Impact of Information Technology on Data Mining Field Methodology in Security Systems and Event Management

Omorogbe Osasu Harry, Asibor Raphael Ehikhuemhen

Abstract: ***This paper gives an overview of data mining field & security information event management system as a technique for knowledge management in business process redesign, in this paper we discuss potential use of data mining techniques in mining. It reviews the basic techniques and methods of data mining and proceeds to identify possible mining applications of this methodology. In particular the paper proposes use of data mining to develop predictive capacity related to condition and performance of mining equipment. Other possible uses of data mining include optimization of mine performance as well as equipment operator training***

Index Terms: Data mining, security information event management system.

I. INTRODUCTION

Modern mine control systems and mine equipment are highly computerized. One result of this situation is that large volumes of data are collected that define mine performance and equipment condition. Some of this data is processed in real time to provide information that allows for optimization of mine performance. Examples are the fleet dispatch systems that develop equipment assignments, best matched to the stated objectives of the mining operation and based on real-time processing of data that defines equipment status and location. Most of the collected data, however, is used for reporting and post-mortem analysis of mine performance, for equipment failure analysis and for prevention of its catastrophic failures only.

An example are the vital signs monitoring systems installed on larger pieces of mining equipment. These systems collect data generated by a variety of sensors and store it to facilitate easy failure diagnostics. In addition these systems have a capability to warn the operator of impending failure or to conduct orderly equipment shut-down if an emergency situation occurs.

One of the major developments which had profound impact on the economic growth pattern in the world in the new millennium has been the strides in the domain of Information Technology sector. The world has observed significant growth of applications in diverting areas of Information Technology. Information Technology has permeated nearly every aspect of modern business operations and communications. Information Technology is a powerful force in today's global society. The advent of computers and Information Technology (IT) has been perhaps the single massive drive impacting organizations during past few decades. Information Technology or IT is revolutionizing all the living ways. No doubt, it has given a new meaning to the word "Convenience". Information Technology has drastically changed the business landscapes and word "IT" has become the "Catchword" of the modern life today. Information Technology has become, within a very short time, one of the basic building blocks of modern industrial society. The effective use of IT is an essential element of competing in a fast-paced, knowledge based economy. Information Technology is the major contributor to the progress of the developed countries. Information Technology has been defined in various ways by different authors. Over the years, IT has been conceptualized and measured differently by different researchers. The majority of the authors, however, parallel Information Technology with computer systems. "Information Technology is the term that describes the organization's computing and communications, infrastructure, including computer systems, telecommunication networks, and multimedia, hardware and software". Also "IT includes hardware, software, databases, networks, and other related components which are used to build information systems" Many other researchers also have come up with the same idea and say that "IT is the technology that supports activities involving the creations, storage, manipulation and communication of information together with their related methods and management applications" similarly, Information Technology can be consider as general term that describes any technology that help to produce, manipulate, process, store, communicate, and/or disseminate information. This definition may be regarded as the comprehensive one, as it covers all aspects discussed by different researchers and includes all the components and processes needed to carry out information processing work in the organization. So it can be said that that IT concept came from a merging of computer with telecommunications

Availability of huge databases and spreading computerization has led to large strides in data processing capabilities and techniques.

Revised Version Manuscript Received on November 06, 2017.

Mr. Omorogbe Osasu Harry, Lectures, Department of Computer Science/Information Technology, Igbinedion University, College of Natural & Applied Sciences, Okada, Nigeria. E-mail: harresarch@yahoomal.com

Dr Asibor Raphael Ehikhuemhen, Lectures, Department of Computer Science/Mathematics, Igbinedion University, College of Natural & Applied Sciences, Okada, Nigeria. E-mail: asibor.raphael@iuokada.edu.ng

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Variety of powerful data processing methods have been developed over the last years that facilitate rapid processing of voluminous data for extraction of user friendly information. One of such methods is data mining. Originally developed by intelligence community to look for information in huge communication databases, data mining has since found a range of commercial and scientific applications. Nowadays it is widely used by retail industry to analyze sales, direct promotion and marketing efforts, by cellular telephone companies to assure client retention, by scientists to search for information in large databases created by Hubble space telescope, and in many other applications.

This paper briefly reviews data mining and the related techniques, and proposes their use for discovery of knowledge in data acquired by a variety of data acquisition systems used in today's mines. In particular the paper suggests that data mining can be used to develop predictive capacity related to equipment condition and its performance. Data mining offers a potential for further, significant improvement of mine performance.

II. DATA MINING

Data mining derives its name from the similarities between searching for gold in mines. In gold mines we search for very small particles of gold in tons of soil. Similarly in data mining we search for valuable information from huge amount of data collected in various ways. Data mining, a synonym to "knowledge discovery in databases" is a process of analyzing data from different perspectives and summarizing it into useful information. It is a process that allows users to understand the substance of relationships between data. It reveals patterns and trends that are hidden among the data. It is often viewed as a process of extracting valid, previously unknown, non-trivial and useful information from large databases [1].

Data mining is an iterative process that involves setting the objectives of the search, selecting and cleaning input data, transforming it, running a mining function and interpreting Security information and event management system is the industry-specific term in computer security referring to the collection of data typically log files or event logs from various sources into a central repository for analysis. Event logs are generated by various networking devices, Operating Systems and Application Servers. Event logs give raw input of all activity happening in IT infrastructure of any organization. This raw data act like input to SIEM system which provides us security alerts, reports as an output. The processing of all raw data is achieved using data mining technique.

[1]. Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales [2].If scope of data mining is applied to all events logs generated by various networking devices, system and application servers then efficiency of enterprise security can be drastically increased.

The real problem in today's enterprise security is amount of logs generated by various systems. Organizations often put too much faith in their new shiny firewalls, IDSs, or antivirus software. Once one or more of these solutions are implemented then IT staff realizes that interpretation of all logs generated by this solution is big challenge. A network could either perform as a well-tuned orchestra or as several pieces that play wonderfully by themselves but give you a headache when they are all brought into the same room. Each individual security component could be doing its job by protecting its piece of the network, but the security function may be lost when it is time to interrelate or communicate with another security component. SIEM system helps us to take an architectural view, where we can look at the data flow in and out of the environment, how this data is being accessed, modified, and monitored at different points, and how all the security solutions relate to each other in different situations.

III. DATA MINING BASICS

Our capabilities of both generating and collecting data have been increasing rapidly. The widespread use of bar codes for most commercial products, the computerization of many business and government transactions, and the advances in data collection tools have provided us with huge amounts of data. Millions of databases have been used in business management, government administration, scientific and engineering data management, and many other applications. It is noted that the number of such databases keeps growing rapidly because of the availability of powerful and affordable database systems. This explosive growth in data and databases has generated an urgent need for new techniques and tools that can intelligently and automatically transform the processed data into useful information and knowledge.

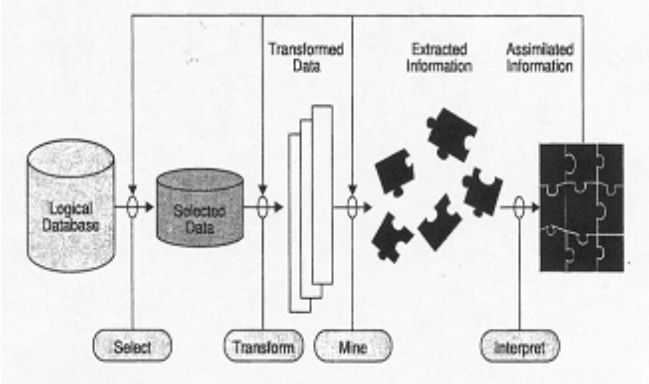
Consequently, data mining has become a research area with increasing importance [3]. We need information but what we have is a huge amount of data flooding around. Because of the amount of data is so enormous that human cannot process it fast enough to get the information out of it at the right time, the data mining technology has been established to solve this problem potentially. The ultimate goal of knowledge discovery and data mining process is to find the patterns that are hidden among the huge sets of data and interpret them to useful knowledge and information

A. Typical Data Mining Architecture

Based on storage & retrieval of data following architecture is possible the results. The schematic in fig.1, adopted from IBM (International Business Machines, 2000), presents these tasks graphically. The selection of data to be analyzed may involve integration of data from various sources and often requires their formatting to fit the format acceptable to the data mining software. In a mining situation where the objective may be optimization of Komatsu truck performance, data on load carried, on cycle times,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

and on truck component performance may be needed, acquired in different formats from engine monitoring system (say Cummins engine monitoring system), from truck dispatch system (say Modular Mining's Dispatch), and from an on-board weigh



measuring system provided by a third party. Major problem may be faced with making data formats compatible with each other and with that of data mining software to be used. The next step, transforming the data or its pre-processing may involve filtration, discretization, data joining and similar actions. It allows organization of the data so that it may be mined efficiently. In the case of Komatsu truck mentioned above the data joining would be a major task, as would its discretization and filtration.

Mining data is done using one or more of data mining techniques briefly discussed below. It needs to be noted that data mining did not originally relate to mining. It is a general-purpose data processing method that permits discovery of information that may exist in various databases. Interpreting the results is the last and a very important step of data mining. Usually various visualization tools are used in the process, which allow for easy viewing of the information and identification of information discovered during the data mining process.

IV. DATA MINING TECHNIQUES

A number of techniques are used in data mining, each with its own interesting applications. Several methods summarize and describe these techniques classifies data mining techniques as follows.

A. Decision Trees

The decision trees are predictive models that an be viewed as a tree, with tree branches representing a classification question and the leaves representing partitions of the data set with their classification. The prediction is made on the basis of a series of sequential decisions. Thus in case of mining trucks the decision tree could be used to identify which trucks are most likely to fail, and when, based on such questions as: what is the truck make, how old is it, how long it has operated, what is its past repair history, who was its operator and the like. A decision tree model can be confirmed or modified by hand and it can be directed based on the expertise of the person constructing it.

The decision tree models are best used for exploration of the data sets and that of the problem at hand. It is done by looking at the predictors and values that are chosen for each split of the tree. They can also be used for data pre-processing for other prediction algorithms.

B. Neural Networks

Neural networks are computer implementations of sophisticated pattern detection and machine learning algorithms used to build predictive models from large historical databases. They allow for construction of highly accurate predictive models that serve to solve a large number of different problems. The main problem with neural modeling is lack of clarity, the price often paid for their complexity and high accuracy. To overcome this problem, various visualization techniques are used in conjunction with neural models to help explain and control the model.

The primary application of neural models in data mining is clustering, the technique that is used to segment a database into clusters, or sub-sets, based on a set of predetermined attributes. The ability of neural models to perform accurate numerical predictions led to variety of applications, including predictions of the stock markets behavior. As related to a mining truck, neural clustering may be used to define and quantify the relations between various data streams collected on this truck, following by clustering of these streams into mutually dependent groups. Thus, for example, the factors that have an impact on cycle time of the truck can be defined and quantified.

C. Nearest Neighbor and Clustering

Both these techniques are very intuitive and between the first used for data mining. Nearest neighbor prediction algorithms are convenient and simple predictive tools that allow for clear explanation of why a prediction was made. The predictions are based on behavior or properties of the "neighbor" data with the highest weight assigned to the data that is closest. Clustering is grouping, or "clustering' together the data that has the same or similar attributes.

Both clustering and nearest neighbor techniques are between the easiest to use and have a variety of applications. Both are primarily used for prediction of new data rather than extraction of rules from an extensive databases. Using the mine truck example, these techniques appear to be most suited for prediction of when and how this truck will fail, a key piece of information for a mine operator.

D. Genetic Algorithms

Genetic algorithms refer to simulated evolutionary systems that dictate how populations should be formed, evaluated and modified. One of a variety of algorithms known as optimization techniques generic algorithms are in their infancy and more experience with them is required before a mine-related use can be proposed.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Fig. 1. Data Mining Architecture.

There are three tiers of data mining Architecture Data Mining Approach in Security Information and Event Management

1) *Data layer* As mentioned above, data layer can be database and/or data warehouse systems. This layer is an interface for all data sources. Data mining results are stored in data layer so it can be presented to end-user in form of reports or other kind of visualization.

2) *Data mining application layer* This layer is used to retrieve data from database. Some transformation routine can be performed here to transform data into desired format. Then data is processed using various data mining algorithms.

3) *Front-end layer* This layer provides intuitive and friendly user interface for end-user to interact with data mining system. Data mining result presented in visualization form to the user in the front-end layer

E. Data Mining Techniques

There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns.

Following tables gives an idea about various data mining techniques.

Table I: Various Data Mining Technique

	Techniques Name	Function
1	Association	a pattern is discovered based on a relationship of a particular item on other items in the same transaction
2	Classification	Classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics
3	Clustering	Makes meaningful or useful cluster of objects that have similar characteristic using automatic technique.
4	Prediction	Discovers relationship between dependent and independent variables
5	Sequential Patterns	Discover similar patterns in data transaction over a period

V. DATA MINING FOR SECURITY APPLICATIONS

In this section we will understand what is role of data mining in security information & event management system.

Table II: Various Network Security Devices & Their Function

	Device Name	Function
1	AAA system	handles user requests for access to computer resources and, for an enterprise, & provides authentication, authorization, and accounting
2	Web Gateway	Perform URL filtering & block malicious sites, provides proxy function
3	Vulnerability Assessment Tools	Data base server, Find vulnerability in operating system, application
4	Firewall	permit or deny network transmissions based upon a set of rules and is frequently used to protect networks from unauthorized access
5	Data Leakage Prevention system	Systems that enable organizations to reduce the corporate risk of the unintentional disclosure of confidential information.
6	Mail Gateway	Used to detect & prevent spam mails & unwanted software attached in emails
7	Network & Host Intrusion prevention system	monitors network and/or system activities for malicious activities

In today's world every IT administrator staff has to deal with millions of events. These events are generated by various devices for example the staff in Network Operation Center (NOC) has to analyze events generated by networking devices like routers, Switches, load balancer similarly staff in Security Operation Center (SOC) has to analyze events generated by security devices like firewall, IPS, AAA server. The management body of every organization wants administrator to analyze each & every events. This gives burden to staff & likely chances to miss critical events which increase threats to entire organization. A defense in depth strategy (industry best practice) utilizes multiple security devices. Each device has specific function for which they deploy in IT infrastructure. As a part of risk management many organization generally deploy following devices in their IT infrastructure. Following tables list various security devices and their functions. Any organization dealing with all these security devices face problem in monitoring all such events. This force the origination to increase security analysts posts. This huge amount of events creates following problem in any organization.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

1) Security administrator has to manage all devices & analyze the events generated by these devices which increase the work load

2) Efficiency decrease by spending long time in finding false alarms.

In order to reduce the number of security events on any given day to a manageable, actionable list and to automate analysis such that real attacks and intruders can be discerned we should apply data mining technique to all such events. If we want see holistic view for enterprise security then we do mining on all security & network events.

VI. MINING USES OF DATA MINING

The focus of data mining is to discover and define hidden patterns and trends. Once a pattern is defined it can be used in many ways, such as a training input into a neural network or encoded as a rule into an expert system. Traditional applications of data mining include those for monitoring medical bill fraud, marketing with coupons, monitoring credit card transactions etc. While some of this data is used to generate information describing truck performance and condition, most of the collected data remains unused and is not analyzed. Very little of it, if any at all, is used to forecast truck condition or performance into the future. Instead the whole data analysis effort directed on assessment of past performance. Use of data mining techniques for information discovery in this huge database appears to be one of the promising ways to improve performance of many mines. Review of current industrial applications of data mining indicates that there are numerous opportunities for its use in mines. Three most obvious applications are (1) mining equipment condition monitoring and failure prediction, and (2) quantification of and prognostication the mining equipment performance (3) training of equipment operators.

A. Equipment Condition

Data mining techniques of clustering and association appear to be the most promising in defining the relations and associations that may be of interest. On the other hand rule induction and polynomial regression, the latter not discussed here, may be the best techniques to develop the

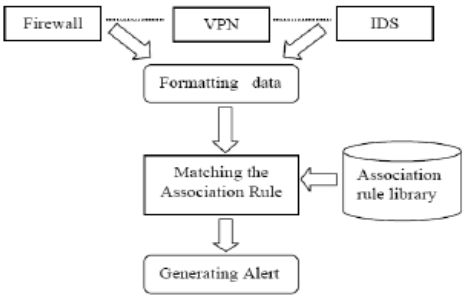
VII. IMPLEMENTATION OVERVIEW

The process of the association analysis can be divided into three parts usually

• Filtering redundant information and formatting the security information.

• Matching the association rules.

• Generating security events.



Framework of the association analysis

VIII. CONCLUSION

This study shows that how data mining can be used in SIEM system [4]. In this work, we introduce the related knowledge, architecture of SIEM system and then the rule of algorithm for the correlation analysis. We have seen various association rules to detect abnormal patterns. One of the areas we are exploring for future research is how we can use other data mining technique like classification, clustering to enhance the system capacity. In addition, we are enhancing the techniques we have mentioned to reduce false positive alerts and to reduce CPU load on system while computing data mining rules. Furthermore we are working to contribute some new modules for open source SIEM project. Modern mines generate huge quantities of data that describe and quantify condition and performance of mine equipment and of the mines themselves. Availability of this data creates a unique opportunity to improve performance of both. Data mining, a set of techniques used to discover hidden relations and trends in large databases, is the likely tool that will permit this to realize this opportunity. The most obvious mining applications of data mining are to prognosticating condition of mining equipment, to prognosticating its performance and to training of equipment operators.

REFERENCES

1. K. R. Rao, "Data Mining and Clustering Techniques," DRTC Workshop on Semantic Web, DRTC, Bangalore, paperk, pp. 1-1, 8th – 10th December, 2003.
2. J. W. Seifert, "Data Mining and Homeland Security: An Overview," CRS Report, pp. 1-1, Jan. 2007.
3. M. S. Chen and J. H. Philip, "Data Mining: An Overview from a Database Perspective," IEEE Trans on knowledge and data engineering, vol. 8, no. 6, pp. 1-1, Dec 1996.
4. S. Yuan and C. Zou, "The Security Operations Center Based on Correlation Analysis."
5. International Business Machines, 2000
6. R. Agrawal and Srikant, "Fast Algorithms for Mining Association Rules," in Proceeding of the 20th VLDB Conference Santiago, 1994
7. J. Han and M. Kamber, "Data Mining Concepts and Techniques. Second Edition," The Morgan Kaufmann Series in Data Management Systems.
8. Rudin, B. Letham, A. S. Aouissi, E. Kogan, and D. Madigan, Sequential Event Prediction with Association Rules.

Mr. Omorogbe Osasu Harry is currently pursuing Ph.D degree in computer science from Ambrose Alli University in Ekpoma, Edo State of Nigeria. Omorogbe Harry. Received Master of Science degree and Bachelor of Science degree from University of Benin, Benin City Nigeria. Currently working as a Lecturer in the department of Computer Science and Information Technology, Igbinedion University, Okada. My area of interest are Web Development, Software Development, Human Computer Interaction (HCI), Computer Networking, Computer Security System, Cyber issues, wireless mobile networking, mobility, resource management and law. Published several national and one international paper.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Dr Asibor Raphael Ehikhuemhen, hails from Uromi in Esan-North East area of Edo State, Nigeria. Holds a Ph.D in Mathematics (Computational Fluid Dynamics). Has his Master's degree from Olabisi Onabanjo University, Ago-Iwoye Ogun State. And a Doctor of Philosophy in Mathematics from Ambrose Alli University, Ekpoma, Edo State, Nigeria. A researcher in Combustion, Electro-osmotic flow, Medical Physics. Haematology, wireless mobile networking, mobility and resource management. A member of IAENG, NAMP, MAN etc. Currently the Chairman, Igbinedion University ICT Committee, Sub-Dean and a member of the institution's Admission.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.