Fraud and Forensic Accounting In a Digital Environment

White Paper for
The Institute for
Fraud Prevention

Conan C. Albrecht
Marriott School of Management
Brigham Young University
conan@warp.byu.edu

Fraud and Forensic Accounting In a Digital Environment

ABSTRACT

This paper discusses four aspects of computer-aided fraud detection that are of primary interest to fraud investigators and forensic accountants: data mining techniques for the detection of internal fraud, ratio analysis for the detection of financial statement fraud, the issues surrounding external information sources, and computer forensics during fraud investigations. It provides an informative background and then details the current status of research in each area. It describes what is currently unknown, and it proposes future research topics.

Keywords: Fraud, Computer Forensics, Proactive Fraud Detection, Digital Accounting

1 INTRODUCTION

The modern digital environment offers new opportunities for both perpetrators and investigators of fraud. In many ways, it has changed the way fraud examiners conduct investigations, the methods internal auditors use to plan and complete work, and the approaches external auditors take to assess risk and perform audits. While some methods, such as online working papers, are merely computerized versions of traditional tasks, others, such as risk analysis based on neural networks, are revolutionizing the field. Many auditors and researchers find themselves working amid an ever-changing workplace, with computer-based methods leading the charge.

Perhaps the most difficult aspect to computer-based techniques is the application of a single term to a wide variety of methods like digital analysis, electronic evidence collection, data mining, and computer forensics. Indeed, computer-based fraud detection involves a plethora of different technologies, methodologies, and goals. Some techniques require a strong background in computer science or statistics, while others require understanding of data mining techniques and query languages.

In the author's experience, discussions about computer-based fraud detection techniques in most accounting circles are revolve around the use of Benford's Law to discover false invoices or other fraudulent amounts in corporate databases. Analysis of data against Benford's distribution is useful, but it is only one of many computer-based fraud detection techniques that should be used by professionals and researched by academics.

This paper reviews the different aspects of computer-aided fraud detection. In particular, it describes each topical area, its research to date, and needed research for both professionals and academics.

1.1 Types of Fraud Detection

Fraud itself comprises a large variety of activities and includes bribery, political corruption, business and employee fraud, consumer theft, network hacking, bankruptcy and divorce fraud, and identity theft. Business fraud—often called occupational fraud—is usually most interesting to accounting professionals and faculty and is the primary subject of this paper (ACFE, 2008). Within the area of business fraud, many find it helpful to separate between internal and external fraud (Albrecht, et. al., 2008). Internal business fraud involves schemes *against a company* (i.e. to steal money from a company). Internal fraud includes schemes like employee embezzlement and kickback relationships. Internal fraud is usually found by internal auditors or dedicated fraud detection teams through hotlines, data mining efforts, and internal audits.

External business fraud, or financial statement fraud, involves schemes *on behalf of a company*. This is most often done by misrepresentation of the financial statements to improve company image and mislead stockholders and other interested parties. Common external schemes involve revenue and inventory overstatements, liability understatements, inadequate disclosure fraud, and other manipulations to the financial statements and company records (Wells, 2002).

The act of fraud investigation is comprised of several activities, including initial discovery, public record search, interviews of various types, document recovery and search, legal prosecution, and computer forensics. The typical fraud investigator is

heavily involved with many of these activities, but he or she generally works with legal counsel or information security professionals for more specialized tasks.

Even within the relatively narrow field of computer-based fraud detection, significant differences in task performance and knowledge requirements exist. For example, computer forensics requires knowledge of disk cloning, operating systems, file and graphics formats, and scripting for automation. In contrast, data theft prevention and investigation requires knowledge of databases, security, intrusion detection, hacking principles, and encryption.

1.2 Paper Scope

This paper focuses on four central themes that run through the topics described above. These are outlined as follows:

1. *Data mining for fraud*: Techniques and methodologies for discovering fraud in corporate databases.
2. *Financial statement fraud:* Ratio analysis and other methods of finding financial statement manipulation.
3. *External information sources:* Information about perpetrator finances and other data, usually found in websites.
4. *Computer forensics:* Investigating by sifting through computer hard drives and other information devices.

This paper does not cover a number of peripheral topics that might be termed computer-aided fraud detection. IT auditing involves the review of technical systems, such as ERP applications and databases, for risks and control weaknesses. Computer-aided audit risk assessment is the process of using expert systems, neural networks, or other computer-based programs to assess various audit and inherent risks during the planning phases of audits. IT control and security planning involves requirements

definition on permissions and security in a corporate systems. Evidence collection techniques are useful in large fraud cases; their associated software houses large sets of documents, data, and other information in fraud investigations. Finally, deep forensics topics such as encryption cracking and intrusion detection are better suited to computer science journals. The aforementioned topics are beyond the scope of this paper and are not included.

1.3 Structure of the Paper

This paper provides a section for each of the four topics noted above. Because these topics are mini-papers in their own right, the sections are split into three areas: a detailed introduction and description of the topic, a literature review and current status, and a summary of areas still unknown and research possibilities.

2 DATA MINING FOR FRAUD

In 2002, Gene Morse found a round $500 million debit to a PP&E account at WorldCom. He discovered the anomaly through searches in a custom data warehouse he had developed in the Essbase multidimensional database. WorldCom would not give Morse access to full financial systems, so he created his own warehouse and used basic data mining techniques to search it. Using a small script and Microsoft Access, Morse followed the account through the financial reporting system and ultimately discovered a $1.7 billion entry of capitalized line costs in 2001 (Lamoreaux, 2007).

The WorldCom fraud discovery is one example of using computer technology to search full populations of data for anomalies, trends, and fraud. Traditional auditing uses techniques like discovery, stratified, or random sampling to determine whether a population contains errors (Albrecht and Albrecht, 2002). This approach works well

when auditors are searching for anomalies—unintentional errors usually caused by weaknesses in controls—because anomalies occur at regular intervals throughout the data set. In contrast, fraud—intentional errors caused by intelligent human beings—can occur in only a few transactions. While a sample of a population containing anomalies should be representative, a sample of a population containing fraud may not be representative. Assuming a fraud is recorded in only a few transactions, a sampling rate of 5 percent results in a 95 percent risk the fraud will not be sampled and will be missed. Fraud detection methods should use full populations whenever possible, and since full populations can be voluminous, they almost always require computers and data mining techniques.

2.1.1 Methodology

One of the assumptions that underlie traditional auditing methods is the presence of an intelligent human being. When an auditor checks items in a sample, he or she is able to apply human reason and common sense to transactions. Fraud investigations often start with the auditor conducting a routine audit task, looking at a transaction, and saying, "that doesn't make sense." This approach can be seen as an inductive approach; the auditor investigates further when anomalies are found.

Data mining routines—run by computer—do not have this innate sense of normality. Queries and scripts do exactly what they are programmed to do. They do not "dig deeper" unless the user specifically programs them to do so. To accommodate this limitation, the fraud hypothesis testing approach has been proposed (Albrecht, et. al., 2000). This approach has also been labeled the deductive or proactive approach to fraud detection; it involves the following six-step approach:

1. Auditors gain a solid understanding of the business processes, controls, and environment. This understanding allows them to proactively predict the frauds that might be occurring.

2. The team brainstorms the possible frauds that could exist in the environment they are auditing. This might result in 50 potential schemes.

3. Once potential schemes are identified, the team outlines the ways these schemes would show up in data. These indicators, or red flags, are the primary indicators that the fraud may be occurring.

4. For each indicator, the team searches corporate databases using queries, scripts, and data mining techniques. Any anomalous transactions are pulled for further investigation. This could be seen as a "sample" (albeit not in the traditional sense) that should be looked at more closely.

5. Auditors analyze the query results to determine possible explanations for the anomalies, which could be fraud, weak controls, or other reasons.

6. The team follows up on those indicators that may be caused by fraud. These further investigations employ additional queries or traditional means to determine the true cause of the anomalies.

The hypothesis testing approach has been used successfully in several case studies, including liquor sales (Loftus and Vermeer, 2003), university environments (Daily Tar Heel, 2007), and health care in India (Albrecht, 2008).

2.1.2 Continuous Auditing

Once computer queries and scripts are written, continuous auditing is possible. Rather than testing on historical data (the normal audit process), tests can be programmed into live corporate systems to provide continuous monitoring of transactions. Continuous monitoring using information technology has been successfully used at a number of companies (Hermanson, 2006). For a full literature review, see the paper by Rezaee, et. al. (2002) which proposes a generalized approach to audit testing and analysis. The method is similar to the hypothesis testing approach desecribed earlier.

2.2 Common Techniques

The techniques used to perform data mining for fraud range from simple statistical averages to complex neural networks and cluster analyses. This section presents some of the more common techniques found in literature.

2.2.1 Digital Analysis

In 2000, Mark Nigrini published an important book called "Digital Analysis Using Benford's Law" (Nigrini, 2000). Although Benford's Law is now a century old and was discussed in fraud literature (Hill, 1995; Busta and Weinberg, 1998; Nigrini, 1999) prior to the book, Nigrini's work introduced the technique to the large audience of auditors.

Benford's Law works because nature produces more small things than large things. There are more insects than large mammals, more small houses than large ones, and more small lakes than large bodies of water. Similarly, businesses produce more transactions with small amounts than with large amounts. Benford's Law predicts that amounts will start with the digit 1 more often than the digit 9, and it even provides a mathematical formula describing the law and percentages. The digit 1 should show up about 30 percent of the time, while the digit 9 should occur less than 5 percent of the time.

Digital analysis is attractive because it is "deeply nonintuitive and intriguing…, simple enough to be described (if not fully explained) even to those without any formal training in math" (Cho and Gaines, 2007). It can be run on data with little regard to context. For example, when data from a certain vendor are routinely outside the expected percentages, further investigation is almost always warranted.

The primary limitation to Benford's Law is business data do not always follow natural patterns; there exist a large number of reasons that transactions may not match Benford's Law. Explanations like recurring fixed expenses, unusual business cycles, and assigned amounts are often found. The author has taught digital analysis to thousands of professional auditors; in ten years of asking participants about their success with digital analysis, only three individuals have reported finding fraud with Benford's Law (others have reported that digital analysis could have been used to find already discovered frauds, but hind sight is not prediction). In some ways, the audit field may have overestimated the usefulness of digital analysis. But despite its limitations, Benford's Law remains one of the most popular data mining techniques for fraud.

2.2.2 Outlier Detection

One of the primary methods of detecting fraud is discovering data values that are outside the normal course of business. For example, a kickback scheme might be the reason purchases from one vendor are twice as high as similar purchases from another vendor.

The simplest method of outlier detection is the statistical z-score calculation. This formula, given as *(value - mean) / standard deviation*, provides a simple and compact method of measuring outliers. The numerator shifts each point to a zero-based scale, and the denominator adjusts the distribution to a standard deviation of one. Once the data are transformed into this standardized scale, generalized statements can be made. In the author's experience, outlier scores of 5, 8, or even 12 are often found in real world data. At times these may be the result of non-normal distributions, but even in those cases, the score provides an indicator to potential problems.

More advanced techniques have been used in specialized areas. For example, credit card fraud can be discovered by identifying transactions through both unsupervised and supervised learning. Bolton and Hand (2001) used behavioral outlier detection with unsupervised learning to detect abnormal spending behavior as well as increased frequency of use. Others have used regression models, Discrete Gaussian Exponential, depth-based techniques, distance-based techniques, and a number of other techniques to identify outliers. These research streams can be found in Agyemang, et. al. (2005) and Kou, et. al. (2004).

2.2.3 Trending

In addition to comparing same-period numbers from different vendors, employees, or customers, fraud can be discovered by comparing numbers over time. Because almost all perpetrators are greedy (Albrecht, 2008), fraud increases exponentially over time. Auditors can easily spot an increasing trend on a line chart—computers are not needed if only one item is being audited (one employee, one vendor, etc.). The need for automation is during the initial phase of a fraud investigation. If auditors do not know which item is increasing, they must look through thousands of graphs to determine which item requires additional investigation. Trending methods allow the computer to determine which trends are increasing so the auditor can focus on those items.

One of the most basic methods of determining an increasing trend is linear regression. Once the computer fits a line to the data, the slope and goodness-of-fit provide a simple measure of trend. Nonlinear regression and Box-Jenkins analysis

provide more advanced methods of measuring trend. Statistical packages like SAS and SPSS provide full trending modules for the interested auditor.

2.3 Advanced Statistical Techniques

Work has been done in statistical and computer science field on advanced methods for fraud detection. These methods include Bayesian Networks, genetic algorithms, state transition analysis, rule matching, and cluster analysis. See Agyemang, et. al. (2005) and Kou, et. al. (2004) for a detailed review of these and other methods. The reader should note, though, that most of these advanced technical methodologies are not generally used in typical business systems (such as payroll, sales, purchasing databases). They are most often used in highly-specialized areas like credit card fraud, health care claims, and voter fraud.

2.4 Applications

While academic outlets generally publish methodologies and techniques, professional publications highlight the common fraud-oriented data mining platforms (Nigrini, 1999; Coglitore and Matson, 2007; Lehman, 2008; SecurityProcedure.com, 2008). The most popular applications are general-purpose audit programs like ACL and IDEA. Specialized fraud-detection software is making an entry into the market, with fraud module releases from both ACL and IDEA, fraud components in SAS and SPSS, and a dedicated fraud detection program in Picalo.

2.5 Future Research

In the past ten years, researchers have published many methods and techniques for fraud detection. However, these techniques are either not developed sufficiently or

too technical for the typical auditor. Research on techniques is needed in two areas. First, the field needs a better understanding of what simple techniques for outlier detection, trending, link analysis, and full-text analysis are useful for fraud detection. To date, only Benford's Law has seen wide use, and while it is interesting, its effectiveness for actually finding fraud is not well proven—especially on discovering frauds that have not otherwise been discovered with other techniques. Second, a large body of advanced statistical and computer-science literature is available. Application of these techniques to the fraud area has been done in a few papers, but significant work still remains. The literature provides little clarity on which technologies can be used to discover fraud. Both empirical and case study research is needed to determine how these techniques can be successfully implemented.

Beyond the techniques themselves, auditors have little training in computer programming, query languages, and statistics. In addition, they do not have sufficient time to perform these algorithms in typical audits. Detectlets are one method of encasing the knowledge, routine, and algorithm into a wizard-like interface (Albrecht, 2008). They may be able to solve both the training and time problem. Development of detectlets or another solution is needed before the audit field can realize success with these advanced techniques. This is a primary area that information systems and technically-savvy accounting researchers can add valuable knowledge to the audit field.

It is often said that the preparation of data is more difficult than the analysis of it. Research is needed into the best tools, techniques, and methodologies that auditors can use to prepare data for analysis from source business databases. Cutoff points, missing values, abnormal trends, and other difficulties arise during preparation. One exciting

prospect is the automatic conversion of databases from one schema to another (currently being worked on in database circles). If programs could automatically interpret database column types and relationships, they could transform the database from any company into a standard schema. Algorithms written to the standard schema could then be run on the standard schema without worrying about differences between companies and locations. This conversion could be the basis for an off-the-shelf, fraud detection solution.

Finally, many auditors simply do not know how to incorporate data mining into their work. Software companies like ACL and IDEA provide workbooks that can be used in courses, but these workbooks have few examples of direct fraud detection—especially with advanced techniques. Methodologies like the hypothesis testing approach are a first step in providing methodology research, but significant additional research—both empirical and field—are needed to validate and extend the existing methodologies. These methodologies can then be taught in university courses to future auditors.

3 FINANCIAL STATEMENT FRAUD

Statement on Auditing Standards (SAS) No. 99, Consideration of Fraud in a Financial Statement Audit (AICPA 2002), requires auditors to assess the risk that fraud may materially misstate financial statements. Despite SAS 99 being an important requirement towards increased fraud detection, a survey by Marczewski and Akers (2005) revealed that CPAs don't anticipate that SAS 99 will substantially increase audit effectiveness. Another study found that while SAS 99 increased auditor responsibility, most auditors had difficulty identifying fraud and its risks (Beasley and Jenkins 2003).

Notwithstanding these difficulties, failure to detect fraudulent financial reporting not only places the audit firm at risk, but also exposes the audit profession to increased public and governmental criticism. Cecchini, et al. (2006) present evidence that research aiding auditors in assessing the risk of material misstatement during the planning phase of an audit can help reduce instances of fraudulent reporting.

3.1 Ratio Analysis

Traditional methods for the detection of financial statement fraud, such as vertical and horizontal analysis, tip lines, analysis of relationships between management and others, comparisons with industry, and analytical symptoms, are well documented in textbooks (Albrecht, et. al., 2009). More recently, research has focused on ratio analysis for the detection of financial statement fraud.

Ratio analysis involves calculating both traditional and nontraditional financial ratios, such as accruals to assets, asset quality, asset turnover, days sales in receivables, deferred charges to assets, depreciation, gross margin, increase in intangibles, inventory growth, leverage, operating performance margin, percent uncollectibles, sales growth, SGE expense, and working capital turnover. Since ratios standardize firms for size and other factors, one would expect firms within an industry to follow similar trends.

Early studies use statistical techniques like probit and logistic regression, while later studies have branched out to neural networks, classification schemes, and rule ensembles. Studies that use both internal and external data have proven more successful, but the goal of most of the research is to use only externally-available data for analysis (limiting the research to data available to most stakeholders). A review of many of these studies follows.

In an exploratory study, Loebbecke, et al. (1989) refined an earlier model developed by Loebbecke and Willingham (1988) using 77 fraud cases and tested the mappings of various red flags to create classifications in three main areas: conditions, motivation, and attitude. They found at least one factor from each area present in 86 percent of the fraud cases.

Hansen, et al. (1996) developed a generalized qualitative-response model to analyze management fraud using the set of cases developed from internal sources by Loebbeke, et al. (1989). The model was initially tested assuming symmetric misclassification costs between two classes of firms (fraudulent and non-fraudulent). Using over 20 cross-validation trials of five percent holdout cases, an overall 89.5 percent predictive accuracy was realized; but accuracy in predicting fraud was only 62.8 percent. A second model adjusted for asymmetric costs to reflect the greater importance of predicting fraud. While the overall accuracy dropped to 85.3 percent; the accuracy in predicting fraudulent companies increased to 88.6 percent. The corresponding reduction in prediction accuracy for non-fraudulent firms was from 95.5 percent to 84.5 percent.

Green and Choi (1997) used a neural network to classify a large set of fraudulent and non-fraudulent data. The researchers used five ratios (allowance for doubtful accounts/net sales, allowance for doubtful accounts/accounts receivable, net sales/accounts receivable, gross margin/net sales, and accounts receivable/total assets) and three account variables (net sales, accounts receivable, allowance for doubtful accounts) to identify risk. The selection of these variables biased their results toward certain types of revenue frauds; but their method showed potential for neural networks as fraud investigative and detection tools.

Eining, et al. (1997) examined how the following three types of decision aids helped auditors find fraud: checklists, logistic regression models, and expert systems. The expert system used a red flag approach to identify high risk areas and proved to be the most effective of the three.

Summers and Sweeney (1998) investigated the relationship between insider trading and fraudulent companies. Using logistic regression, an initial model was developed based solely on firm-specific financial data. Overall prediction accuracy with this model was 60 percent, with 68 percent accuracy achieved in predicting fraudulent companies. A second model that incorporated both firm-specific financial data and insider trading factors was able to improve overall classification accuracy to 67 percent, with 72 percent accuracy in predicting fraudulent companies.

Beneish (1999) developed a probit model and considered several quantitative financial variables for fraud detection. Five of the eight variables involved year-to-year variations. The study considered differing levels of relative error cost. With 40:1 asymmetric costs, 56 percent of the manipulators were correctly identified in a holdout sample. Results showed that the Days Receivable Index and the Sales Growth Index were most effective in separating the manipulators from the non-manipulators. Most of the ratios used in our research were adapted from this study.

Bell and Carcello (2000) constructed a logistic regression model to predict the likelihood of fraudulent financial reporting. This analysis relied on risk factors identified as weak internal control environment, rapid company growth, inadequate or inconsistent relative profitability, management placing undue emphasis on meeting earnings projections, management lying to the auditors or being overly evasive, the ownership

status (public vs. private) of the entity, and an interaction term between a weak control environment and an aggressive management attitude toward financial reporting. The model was tested on the same sample of 77 fraud engagements and 305 non-fraud engagements used by Hansen, et al. (1996). Fraud was predicted with 81 percent accuracy, and non-fraud was detected with 86 percent accuracy. The model scored better than auditing professionals in the detection of fraud and performed as well as audit professionals in predicting non-fraud outcomes.

More recently, research has focused on ratio analysis with more advanced statistical and computer science methods or with more ratios. Grove and Cook (2004) test the viability of new ratios as red flags. For a detailed review of statistical techniques used to analyze ratios, see Phua, et. al. (2005) and Kirkos, et. al. (2007).

3.2 Future Research

Comparing financial ratios between companies, industries, and generated models still requires much research. To date, the most successful techniques (using external data and holdouts through the entire learning process) have achieved only 70 percent success. Since 50 percent success can be achieved with random guessing, more research is needed to improve the models.

Several potential areas of improvement should be noted. First, traditional financial ratios (like asset turnover and days sales in receivables) have been used to differentiate fraud from nonfraud firms. Research into the development of new fraud-specific ratios might provide insight into formulas that are more successful at detecting fraud.

Second, most research has compared known fraud firms against a similar firm. What defines similar? Should the paired firm be selected based on similar revenues, profits, assets, or liabilities? Can industry models be created that better approximate the industry (and create a better paired "firm")?

Third, the correct time frame of comparison needs research. Is it sufficient to analyze data for 5 years, or do longer or shorter terms make sense? Should yearly or quarterly reports be used? The determination of where the anomalies are found in external data has seen considerable research, but it is still not well understood.

Fourth, most research on ratio analysis does not approximate the real world. Most research makes one to one comparisons, essentially assuming that 50 percent of firms are fraudulent. In reality, fraudulent firms are probably less than a few percent of the total number of firms that report financial statements. This difference becomes important in learning algorithms. Also, the costs associated with Type 1 and Type 2 errors must be factored into many algorithms. The cost of missing an existing fraud is generally regarded as much higher than rejecting a potential client because it might be fraudulent (but, in reality, was not), yet most research simply assumes a 50/50 cost.

Finally, while ratio analysis is the current focus, entirely different streams may provide better results. Exploratory research into new methods of combining publicly-available company information could prove effective. In the end, research may show that financial statements are simply too summarized for effective data mining for fraud. However, financial professionals, shareholders, and indeed, the entire audit profession would be forever changed if computer-based research produced a reliable and repeatable method of discovering fraud in financial statements.

4 EXTERNAL INFORMATION SOURCES

In recent years, a storehouse of publicly-available data has become available on the Internet. Prior to the 1990's, investigators wanting public information had to search through records at courthouses, libraries, universities, and public records offices. Many times, these records were in paper form, and even when they were electronic, the task was daunting to most investigators. Indeed, searching each country courthouse in Texas meant traveling through the state to visit at least 254 locations!

Companies like Lexis-Nexis have long collected news articles, legal documents, and other information. However, the 1990's saw an explosion of information collecting by companies like ChoicePoint, Accurint, and Lexis-Nexis. While the information had always been public, bringing it together into one place represented a leap forward in its accessibility; it brought external information searching into the computer-oriented fraud detection area.

Searching for external information on people and companies becomes important in at least two areas. First, investigations that have identified key individuals can search public records for information on those individuals. These individuals are normally identified through internal data mining efforts, anonymous tips, or other methods. Second, regulations like Sarbanes-Oxley require that auditors pay close attention to top-level management. External information can be useful in determining in top-level management are spending excessively or otherwise engaging in activities that cause concern.

4.1 Available Information

Few academic papers exist on the availability of external information. Because the information is published by private companies that collect the public information into online databases, professional outlets generally carry information on product and content availability (Frost, 2004; Cameron, 2001). Books and textbooks also provide links to web sites with information (Tyberski, 2004; Albrecht, et. al., 2009).

The following is a sampling of the available information for investigators on these web sites. See the above references for a full treatment of the subject.

- Incorporation records
- Property and other asset records
- Civil lawsuits
- Criminal records
- Tax liens
- Civil judgment records
- Bankruptcy filings
- Home values, loans, neighbor contact information
- News articles, current events

The information storehouses like ChoicePoint are not the only source of online information. Mapping, satellite photos, birds-eye views, and entire street views are available from Google and Microsoft. Deep web search engines like *CompletePlanet* and *DeepDyve* provide one-stop searching for sites that normally require logins (and are thus not included in traditional search engines). Other, specialized services exist for most information needs.

4.2 Privacy Issues

One area that has seen considerable research is the privacy issues raised by these new web sites. The new information web sites are not providing any information that had not been previously available. However, by both bringing the information together into

one place and by digitizing paper records, the information is significantly more available to the public. In addition, techniques like link analysis and correlations can be done on the records, effectively creating information that was not previously available. Many worry about the significant loss of privacy that has occurred in recent decades (Black, 2002; Paletta, 2005; Lovett, 1955). Complicating matters is the fact that different states disagree about the online access to court records, making legislation difficult (Swartz, 2004).

For example, when Google released street view—a tool which provides pictoral views of most streets in the USA—the US military asked it to remove all pictures within a few miles of its bases (Technology Expert, 2008). Web sites like *GoogleSightseeing.com* (not affiliated with Google) publish the most interesting, popular, and sometimes embarrassing and intrusive pictures taken by Google cameras. In 2008, Google launched a satellite capable of taking pictures at resolutions previously unknown to private entities (Jones, 2008). Finally, Google Docs, Gmail, and other cloud-based services give increasing amounts of data to a single company. While this paper is not proposing a conspiracy theory or trying to discourage the use of Google services, the company serves as an example of the potential privacy concerns that is occurring within private companies and governments. Both academic researchers and professionals are concerned with the loss of privacy new technologies bring (Stafford, 2005; Samborn, 2002).

Highlighting the risks of compiling too much information into one place is the highly-publicized ChoicePoint breach. In 2005, hackers grabbed data on 163,000 individuals from ChoicePoint's databases (Swartz, 2007; Freeman, 2006). While the

ChoicePoint hacking was extremely serious, it is certainly not the only significant breach; successful hacking attempts are reported weekly in the popular media.

4.3 Future Research

Little academic research has been done on external information gathering. Most needed is methodologies for searching, compiling, and using these data in fraud investigations. Most sources only list the available sources; more structured ways of working with these data are needed. These methodologies can be included in textbooks and university courses.

There has been some work done on privacy issues, especially by the law field. However, more research is needed on how to regulate data providers, how to keep information safe, and where the lines of privacy should be drawn. As these web sites continue their move to international data, differences in country policies and cultures will become important research topics.

5 COMPUTER FORENSICS

Today, almost every financial fraud incorporates the use of a computer, whether the fraud is falsifying invoices or electronic money laundering (Smith, 2005). In the case of financial statement fraud, entries probably exist as electronic journal entries, login records found in log files, and electronic correspondence between involved individuals. In recent years, auditors find themselves increasingly involved in evidence collection through computer forensics.

As it pertains to fraud detection, computer forensics is the process of imaging data for safekeeping and then searching cloned copies for evidence (Gavish, 2007; Dixon, 2005). Perhaps the most common example is seizing the computer of a suspect for

analysis. In gaining access to or auditing the data on a digital device, computer forensics can also involve white-hat (legal) hacking, password and encryption cracking, key logging, digital surveillance, and intrusion detection.

Arguably the most common form of electronic evidence is email correspondence. When an individual sends an email, a copy is normally kept in at least four places: on his or her workstation, on the sending mail server, on the receiving mail server, and on the recipient's computer. Computer forensics on any location should provide the full text of the email, including any attachments.

5.1 Tools of the Trade

As with external data sources, most of the written information about computer forensics is contained in textbooks and professional outlets. The two leading products, EnCase and FTK (described below), are mature software suites. Both software companies provide comprehensive training on forensic software and techniques. Academic outlets are generally reserved for encryption-cracking algorithms, steganography, hacking, operating system weaknesses, and protocols. While these ideas may be useful in some forensic audits, they are beyond the scope of this paper. They can be found in computer science, mathematics, and statistical literature.

5.1.1 Forensic Suites

The two leading software packages are EnCase by Guidance Software and the Forensic Toolkit (FTK) by AccessData (Kuchta, 2001). These suites provide shorter learning curves than previous single-purpose utilities and bring a greater number of professionals to the field more quickly. Both packages provide task-oriented processes

for securing and cloning the hard drive, calculating *md5* or *sha* checksums, searching for graphics, and keyword search.

In recent years, Linux-based tools have become popular as free alternatives to the traditional suites. Helix, the Penguin Sleuth, and Security Tools Distribution are Linux distributions that run directly from CD, providing clean environments for searching a computer without the need for cloning (Causey, 2005). These tools boot a suspect computer directly to Linux and mount the user hard drives in read-only mode, essentially bypassing most passwords and security protections. While Linux-based tools are more difficult to use and do not have the same precedent in court as EnCase and FTK, they have become popular with some auditors.

More specialized tools—usually geared to a single purpose like password cracking or file undeleting—are available (Kuchta, 2001). These tools make up the toolkit of a forensic investigator. Most auditors only need to understand the general categories to effectively work with dedicated forensic personnel.

Forensic investigators generally look for information in the following areas:

- Office suite files in computer directories
- Graphics in directories and in the browser cache
- Email, instant messaging logs, and other computer-based communications.
- Cell phone text message logs and call records.
- Memory and disk caches
- Deleted space on hard drives
- Other digital devices like USB flash drives

5.2 Methodologies

Beyond the training provided by software companies, some academics have researched methodologies and techniques for computer forensics. In particular, Waldrup, et. al. (2004) proposed a generalized five-step process for a defensible forensic

accounting process. They state their process is defensible in court and robust from a technical standpoint.

Smith (2005) researched the relationship between the use of digital data by auditors and computer forensic specialists. By focusing on the roles of the two individuals, Smith investigated how collecting digital evidence can benefit fraud-oriented audits.

Link analysis is a methodology that analyzes links between data found in data mining, from external sources, and during forensic investigations. Some researchers are investigating how link analysis can be effectively used to correlate complex evidence in forensic accounting cases (Kovalerchuk, et. al. 2007).

5.3 Future Research

While significant research has been done on individual forensic techniques in the computer science and mathematics fields, additional research is needed to apply these techniques to the forensic accounting field. Most auditors do not have the background needed to understand how to apply rigorous forensic algorithms; if techniques are not included in EnCase or FTK, they are not generally available to the field. Academics with this knowledge could bring additional techniques to the accounting field.

As can be seen in the methodology section above, only preliminary work has been done in creating robust, researched methodologies for forensic auditors. The proposed methodologies need to be validated and tested, and new methodologies should be proposed.

The proliferation of data-capable devices presents new opportunities for the application of forensics research. Devices like cell phones, iPods and other MP3 players,

digital cameras, PDAs, USB key fobs, external hard drives, and even writing instruments with embedded storage are potential forensic search points. For example, the relatively large hard drive in today's iPods—despite their innocuous and ubiquitous nature—has surely been used to steal data or sensitive information from companies and individuals. As this trend continues, regulation, privacy, and protection will become increasingly-important issues. In addition, the increasing availability of cloud (internet) storage and backup presents new opportunities for forensic auditors.

Finally, forensic auditing is a relatively new field. Research into how forensics should be included in accounting curricula would be useful. Few teachers know which topics should be taught to accounting students and which should be left to information systems or computer science classrooms. Indeed, if dedicated fraud courses are only now entering the curriculum of university programs, computer forensics for accounting majors—if research shows that such a thing should be done—is relatively unknown.

6 CONCLUSION

Computer-aided fraud detection is a new, exciting field for accounting researchers. Topics like data mining techniques, ratio analysis for the detection of financial statement fraud, issues surrounding external information sources, and computer forensics bring opportunities for robust research and for collaboration between accounting faculty and information systems, legal, computer science, mathematics, and other researchers.

Currently, the research is spread across a wide variety of journals, from auditing to information systems to investigative outlets. The goal of this paper is to introduce what is known about each topic and propose needed areas of study; it is not meant to be a

comprehensive literature review on each topic, but it references other reviews where possible.

7 REFERENCES

ACFE (2008). 2008 Report to the Nation. The Association of Certified Fraud Examiners.

Agyemang, M., Barker, K., & Alhajj, R. (2006). A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis, 10, 521-538.

AICPA (2002). SAS No. 99: Consideration of Fraud in a Financial Statement Audit Summary.

Albrecht, C. C., Albrecht, W. S., & Dunn, J. G. (2000). Conducting a Pro-Active Fraud Audit: A Case Study. Journal of Forensic Accounting, II, 203-218.

Albrecht, C. C. (2008). Detectlets: A New Approach to Fraud Detection. In European Academy of Management at Ljubljana, Slovenia.

Albrecht, W. S., & Albrecht, C. C. (2002). Root Out Financial Deception. Journal of Accountancy, 30-33.

Albrecht, W. S., Albrecht, C., & Albrecht, C. C. (2008). Current Trends in Fraud and its Detection. Information Security Journal: A Global Perspective, 17(1).

Albrecht, W. S., Albrecht, C. C., Albrecht, C. O., & Zimbelman, M. (2009). Fraud Examination (3). South-Western Cengage Learning.

Beasley, M. S., & Jenkins, J. G. (2003). The Relation of Information Technology and Financial Statement Fraud. Journal of Forensic Accounting, 4, 217-232.

Bell, T. B., & Carcello, J. V. (2000). Research Notes, A decision aid for assessing the likelihood of fraudulent financial reporting. Auditing: A Journal of Theory and Practice, 19(1), 169-175.

Benish, M. (1999). The Detection of Earnings Manipulation. Financial Analysts Journal, 55, 24-36.

Black, J. (2002). Public Records in Public View--Online?. Business Week Online.

Bolton, R. J., & Hand, D. J. (2001). Unsupervised profiling method for fraud detection. In Conference of Credit Scoring and Credit Control VII, Edinburgh, UK.

Busta, B. (1998). Randy Weinberg. Using Benford's law and neural networks as a review procedure, 13(6), 356-366.

Cameron, P. (2001). You Can't Hide from Accurint. Law Technology News, 8(9).

Causey, B. (2005). How To Respond To Attacks. Certification Magazine.

Cecchini, M., Aytug, H., Koehler, G., & Pathak, P. (2005). Detecting Management Fraud in Public Companies. University of Florida Fisher School of Business.

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. The American Statistician, 61(3), 218-223.

Coglitore, F. J., & Matson, D. M. (2007). The Use of Computer-Assisted Auditing Techniques in the Audit Course: Further Evidence. Journal of Forensic Accounting, VIII, 201-226.

Daily Tar Heel (2007). Report on Social Security Number Fraud. University of North Carolina at Chapel Hill, April 26.

Dixon, P. D. (2005). An overview of computer forensics. IEEE Potentials, 24(5), 7-10.

Eining, M. M., Jones, D. R., & Loebbecke, J. K. (1997). Reliance on Decision Aids: An Examination of Auditors' Assessment of Management Fraud. Auditing: A Journal of Theory and Practice, 16(2).

Freeman, E. H. (2006). Disclosure of Information Theft: The ChoicePoint Security Breach. Information Systems Security, 11-15.

Frost, M. (2004). Finding Skeletons in Online Closets. Searcher, 12(6), 54-60.

Gavish, A. (2007). The Hidden Costs of Computer Misconduct. Security.

Green, B. P., & Choi, J. H. (1997). Assessing the Risk of Management Fraud Through Neural Netowrk Technology. Auditing: A Journal of Theory and Practice, 16(1).

Grove, H., & Cook, T. (2004). Lessons for Auditors: Quantitative and Qualitative Red Flags. Journal of Forensic Accounting, V, 131-146.

Hansen, J., McDonald, J., Messier, W., & Bell, T. (1996). A Generalized Qualitative-Response Model and the Analysis of Management Fraud. Management Science, 42, 1022-1033.

Heel, D. T. (2007). Report on Social Security Number Fraud. University of North Carolina at Chapel Hill.

Hermanson, D. R., Moran, B., Rossie, C. S., & Wolfe, D. T. (2006). Continuous Monitoring of Transactions to Reduce Fraud, Misuse, and Errors. Journal of Forensic Accounting, VII, 17-30.

Hill, T. P. (1995). A Statistical Derivation of the Significant Digit Law. Statistical Science, 10, 354-363.

Jones, M. W. (2008). Does Google Have Its GeoEye On You. . Electronic document, http://tech.blorge.com/Structure:%20/2008/10/10/does-google-have-its-geoeye-on-you/, accessed .

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining for the detection of fraudulent financial statements. Expert Systems with Applications, 23.

Kou, Y., Lu, C., & Sirwongwattana, S. (2004). Survey of Fraud Detection Techniques. In 2004 International Conference on Networking, Sensing, and Control (pp. 749-754).

Kovalerchuk, B., Vityaev, E., & Holtfreter, R. (2007). Correlation of Complex Evidence in Forensic Accounting Using Data Mining. Journal of Forensic Accounting, VIII, 53-88.

Kuchta, K. J. (2001). Your Computer Forensic Toolkit. Information Systems Security.

Lamoreaux, M. (2007). Internal Auditor Used Computer Tool to Detect WorldCom Fraud. Journal of Accountancy, 35.

Lehman, M. W. (2008). Join the Hunt. Journal of Accountancy, 46-49.

Loebbecke, J., Eining, M., & Willingham, J. (1989). Auditors' Experience with Material Irregularities: Frequency, Nature, and Detectability. Auditing: A Journal of Theory and Practice, 9, 1-28.

Loebbecke, J. K., & J. J. Willingham, J. (1988). Review of SEC Accounting and Auditing Enforcement Releases. University of Utah.

Loftus, J. T., & Vermeer, T. E. (2003). Pro-Active Fraud Auditing: Technology, Fraud Auditing, and Liquor. Journal of Forensic Accounting, IV, 307-310.

Lovett, R. W. (1955). Looking Around. Harvard Business Review.

Marczewski, D., & Akers, M. (2005). CPA's Perceptions of the Impact of SAS 99. CPA Journal, 75, 38-40.

Nigrini, M. J. (1999). I've Got Your Number. Journal of Accountancy.

Nigrini, M. J. (2000). Digital analysis using Benford's Law. Global Audit Publications.

Paletta, D. (2005). Regulating ChoicePoint: Whose Job Is It, Anyway?. American Banker.

Phua, C., Lee, V., Smith, K., & Gaylor, R. (2005). A comprehensive survey of data mining-based fraud detection research. Working Paper.

Procedure, S. (2008). Retrieved November 17, 2008 from http://www.securityprocedure.com/download-picalo-open-source-alternative-acl-audit.

Rezaee, Z., Sharbatoghlie, A., Elam, R., & McMickle, P. L. (2002). Continuous Auditing: Building Automated Auditing Capability. Auditing: A Journal of Theory and Practice, 21(1).

Samborn, H. V. (2002). No Place To Hide. ABA Journal.

Smith, G. S. (2000). Black Tech Forensics: Collection and Control of Electronic Evidence. Journal of Forensic Accounting, I, 283-290.

Smith, G. S. (2005). Computer Forensics: Helping to Achieve the Auditor's Fraud Mission. Journal of Forensic Accounting, VI, 119=134.

Stafford, A. (2005). Information Brokers: Privacy In Peril. PC World, 101-104.

Summers, S. L., & Sweeny, J. T. (1998). Fraudulently Misstated Financial Statements and Insider Trading: An Empirical Analysis. The Accounting Review, 73(1), 131-146.

Swartz, N. (2004). U.S. States Disagree About Online Access to Court Records. Information Management Journal, 11.

Technology Expert (2008). Google's Street View "Off the Map" at U.S. Military Bases. Retrieved November 17 from http://technologyexpert.blogspot.com /2008/03/google-street-view-off-map-at-us.html.

Tyburski, G. (2004). Introduction to Online Legal, Regulatory, and Intellectual Property Research. South-Western Educational Pub.

Waldrup, B., Capriotti, K., & Anderson, S. C. (2004). Forensic Accounting Techniques: A Defensible Investigatory Process for Litigation Purposes. Journal of Forensic Accounting, V, 1-16.

Wells, J. T. (2002). Occupational Fraud: The Audit as Deterrent. Journal of Accountancy.