Abstract: In the last several years, there has been a greater increase in the prominence of fraud detection. Today auditors are using data mining tools and techniques to examine the entire population of transactions in order to select samples for test controls and identify fraud. While the greatest gains in efficiency can be achieved by maximizing the use of technology, auditors must have access to the data and software tools as well as the techniques and knowledge necessary to make intelligent use of vast amounts of financial and non-financial information in their mission of detecting and deterring fraud. This paper presents the concept of economic fraud, categorizes and compares the different missions of fraud review like financial audit, fraud examination and fraud audit. It defines the concept of data mining and its life cycle from a fraud auditing point of view.

Key-Words: data mining, data mining life cycle, economic fraud, fraud audit, data mining program

1 Introduction
In the last several years, there has been a greater increase in the prominence of fraud detection.

The problems at WorldCom and Enron have emphasized not only the importance of audit but also the devastating effects that fraud can have on a company and its auditors, and highlighted the importance of preventing and detecting fraud. Many corporations’ managers, employees, stockholders, and government oversight agencies continue to fight with the unprofessional, unethical, and/or downright criminal conduct perpetrated by miscreants from inside and outside corporations. In addition, governments have developed new rules and regulations to ensure accurate financial reporting, such as the Sarbanes-Oxley Act.

The Report to the Nation on Occupational Fraud and Abuse, a study conducted by the Association of
Certified Fraud Examiners (ACFE) in 2010, suggests that the median losses for the company were about $160,000. Nearly one third of the fraud schemes caused a loss to the victim organization of more than $500,000 and almost one quarter of all reported cases topped the $1 million threshold. The report by the ACFE also measured the common methods of detecting fraud. Tips and complaints have consistently been the most effective means of detecting frauds, accounted for 35.8%, management review 15.4%, internal audit 11.6%, by accident 11.2%, account reconciliation 8.2%, document examination 6.0%, external audit 5.2%, monitoring 2.6%, police 2.5%, confession 1%, IT controls 0.5%.

The ACFE defines “occupational fraud and abuse” (employee frauds) as: “the use of one’s occupation for personal gain through the deliberate misuse or theft of the employing organization’s resources or assets.” and financial statement fraud as: “the deliberate misrepresentation of the financial condition of an enterprise accomplished through the intentional misstatement or omission of amounts or disclosures in the financial statements in order to deceive financial statement users.”

In recent years, professional auditing bodies have developed new audit standards and statements like Statement on Internal Auditing Standards 3 (SIAS 3) “Deterrence, Detection, Investigation and Reporting of Fraud”, SAS 53 “The Auditor’s Responsibility to Detect and Report Errors and Irregularities (provides guidelines for auditors in detecting fraud)”, SAS 99 “Consideration of Fraud in a Financial Statement”, pressuring more on auditors to deter and detect fraud. The International Federation of Accountants (IFAC) addressed one of the most important issues facing auditors today—the responsibility for detecting fraud—by releasing an International Standard of Auditing (ISA) entitled The Auditor’s Responsibility to Consider Fraud in an Audit of Financial Statements (ISA 240). It states that auditors should be alert to risks of material misstatement due to fraud and are required to assess any such risks encountered during the course of an audit.

In exercising the “due professional care”, external auditors must be ever alert to the possibility of criminal activity, wrongdoing, conflict of interest, inefficiencies, and other abuses, they need to be alert to the possibilities of intentional wrongdoing, errors and omissions, inefficiencies, waste, ineffectiveness, and conflicts of interest (Singleton T., 2010).

Financial auditing typically refers to the process of evaluating compliance of financial information with regulatory standards, usually for public companies, by an external, independent entity. The well-publicized SOX incorporates concepts and procedures to deter and to catch fraud in audits of internal controls over financial reporting. However, the focus of financial audits and financial reporting ultimately is concerned with providing reasonable assurance that a material misstatement to financial statements has not occurred, regardless of the reason (Singleton T., 2010).

Fraud auditing is a very different term, encouraging the detection and prevention of frauds in commercial transactions. Fraud auditing is the process of detecting, preventing, and correcting fraudulent activities. In the broadest sense, it is an awareness of many components of fraud, such as the human element, organizational behavior, knowledge of fraud, evidence and standards of proof, an awareness of the potentiality for fraud, and an appreciation of the red flags (Singleton T., 2010).

The deterrence and detection of fraud, waste, and abuse requires an increasing amount of auditor time and energy. Fraud examiners and investigators are also finding more demand for their services, as companies facing difficult economic times can not afford to have profits wasted or stolen.

While the greatest gains in efficiency can be achieved by maximizing the use of technology, auditors must have access to the data and software tools as well as the techniques and knowledge necessary to make intelligent use of vast amounts of financial and non-financial information.

### 2 Data mining – basic concept

The electronic environment in which companies operate, along with the controls on that environment, presents an array of complex systems, real-time variances, and worldwide applications. It is a major challenge for auditors to detect fraud in the IT systems. But it also provides a broad range of opportunities for the use of powerful interactive audit software and advanced auditing techniques. Thus, while the business environment is rapidly becoming more complex, there is also an increasing array of audit software, tools, and techniques to assist in fraud investigations.

With the ever increasing system complexity, especially the computer-based accounting information systems including enterprise resource planning (ERP) systems, and the vast amount of transactions, it is impractical for auditors to conduct the overall audit manually. Auditing with the computer describes the employment of information technologies by auditors to perform some audit
work, referred to as computer assisted auditing tools (CAATs).

Generalized audit software (GAS) is an automated package originally developed in-house by professional auditing firms. It can be used by the auditor mostly in the execution and documentation phase (planning, analysis tools, calculation tools, sample selection tools, data manipulation tools, documents preparation tools). The techniques used by GAS should be employed appropriately to accomplish the audit objectives. Those techniques are: test data, integrated test facility, parallel simulation, system testing, continuous auditing).

Data mining is a set of computer-assisted techniques designed to automatically mine large volumes of integrated data, for new, hidden or unexpected information, or patterns. In recent years, data mining has been studied extensively, especially on supporting customer relationship management (CRM) and fraud detection. The limitation of current CAATs combined with significant levels of fraud suggests the need for additional analytical procedures for an effective detection of economic fraud.

### 3 Fraud data mining life-cycle

In the traditional sense, data mining is the use of data extraction software to examine data. Today, auditors are using data mining tools and techniques to examine the entire population of transactions in order to select samples for test controls and identify fraud. The general process model for data mining (CRISP-DM) can be successfully extended for fraud detection data mining projects.

#### 3.1 Business understanding

This initial phase converts project objectives and requirements from a business perspective into a data mining problem definition.

The fraud audit is defined as a set of audit procedures performed over a business transaction population in order to increase the likelihood of identifying fraud. Using data mining in fraud audit or “Fraud data mining” is “the process of obtaining and analyzing transactional data to identify anomalies or patterns indicative of a specific fraud scheme” (Vona L., 2008).

The objective is to find a discrete number of transactions that can be examined using fraud audit procedures. Fraud auditing final purpose is to identify one fraudulent transaction and afterwards have the audit plan dictate how the sample containing the transaction will be extended. Various data mining tools and techniques are used to identify transactions consistent with a specific fraud scheme. The resultant data provides the basis for identifying the fraud scheme variations, concealment strategies and associated red flags, and the fraud opportunity. (Vona L., 2008)

#### 3.2 Data understanding

Starting with an initial data collection, the auditor proceeds with various activities helping him get familiar with the data structure and actual data content. Multiple data sources are not uncommon and there may be more than one type. Interrogating each data source can also produce results in different formats. It is during this phase the auditor understands how data is being populated and retrieved across all data sources.

For data sources directly accessible from a database, the database schema is being analyzed including table relations consisting of primary and foreign keys, data insertion procedures, various SQL triggers on different types of SQL operations. Not all data sources can be queried directly. In some cases the data can only be accessed through various third party software applications or web services. In order to retrieve the data, the auditor needs to be familiar with the corresponding APIs (application programming interface) or WSDL (web service description language) models.

#### 3.3 Data preparation

This phase covers all activities to construct the final data set from the initial raw data. It also includes data integration combining all data sources (databases, third party software applications, web services) into a single, central repository. Not all available data is relevant for fraud auditing. Restricting the initial data set to only the necessary, complete and consistent data will drastically improve the accuracy and reduce the resource requirements of subsequent data mining operations in terms of memory, disk space and cpu usage.

Table and attribute selection are being applied to only target relevant data. With the knowledge of database schemas containing detailed information about each database table field, correlations can be established between a given fraud scheme and existing data. Table fields containing different type of information can be used for different type of fraud analysis. Duplicate and sequential data analysis can be performed on transaction control numbers, stratified data analysis can be performed on transaction codes allowing various types of data aggregations, change and circumvention analysis can be performed on various account types.
Exclusion and inclusion filters further reduce the data set. Based on exclusion and inclusion theory, homogeneous transaction populations are being generated, excluding non relevant transactions for a given fraud schemas and including relevant ones. Both filters need to be applied in order to have full control on data reduction. In addition to the above, data merging can be performed with the help of aggregation functions, such metrics being widely used in fraud detection.

Inconsistent and incomplete data is being removed as part of the cleansing process which stands out for fraud auditing compared to other data mining projects. Because in fraud auditing incomplete and inconsistent data can signal a fraud schema, special precautions need to be taken in order to ensure the data cleansing process doesn’t eliminate fraud evidence.

3.4 Modeling against fraud data profile
For each fraud schema various data mining techniques are being applied with their parameters being calibrated to optimal values. These techniques can be largely classified into four main classes with clustering, classification, regression and association rule learning tasks. Data mining methods used in the fraud audit process belong to all four classes and can be classified into the following categories: data description, dependency analysis, classification analysis, cluster analysis, outlier analysis, evolution analysis.

3.5 Sampling plan evaluation
Each data mining technique can generate different data groups or samples containing transactions consistent with the given fraud scheme. In this phase a data mining technique or a combination of data mining techniques producing the best results is being selected as the sampling method for the given fraud schema. The main selection criteria is the rate of false positives, transaction identified as matching the fraud risk but being in fact non fraudulent.

3.6 Sample generation
With the selection of data mining techniques for the given fraud schema, sampling is now possible. Fraud data mining is all about searching for a specific error. The approach to sampling is focused, nonrandom, and biased toward the given fraud scheme. The intent is not to offer an opinion on the effectiveness of the internal controls, but to search for fraudulent transactions. (Vona L., 2008).

3.7 Audit analytical procedures
Performing audit analytical procedures over the data mining generated sample will yield audit evidence whether or not fraud is present within the sample. Audit analytical procedures typically involve performing test controls including observations, investigations, and account reconciliations. Following the conducted tests a “reasonable assurance” can be obtained against the verified data, allowing auditors to emit the audit opinion.

3.8 Sample expansion policies
Each fraud schema has its own expansion policy allowing additional, related data to be retrieved. Such policies are conduction following an unqualified audit opinion in order to discover the full fraud extent throughout the company.

4 Conclusions
Dealing with intricate transactions in large volume requires a considerable more effort from an increasing qualified personnel, more advanced tools to capture, analyze, present and report data. This increases costs and reduces the overall profitability. Covering the entire transaction data population, data mining techniques can increase fraud auditing accuracy at a fraction of the cost. Some techniques capable of unsupervised learning are even capable to automatically adapt to new fraud schemas and patterns.

No data mining technique can produce meaningful results from invalid or inaccurate data. Prior to any data modeling, it is vital to prepare the raw data by performing selection, cleaning and formatting operations. Given a population of transactions, the business understanding and data understanding life cycle stages are conducted once for all fraud schemas. The remaining life cycle stages are influenced by the targeted fraud schema and are conducted separately, multiple times for each fraud schema. This keeps the number of false positives to a minimum having different data mining implementations for different fraud schemas.

References:


