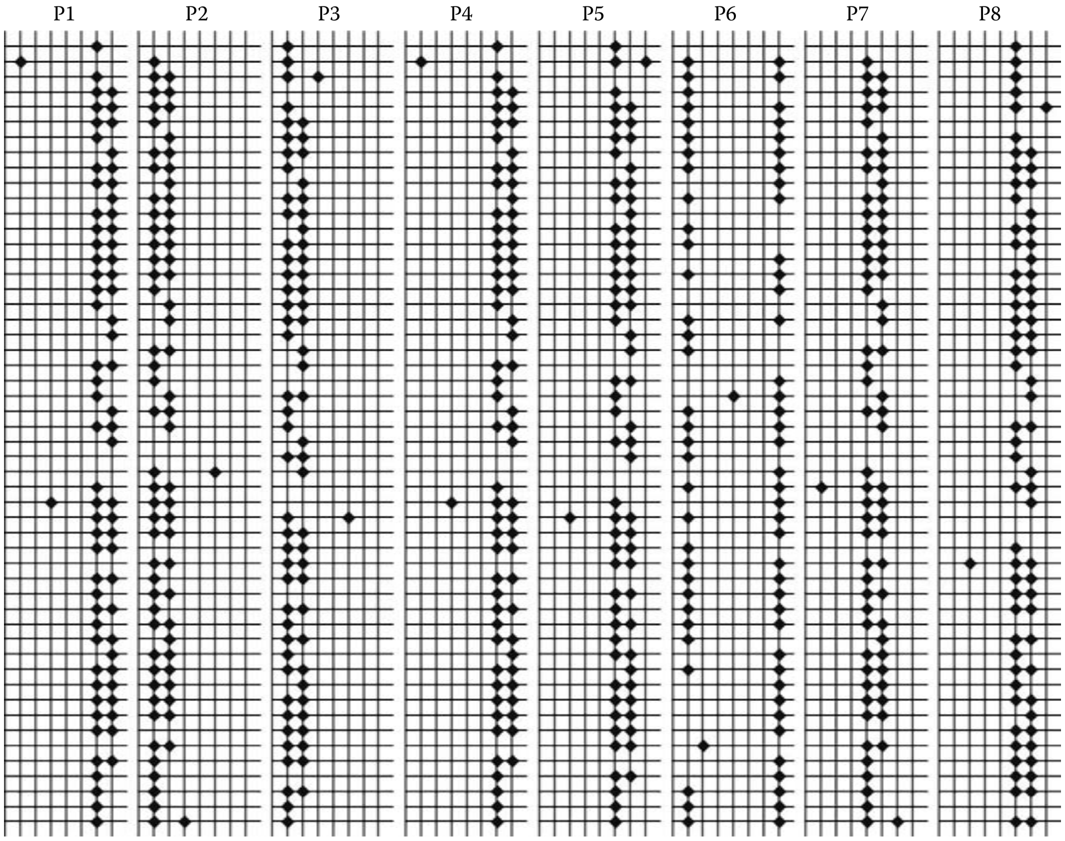


Figure 7.39 Date grid for ICD code = 30002.

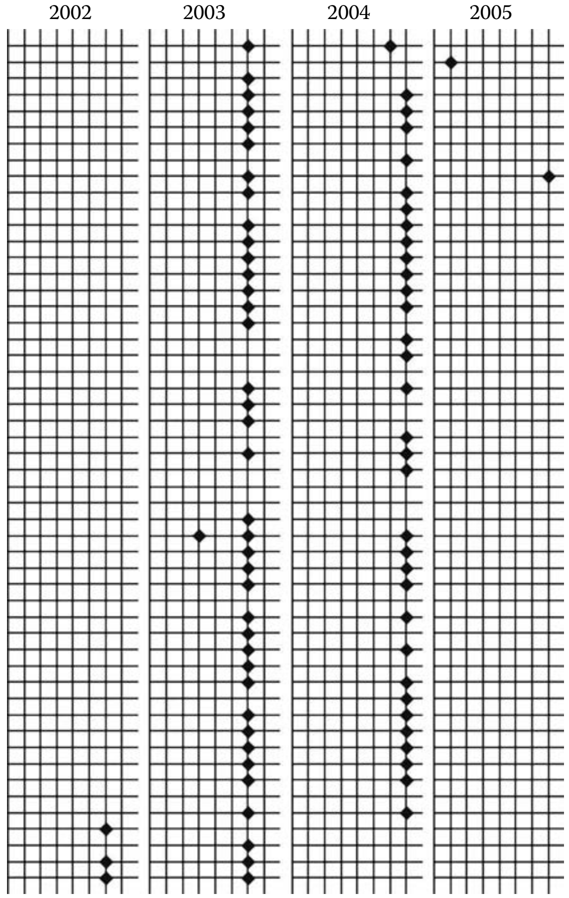


Figure 7.40 Claims for patient P1 grouped by claim year.

treatment for "Diabetes Mellitus without Complication Type I" (ICD #25001). This represents a 64:1 ratio for this provider.[55] Diabetes is certainly a more involved process of diagnosis and treatment, and, therefore, it could be expected that the claim ratio might be elevated. However, this must be verified by looking at the patterns contained in the database. Figure 7.42 shows all of the claim detail presented as a temporal display grouped according to the nine patients (e.g., members) identified by the summarization.

A very distinct pattern is exposed and is repeated for each group: an initial claim is made; exactly four weeks later another claim is filed; followed by another claim three weeks and one day later; then two more claims about two weeks later; followed by another claim made a week

[55] This same ICD was reported by another provider with only an 11:1 ratio.

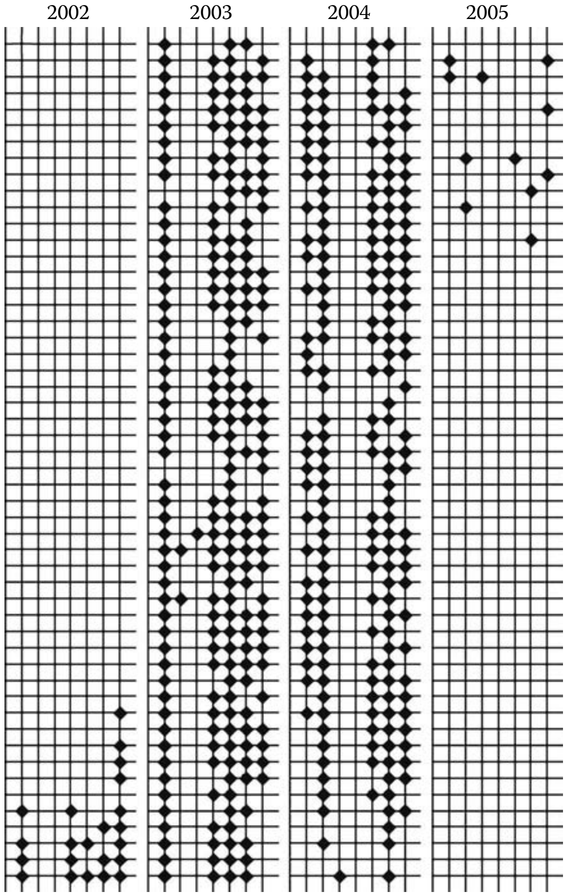


Figure 7.41 Temporal display for ICD code = 30002 grouped by year.

later; eight weeks beyond, another claim is made; and finally, six to seven weeks later, a last claim is made. Using a slightly modified and condensed layout with a manual overlay, the pattern becomes very explicit, as presented in Figure 7.43. Obviously, the treatment regimen prescribed by this provider (a physician of internal medicine) is very consistent.

Upon further evaluation of the diagrams, each patient (pattern) has eight unique claim dates shown in the temporal display. However, there is a 64:1 ratio for this data, meaning that each date must represent multiple claims. Drilling down on the diagram confirms this fact. Figure 7.44 shows the same data from a rotated perspective, where each column corresponds in height to the number of claims submitted. Each date supports multiple claims because every visit, test, and procedure is submitted as a separate Current Procedural Terminology (CPT) code, a convention defined by the American Medical Association to describe medical, surgical, or

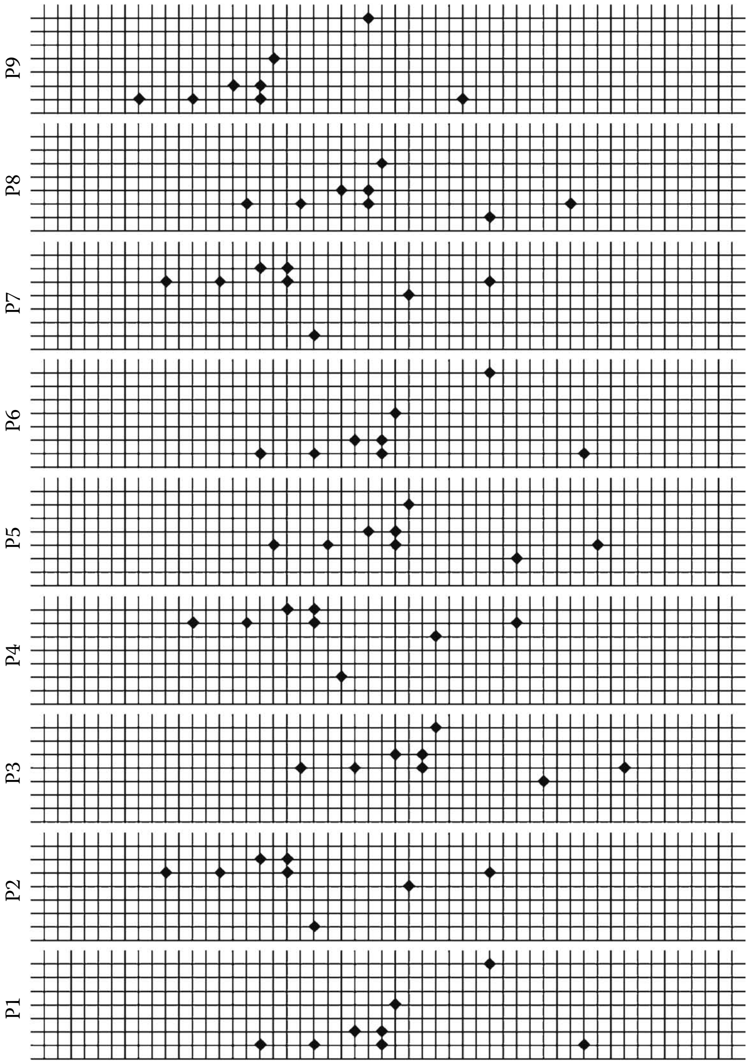


Figure 7.42 Temporal display for ICD code = 25001.

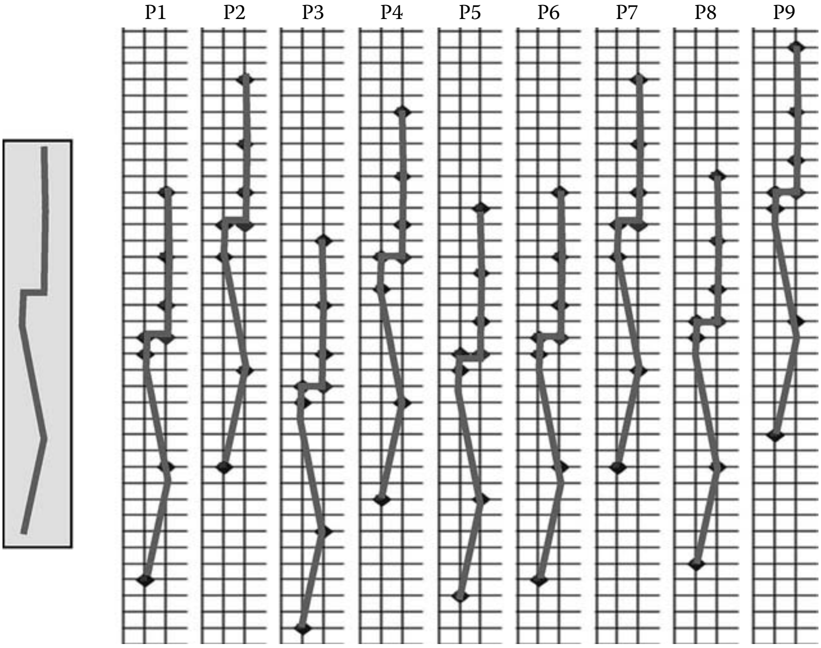


Figure 7.43 Pattern emphasized for ICD code = 25001.

diagnostic services. For example, the following are some codes reported for one of the claim dates: 99213 = Office visit, 83036 = Hemoglobin; glycosylated, and 81001 = Urinalysis.

Finally, a single patient (P#1) is extracted and temporally regrouped on the claim year. The actual pattern exposed for this physician for the ICD code becomes more explicit, as shown in Figure 7.45, and forces a reinterpretation of the original sequence of claims previously defined. The initial diagnosis occurs (1:11) with 11 claims; a week later a follow-up is conducted (2:3) with three more claims; 14 weeks later an additional office visit occurs (3:8) with eight claims; the following year the fourth visit occurs (4:8) with the same identical eight claims made on the previous visit; the fifth claim occurs exactly four weeks later and represents just a single office visit (5:1); and the sixth (6:8), a few weeks later, is again a repeat of the same eight claims performed previously; two weeks later the seventh (7:17) and largest number of claims submitted, at 17, occurs; and finally, after two more months the last claim (8:8) repeats the same eight claims made previously.

The importance of this level of detail allows the investigator to see what is actually occurring and to determine if the process of treatment

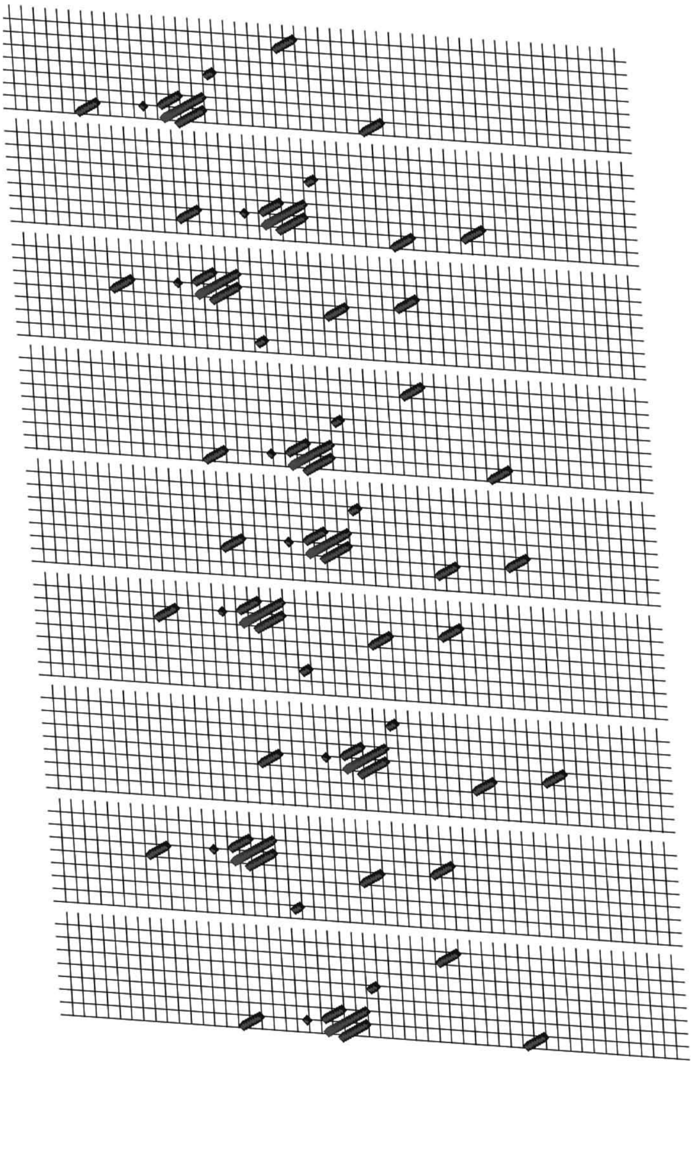


Figure 7.44 Temporal display showing multiple claims.

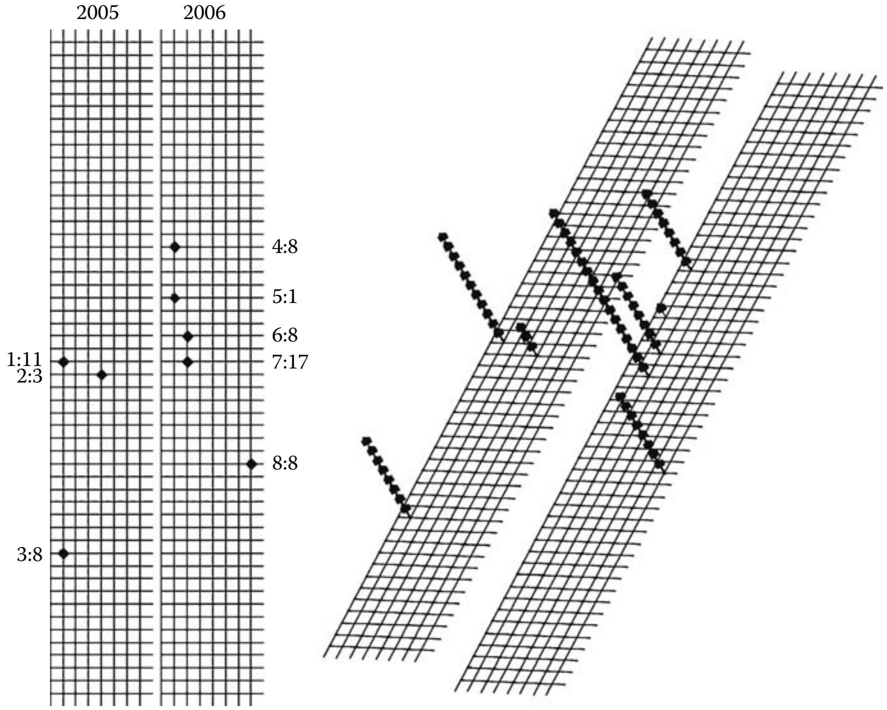


Figure 7.45 Multiyear claim Review for P#1.

is within normal operating parameters, if something needs to be flagged for additional review, or if there is obviously an error present. This can lead the insurer to deny or reduce payments for claims that are questionable or miscategorized,[56] and occasionally, albeit infrequently, it results in the provider receiving additional payments. More important, it provides a method to quickly review a large quantity of data and form an opinion without an overly complicated or extensively computational process, providing a means to adapt, refine, and update a knowledge base of patterns.

This type of high-ratio, transaction-to-entity pattern not only applies to medical claims review, but can also be used to describe activities involving credit card frauds. In the case of those kinds of fraud, if the number of entities (e.g., unauthorized or stolen credit cards) has a high ratio compared to the total number of overall transactions conducted on a merchant account, it would be an indication that the behavior was suspect. This type of situation is common for the bust-out scheme patterns previously discussed. For example, if a merchant conducted, say, 100 transactions

[56] "Appeal That Claim: Be informed, Be Approved," *American Medical Association*, 2007, http://www.amassm.org/ama1/pub/upload/mm/368/appeal-that-claim.pdf.

on 90 unique card numbers in a single day, it would be considered normal business activity. However, if those 90 card numbers comprised, say, 50 invalid numbers (e.g., stolen, counterfeit, unauthorized use, etc.), then the 90:50 (almost 2:1) ratio for this merchant would be of serious concern to the banks underwriting those accounts. For this pattern, the ratio is much lower than the medical billing example; however, the same overall process is invoked to expose questionable behaviors. Obviously, the use of ratios is only a single dimension or viewpoint on the data that can often result in identifying qualified targets of interest.

To expand on the medical discussions, once a “target” entity (e.g., a provider) has been identified for, say, improper billing, code bundling, misclassifying diagnoses, or some other error, further checks can be performed on related entities. If errors were found for one member of a medical practice, there is reason to believe that similar types of “inconsistencies” might be found among the other members of the practice. In the data set there is relevant contact information for the provider, including phone numbers, faxes, and correspondence addresses. Figure 7.46 shows the direct relationships between the provider, an address and a fax number. The thickness of the links is indicative of the 576 claims submitted because each consistently listed the same address and fax number. In reality there are 772 links because this provider also submitted other claims for different ICD codes not covered in the immediate investigation.

Expanding the network reveals that there are additional linkages contained in the database, as shown in Figure 7.47. For this example there are three additional providers connected to the same address as our target entity and no further connections stem from the fax number. Two of the providers shown in this diagram are ophthalmologists and have few claims submitted and are therefore of little “value” to the

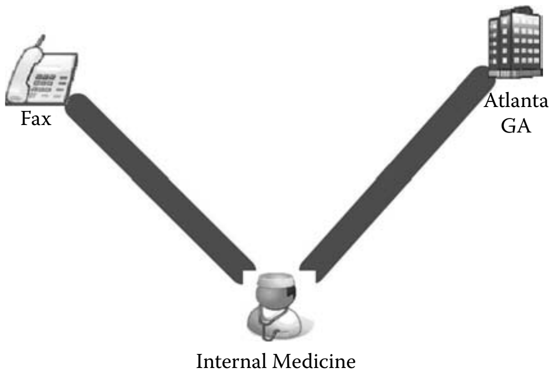


Figure 7.46 Direct connections to provider showing address and fax.

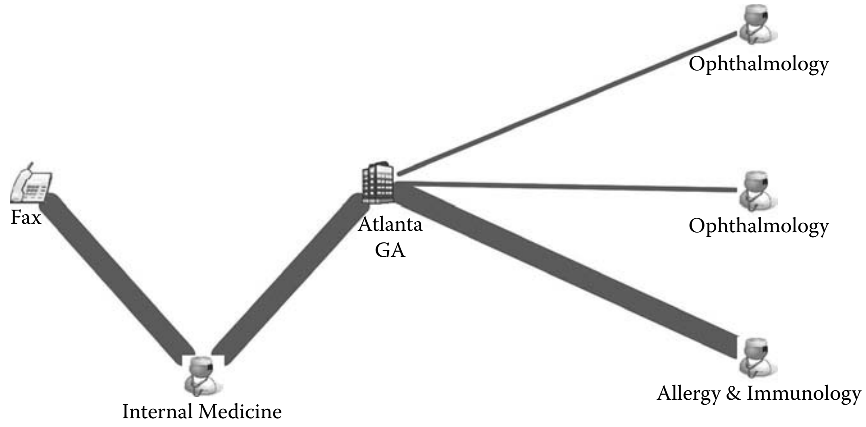


Figure 7.47 Indirect connections to provider via an address.

investigator. The third provider deals with allergies and immunology, and based on the link thickness, has a considerable number of claims submitted. Thus, a drill-down of this provider's claim detail could be performed to determine if everything appeared legitimate and aboveboard. The network can be expanded for as long as the data supports connections and the investigator feels continued analysis is warranted.

One final point to make on this type of analysis is that virtually any dimension contained in the data can be used to help expose anomalies. Many times, some type of metadata extraction or referential source can be used to add value to the core analytical data. Using the address of the providers and the members is one such dimension where distance calculations can be made using the centroid of a ZIP code, or more accurate, street-level geo-coding can be performed. This allows investigators to target providers based on their geographical proximity to the member addresses. As shown in Figure 7.48, if the provider's business is more than a certain distance away from their member's residential addresses, and assuming the provider is not a "specialist" per se, then one might question why the members would not seek out medical attention closer to their homes. This type of situation might be an indicator of a fraud where the provider solicits business from underserved populations, as was shown when several hospitals and medical centers in the Southern California area were accused of submitting claims on indigents and defrauding Medicare and the Medi-Cal systems of millions of dollars.[57]

[57] Cara Mia DiMassa, Richard Winton, and Rich Connell, "3 Southern California Hospitals Accused of Using Homeless for Fraud," *Los Angeles Times*, August 7, 2008. http://www.latimes.com/news/local/la-me-skidrow7-2008aug07,0,6,5921372.story.

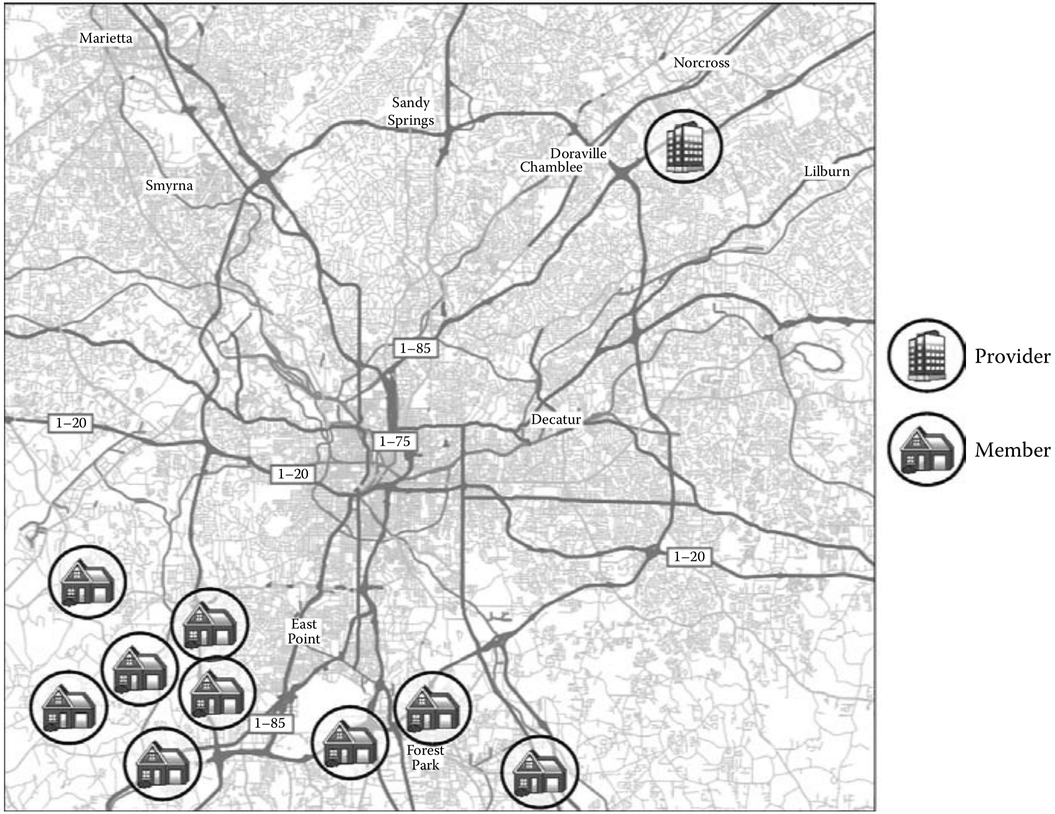


Figure 7.48 Large distances between provider and member addresses.

Conclusion

A number of fraud-centric patterns were discussed and presented in this section, often enhanced through the use of visualization diagrams. Of particular importance is defining the protocols, parameters, and conditions that go into exposing the anomalies. In many of the cases presented, there was no clear-cut right or wrong answer, rather an irregularity in the data, a variance in the values, or an inconsistency in the expected results that stood out as unusual. The majority of the patterns are not particularly complicated to discover and often there are many instances from which to choose and review. Eventually, the data involved in these patterns must be manually reviewed to determine if actual fraud exists. This is analogous to a metal detector signaling an alert that there is something hidden under the surface, and not until the object is dug out of the ground and closely inspected can its true value be determined.

Fraud is a very dynamic entity and is constantly changing, adapting, and morphing itself to take advantage of vulnerabilities and flaws within the oversight and control systems that are established to minimize their presence. For example,[58] a South Carolina parts supplier found a flaw in a purchasing system used by the U.S. Department of Defense, and was able to charge almost $1 million for shipping two 19¢ washers (slated for priority deliver to military operations in Iraq and Afghanistan). The automated system was not outfitted with any type of boundary parameters or internal checks to limit or detect discrepancies in the amounts charged for shipping items that had a “priority” status. The pattern was actually discovered through a manual review when a purchasing agent saw the excessive amounts being charged and rejected the claim.

These types of situations are prevalent throughout many systems and processes. Ultimately, their detection, interpretation, and resolution are up to the customer; they are the ones that determine the tolerance on how much fraud is acceptable and eventually bear additional costs in trying to minimize their losses. Approaching the problem space from different angles, from new starting points, and with nontraditional methods will most likely yield a better return on investment. Additionally, applying analytical techniques from different industries can help increase yields. The trick is in recognizing where and when they should apply.

[58] Tony Capaccio. “Pentagon Paid $998,798 to Ship Two 19-Cent Washers,” *Bloomberg.com*, August 16, 2007, http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a_pIZ20xQxeU.