The leading edge of high tech
The collaboration between data mining and digital forensics

As fraudsters become more sophisticated, schemes more complex and technology more innovative, fraud examiners’ approaches must also evolve. Common investigative techniques include interviews, document review, data mining and digital forensics. With technology playing a larger role in fraud schemes, digital forensics and data mining are becoming increasingly important tools. This column is a six-part series for educators and practitioners that will examine the importance of integrating digital forensics and data mining aspects of investigations, key technologies, methodologies and future applications and how to present them to budding fraud examiners. For some readers, these columns will contain review material, but hopefully all will benefit from information that we need to transmit to higher-education students.

The rate of change in most, if not all, areas of business is accelerating with each passing year. Most professionals view these changes with mixed emotions. On the one hand, new innovations and technology tend to enhance professionals’ abilities to provide their services effectively and efficiently. On the other hand, the more that’s possible, the more that’s expected.

CFEs can’t be idle bystanders with a passive interest as innovations arise. They must openly embrace new knowledge that will drastically affect their professional services. Educators should stress to their students that the rate of change in the technology world — particularly, data mining and digital forensics — has major implications for the fraud examination profession.

Digital forensics primer
Digital forensics is the collection, preservation, analysis and reporting of digital evidence in an investigation. Prominent forms of digital evidence include computers, email, servers, mobile devices such as smartphones and tablets, cloud storage and external storage devices.

The sources of information available through forensic analysis are rich. On many forms of digital media, examiners can recover deleted activity and find communications never intentionally saved by the user in areas such as the computer’s memory. Further, with the advent of smartphones, recovered photos can bear GPS coordinate information, and “call detail reports” from a cell phone provider can assist with...
determining the location of a phone and user at a given time. Other traditional sources include but aren't limited to:

- Internet history and related artifacts.
- Chat/instant message history.
- iPhone backups.
- Chronological timeline of events on devices.
- Financial records and software.

Forensic software allows a fraud examiner to search digital media in a relatively efficient manner — usually beginning with simple keyword searches and analysis of logfiles generated by the digital device. The software can also present historic activity in a chronological timeline to assist in corroborating digital activity with other events.

Digital forensics enables fraud examiners to use more exotic forms of analysis, such as:

- Artificial intelligence-assisted searches for relevant content.
- Detection of emotional tone of communications.
- Collection of names, places, events and dates to construct "relationship maps" of possible related parties, which fraud examiners couldn't otherwise detect because they might be too obscure or separated by too many degrees.

**Data mining**

The definition of data mining varies depending on the industry, business process, application or expert. In fact, many individuals and organizations use data mining and data analytics interchangeably. There's nothing wrong with this — assuming you define these terms in the same way.

Data mining, at its core, is using sophisticated technology to analyze a large set of data to identify trends, patterns or relationships. These patterns, trends and relationships may be known or previously unknown. The foundation for the analysis is the desire to learn something, or answer a question, about the data set.

Therefore, at its foundation, data mining is the use of technology to answer a strategic question(s) regarding a set of data. In the context of occupational fraud then, data mining helps answer the strategic question: "Where in my organization does/could fraud exist?"

Some of the most common applications of data mining in a fraud investigation context include:

- Identification of employee/vendor relationships.
- Identification of fictitious vendor indicators.
- Geographic proximity analysis.
- Trend analysis of vendor payments.
- Identification of ghost employee indicators.
- Trend analysis of payroll hours and earnings.
- Analysis of purchasing card activity for indications of personal use.

**Marrying the two**

Historically, digital forensics and data mining have existed separately in the fraud examination arena. Investigators used digital forensics to analyze unstructured data and data mining for structured data. In fact, mining structured data has become an established component of fraud examinations and forensic accounting investigations.

Structured data, as the name implies, is any kind of data with a consistent, reliable structure and includes sources such as spreadsheets, databases and most data in accounting information systems. Unstructured data, on the other hand, includes everything else, such as text, email, documents, social media, audio, video and many forms of web-based content. Much of this data is generated by humans, contextual and often filled with emotion.

Gartner Research, an information technology research and advisory company, indicates that unstructured data now accounts for about 80 percent of all available data in an organization. Also, technologies in both digital forensics and data mining have reached the stage where they can work together. (See Gartner’s May 16, 2005, report, “Introducing the High-Performance Workplace: Improving Competitive Advantage and Employee Impact” at https://tinyurl.com/oqzc738.)

Fraud examiners may now extract the results of digital forensics analyses and export them into formats readable by most data mining software. Data mining software also has the ability now to read much more than just the traditional structured data format. This results in an abundance of opportunities to integrate these technologies for an even more powerful investigation.

**Applications**

One of the most frequent applications of this integrated approach is email analytics. Email, as a source of evidence, not only contains word-for-word communications but also possesses a date/time element, metadata and emotional tone as expressed through various idioms, phrases and adjectives. Therefore,
email analysis shouldn’t be limited to keyword searches alone. It also should include extraction of meaning and topics; emotional tone of conversations; the creation of relationship networks to visualize how key players and topics interact, influence and evolve over time; and integration into conflict of interest and related party testing based on new relationship information learned.

The underlying concept of this application is text mining or text analytics. This is the central concept in the integration of digital forensics and data mining — accomplished through numerous methods and technologies, which we’ll cover over the next few columns.

**Importance in higher education**
The increased application of digital forensics and data mining in the investigation of occupational fraud means a shift in the desired skillsets of both experienced professionals and college graduates. While organizations with more resources might continue to use programming specialists, digital forensics specialists, and analyst and accounting professionals for investigation teams, smaller organizations are looking for well-rounded individuals who can utilize all the latest technological tools.

Many higher-education institutions already recognize this shift; curricula now include accounting principles, data analysis software and basic digital forensics. As the technology continues to evolve, educators must keep pace with changing skill sets.

**Importance to practitioners**
Experienced professionals are finding it increasingly important to not only be familiar with the capabilities and basic structure of both digital forensics and data mining but also be able to explain the concepts and methodologies. Hopefully, this column series will help both educators and practitioners stay current on technology for occupational fraud and abuse examinations.

**Upcoming columns**
Over the next few issues, we’ll be covering the concepts, methodologies and technologies for accomplishing the successful integration of digital forensics and data mining.

Upcoming topics include:
- Text mining.
- Augmented intelligence and data visualization.
- Network relationship analysis.
- Technologies on the horizon.
- Future of digital forensics and data mining collaboration.

Les Heitger, Ph.D., Educator Associate, is BKD Distinguished Professor of Forensic Accounting in the School of Accountancy at Missouri State University in Springfield. He’s chair of the ACFE Higher Education Advisory Committee. His email address is: LesHeitger@MissouriState.edu.

Jeremy Clopton, CFE, CPA, ACDA, is senior managing consultant in the Forensic Practice of BKD Forensics, LLP. His email address is: jclopton@bkd.com.

Lanny Morrow, CFE, is a technology expert in the Forensic Practice of BKD Forensics, LLP. His email address is: lmorrow@bkd.com.

**AUDITING FOR INTERNAL FRAUD**

**DENVER, CO August 18-19, 2014**

Explore the challenges that auditors traditionally face in identifying fraud and learn the fundamentals of auditing for fraud, with an emphasis on understanding the common schemes, detection techniques and methods of preventing occupational fraud.

Register by July 18 and SAVE $95! ACFE.com/AIFDenver