DE GRUYTER
OPEN

Data Mining Usage in Corporate Information Security: Intrusion Detection Applications

Masoud Al Quhtani
Embassy of the Kingdom of Saudi Arabia in Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina

Abstract

Background: The globalization era has brought with it the development of high technology, and therefore new methods of preserving and storing data. New data storing techniques ensure data are stored for longer periods of time, more efficiently and with a higher quality, but also with a higher data abuse risk. **Objective:** The goal of the paper is to provide a review of the data mining applications for the purpose of corporate information security, and intrusion detection in particular. **Methods/approach:** The review was conducted using the systematic analysis of the previously published papers on the usage of data mining in the field of corporate information security. **Results:** This paper demonstrates that the use of data mining applications is extremely useful and has a great importance for establishing corporate information security. Data mining applications are directly related to issues of intrusion detection and privacy protection. **Conclusions:** The most important fact that can be specified based on this study is that corporations can establish a sustainable and efficient data mining system that will ensure privacy and successful protection against unwanted intrusions.

Keywords: data mining, instruction, privacy, corporate information security
JEL classification: O32, O33, G34
Paper type: Research article

Received: July 07, 2016
Accepted: Nov 18, 2016

Citation: Al Quhtani, M. (2017), "Data Mining Usage in Corporate Information Security: Intrusion Detection Applications", Business Systems Research, Vol. 8, No. 1, pp. 51-59.
DOI: 10.1515/bsrj-2017-0005

Introduction

Information security is one of the most important issues in every company. It is important to note that information security is not just a technology issue. This is a business issue as well. In today's high-tech and interconnected world, if companies want to protect information they need well-thought-out security policies. The importance of information security is best illustrated by the fact that "corporate investments on information security are highly evaluated as intangible assets in the stock market especially for IT-oriented firm" (Ishiguro et al., 2006).

Companies have to take special precautionary measures regarding two types of information. The first type includes internal secrets of the company (recipes, projects, models, strategies), and the second type refers to business relationships and contracts with clients. Both types of data are very important for every company and therefore the investment and the establishment of information security is completely justified. The increased data security risk has led to increased obligations of managers. Managerial responsibilities nowadays do not only include improving business operations through quality management decision, but also finding ways to adequately protect information. Results in this segment, has become an integral part of the management engagement, measured in an effort to protect the information assets of the organization (Trompeter et al., 2001). Regarding the importance of studying corporate information security it is important to note that this is a very complex area that does not involve only the use and storage of usernames and passwords. The concept of information security has been significantly expanded compared to an earlier definition that included only a username and password. (Von Solms et al., 2004). At the same time, organizations are threatened from various sides. These raids result in large losses for organizations. These losses include data theft and disclosure of business strategies (Toval et al., 2002). In addition to these requirements, there are also related damages, as well as the fact that the competition becomes familiar with the business strategy of the organization. For all these reasons we can say that any organization that wants to effectively manage its information assets must actively use an information security management system (Kim et al., 2014).

Data mining and privacy are the two opposing goals. Yet, it is possible to satisfy the criterion of privacy and to use data mining. The main tendency in the study of data mining is to develop tools that will ensure private data are protected, and that will enable the efficient use of data mining capabilities. This issue is especially relevant in the field of e-commerce. The problem of data privacy protection is particularly pronounced in e-commerce. Solutions such as the Secure Socket Layer (SSL) can be useful, but in addition to e-commerce, there are many other traps that increase the risk of loss or misuse of personal information (Fienberg, 2006). Although significant efforts have been invested into establishing a system of complete security policy, in the area of e-commerce, a perfect solution still has not been found. It is important to note that privacy protection is guaranteed in separate databases, while integrated databases present a more complex situation.

In recent years, in which the application of computers has increased, the number of intrusions has increased significantly as well, resulting in the need for a system which will identify those intrusions. Intrusion detection is a process based on monitoring the event of a computer or network, recognizing the signs of intrusion and analysing data on which signs of intrusion have been observed (Vigna et al., 1999). Intrusion detection has been present in research since 1980s, but in recent years this area has become an extremely active area of research (Kim et al., 2007).

Information technologies have become a key component of business support, and corporate governance has become unthinkable without the application of data mining. As noted above, the increased use of new technologies has increased the risk of data misuse. The concept of intrusion detection includes events in a computer or network, and analysing and recognizing signs of possible incidents. Signs of possible incidents are based on the recognition of threats to information security policies, violations of acceptable use policies or deviations from standard safety practices (Sayed et al., 2014). In practice, we encounter two types of intrusion detection: anomaly detection and misuse detection. Anomaly detection involves identifying behaviour that deviates from normal behaviour patterns. The detection of misuse can be ensured by the software which uses clearly defined patterns to identify misuse (Chen et al., 2007). In companies that follow modern developments data protection techniques present a priority. The reason for this approach is the fact that data security is of great importance for the company for several reasons. First, it is important to ensure data protection against intrusions of competition that could use certain information to gain a market advantage. Second, data security helps leaving a strong impression of a company that cares about its customers. And finally, by establishing information security management a company protects its own dignity and autonomy that could be endangered by the data abuse. The abovementioned facts are the key reasons for investing into and studying the area of intrusion detection and privacy protection.

Literature review

Data mining

Data mining, or as it is often referred to as knowledge discovery, is the modern process of analysing vast amounts of data and extracting those most important and most relevant. Data mining tools allow the process of data analysis and forecasting of future developments spending much less time and energy compared to using the traditional method. Data mining is an area that encompasses a variety of fields such as technology, databases, statistics, information, artificial intelligence, data visualization and so on (Thamaraiselvi et al., 2004). Given the importance of understanding the term data mining, we will list a few explanations and definitions of the said term referring to the eminent names in the field. One such definition is given and an article published in the International Journal of Computer Trends and Technology (Matatov et al., 2010). From the previous definition it can be concluded that data mining includes a wide range of tools which in a very short time provide very useful and specific information that can form the basis for making managerial or other decisions.

Data mining tools are incredibly practical in everyday use. Data mining is applied in the areas unimaginable to many people, but due to the fact that in a short period of time a huge amount of data is analysed, and that as a result of this analysis, summary data are obtained and used further as the basis for many decisions, data mining has found its application in many areas such as: market segmentation, customer churn, direct marketing, interactive marketing, market basket analysis, trend analysis and others.

Application of data mining in companies is increasingly present. The ultimate goal is profit, and it all starts with the customer information. When companies have more information on customers (their habits and needs) they can provide more value to the customers. The higher the value a company provides to customers, the higher the profit that can be achieved. The best possible data mining application is for achieving that aim. By analysing a large number of simple data on the client conclusions about their behaviour and their needs are created. Based on these data, a company gives them the value they look for. Although there are many areas in which companies can apply data mining, most buyers of these technologies emerge from information intensive industries such as the financial and the marketing sector. These technologies are used by companies that want to take advantage of a large database to enhance their relationships with customers. The main prerequisites for a successful application of data mining are an extensive and well-integrated database, and a well-defined understanding of the business processes within the company. Data mining applications are most commonly used by the following types of companies: marketing companies, pharmaceutical companies, credit card companies, transportation companies and large consumer package goods companies (to improve the sales process to retailers). All these organizations have something in common. They use their knowledge of clients to help them offer more value with lower costs. As a result, these organizations are able to fully define customer needs and create a marketing campaign adjusted to the needs and desires of customers. A common application of data mining in companies is in marketing and in creating a marketing strategy. The reason for this lies in the fact that simple data mining applications get very sophisticated and reliable data that are the key to the future strategy. There are many examples of companies using new technology to get closer to the customer. Banks use extremely intensive data mining applications especially in creating marketing campaigns and tracking marketing campaigns results. Also, the telephone companies are among the most intensive users of new technologies, especially in the creation of appropriate offers to their clients. Offers are created based on previous consumption and customer behaviour. Data mining analysis is based on the use of large amounts of data. The working principle of a data mining application is easy to understand. Such an application analyses valuable data about customers and the results of the analysis are used as a basis for quality decisions that will positively affect the revenue and profits of the company.

The following list contains some of the commonly used data mining software: R, Rattle, SAS Enterprise Miner, Rapid Miner, and Weka. Data mining applications have certain prerequisites for a successful implementation. First, the computer memory must be adequate. Second, the use of data mining applications requires an understanding of marketing problems and certain statistical skills.

Intrusion detection system

Whitman et al. (2010) define computer information security (CIS) as the "protection of information and its critical elements, including systems and hardware that use, store, and transmit that information", with using tools such as policy, awareness, training, education, and technology. CIA triad is a broadly accepted information security model that is based on the confidentiality, integrity and availability of the information (Greene, 2006). Therefore, these are the three mail objectives of CIS. In order to full-fill these objectives, four layers are used: (1) application access layer, (2) infrastructure access layer, (3) physical access layer, and (4) data-in motion-layer. Application access layer is based on the principle that not every user in the corporation should have access to read, write and store all of the data available, but only the data to which she or he is entitled based on the description of their working place (Poirier et al., 2011). Infrastructure access layer is based on the principle that the access infrastructure components, e.g. servers should be protected from both outside and inside intruders (Mlitwa, et al., 2011). Physical layer is based on the principle that the physical access to any system, computer, data, should be available only to the authorized persons. Data-in-motion layer is based on the principle that the data outside the organization (e.g. e-mails, lap-tops, smart-phones) should be also protected according to the same rules, as the data inside the organization. Therefore, corporation information security effectiveness could be defined as the extent to which the corporation achieves to full-fill the goals of confidentiality, integrity and availability of the information, with the means available in application, infrastructure, and physical and data-in-motion access layer (Green, 2006).

In order to achieve CIS effectiveness, organizations typically based their corporation information security measures on two types of efforts: deterrent efforts (Yeh et al., 2007; Herath et al., 2009) and preventive efforts (Willison, 2006). Both efforts are effective on insider threats (Theoharidou, et al., 2006; Dlamini et al., 2009) and intruder threats (Workman, 2007; Viega, 2009).

Malicious intrusions are focused on networks, web clients and servers, databases, and operating systems. There are several types of intruders. Net-work intrusions are intruders who attempt to enter into the networks in order to find and use important information. There are two types of intruders: human intruders or automated malicious software. There are also intruders who are focused on files or on database. Cyber security presents security mechanisms in order to offer explanations which enable prevention for cyber-attacks. The best example is Intrusion Detection Systems (IDS). The purpose of the IDS is to monitor the system activity and inform responsible person. Intrusion detection tools should be strategically located at the network and application levels. However, the main purpose of the intrusion detection tools is to differentiate usual system activity from criminal activity.

Methodology

This paper evaluates the state of data mining applications in corporate information security from 1995 to 2013. It includes data mining applications in corporate information security regarding issues such as: malicious executables, anomalous insiders, intrusion detection.

As a basis for writing this paper we used previous research in this area elaborated in the literature, studies and other relevant sources. In accordance with the practice of scientific investigative work, the following scientific methods have been applied in this article:

Methods of analysis and compilation methods have been used in the theoretical part of the paper for the purpose of analysing and defining data mining multidimensionality, and the importance of data mining in modern corporations.
The method of deduction has been applied in order to be the basis of general conclusions of data mining role in the improving privacy protection and intrusion detection.

Results

In the recent years, intelligent tools have also emerged as one of the important leverages towards CIS (Stoel et al., 2011; Yen, 2007). Data mining has been widely used in establishing CIS (Baesens et al., 2009). These applications employ intelligent tools like neural networks, cluster analysis, nearest neighbors, outlier detection systems and association rules, with the goal of increasing CIS. These intelligent tools are based on the exploration of the data available in vast number of data sources (e.g. transaction data, web logs, databases logs), and could therefore be referred to as exploratory corporate information systems efforts. Number of applications has been developed, that aim towards early warnings on intrusion attempts and frauds (Bose, 2006; Pejic-Bach, 2010). Data mining can be used to monitor large databases, to improve efficiency and quality of data. However, knowledge discovery methods can be also used in order to detect unusual behavior and threats in corporations.

Intrusion detection can give better results when is combined with some data mining techniques, such as a flexible neural tree model (Chen et al., 2007). The model showed significant improvements in intrusion detection because of reducing the input characteristics and because of hybrid approach for combining base classifiers. Data mining techniques can be used to predict intrusion in local area network (Mohammad et al., 2011). Data mining can generate accurate and applicable intrusion patterns from a huge amount of data which means that intrusion detecting system can be used to any logical or network environment (Yu et al., 2007). Chen et al. (2005) investigated how data mining methodologies can be used to detect intrusions of information systems. Their research showed that support vector machine method achieved the best performance, while artificial neural networks with simple frequency-based scheme achieved the worst. Similar research was made also by Zhu et al. (2001). They investigated three different knowledge discoveries in databases methods in detecting network intrusion. Their results showed that data mining method had a important influence on classification accuracy. Among knowledge discoveries in databases techniques, rough sets have higher accuracy followed by neural networks and then inductive learning.

The main cause of threat to privacy is the rapid development of modern technology, which is not accompanied by sufficiently rapid development of the related education and legislation.

However, when we talk about data mining applications and their use in corporations, the facts elaborated in this paper as well as those that can be found in the used literature lead to the conclusion that these applications protect the privacy of customers. Large corporations use data mining in their business practice in the way that they remove data associated with the identification of the customer or client while analysing relevant data (on shopping, habits, etc.). In this way private identification information is completely ignored in the analysis, and only those data which are essential for future management and marketing decisions are used. By using these approach companies foster good relationships with their customers and clients who have confidence in the company and are willing to come back. This relationship is perfect. Why? The company will protect the privacy of the client. The customer will be satisfied, and the company will use analysed data to form an offer that suits the client's wishes. The customer will be satisfied. The client will have confidence in the company. The company will benefit.

For such a system to be sustainable companies have to constantly care for data security. Companies continually need to analyse potential threats and invest in data protection. Companies are aware of the importance of investment into data mining applications and applications for data protection because these investments ensure profit. By investing in the field of contemporary data protection, corporations protect against competition because if the competition came to their data their business would be totally compromised, and customer confidence shaken. It is clear that companies have a huge interest in establishing data mining applications and their use in business because these applications provide incredible opportunities for analysing huge amounts of data critical for making future decisions.

Intrusion detection refers to security management system for computers and networks. This system operates in a way to collect and analyse data from different areas of a computer or a network in order to determine security breaks. A security breach can occur in two different forms: (i) intrusions - attacks from outside the organization; (ii) misuses - attacks from within the organization.

A key feature of this system is using vulnerability assessment when checking the security of computer systems or networks. Intrusion detection functions include: monitoring and examining the user and the system activities, examining system configurations and weaknesses, evaluating the system and the file reliability, capability to recognize patterns typical of attacks, examination of irregular activity forms and following user policy violations.

Intrusion detection has become one of the major tasks in corporate information security because a number of attacks to computer systems have occurred in the last few decades causing significant damage. The intrusion detection system consists of two steps: (i) *Host Intrusion Detection System (HIDS)* can be run on individual hosts or devices on the network. A HIDS observes the inbound and outbound packages from the device in order to warn the user when a questionable action is noticed; (ii) *Network Intrusion Detection System (NIDS)* is located at a strategic point in the network in order to monitor traffic.

Intrusion detection systems (NIDS or HIDS) cannot be considered an unbeatable protection against intrusion, but if used in combination with some of the physical methods such as firewalls or security personnel then they provide much greater security. Firewalls are mostly used to stop disputable traffic.

As we have said, implementing intrusion detection which results in protection from unwanted intrusions is in every company's best interest as it helps ensure privacy protection. It is certainly reasonable to establish an intrusion detection system that will help to achieve higher level of security of the database because the losses from the misuse of valuable and confidential data are much higher than the cost of investment in intrusion prevention. The costs of these investments are generally insignificant compared to the overall volume of business, and therefore managers of modern corporations easily opt for improvement of the protection system.

Conclusion

Modern corporations have recognized the importance of data mining applications and use them in many business areas. Data mining applications are very often used in marketing. There are many marketing campaigns, and even political ones, which are fully designed according to the analysis of specific surveys, questionnaires or other forms of primary data. A marketing campaign created in this way is extremely effective because clients are presented the values they appreciate and companies gain customer confidence finally resulting in successful business operations.

This research was fully based on secondary data and the already available literature. The quality of such research would be much higher if it were accompanied by the presentation of data and information obtained through interviews with some of the eminent personalities in the field of data mining. This approach would be much more complete and much more intelligible to the reader. In the end we can conclude that such research contributes to the understanding and popularization of data mining. After this complex area becomes familiar and understandable, managers can expect a wider application of data mining applications with all the qualities and benefits that they offer.

References

1. Baesens, B., Mues, C., Martens, D., Vanthienen, J. (2009), "50 years of data mining and OR: upcoming trends and challenges", Journal of the Operational Research Society, Vol. 60, pp. S16-S23.

2. Bose, R. (2006), “Intelligent technologies for managing fraud and identity theft“, in Third International Conference on Information Technology: New Generations, (ITNG 2006), IEEE, pp. 446-451.
3. Chen, W. H., Hsu, S. H., Shen, H. P. (2005), “Application of SVM and ANN for intrusion detection“, Computors & Operations Research, Vol. 32 No. 10, pp. 2617-2634.
4. Chen, Y., Abraham, A., Yang, B. (2007), “Hybrid flexible neural-tree-based intrusion detection systems”, International Journal of Intelligent Systems, Vol. 22 No. 4, pp. 337-352.
5. Dlamini, M. T., Eloff, J. H., Eloff, M. M. (2009), “Information security: The moving target“, Computers & Security, Vol. 28 No. 3-4, pp. 189-198.
6. Fienberg, S. E. (2006), “Privacy and confidentiality in an e-commerce world: Data mining, data warehousing, matching and disclosure limitation”, Statistical Science, Vol. 21 No. 2, pp. 143-154.
7. Greene, S. S. (2006). Security Policies and Procedures, New Jersey, Pearson Education.
8. Herath, T., Rao, H. R. (2009), “Protection motivation and deterrence: a framework for security policy compliance in organisations“, European Journal of Information Systems, Vol. 18 No. 2, pp. 106-125.
9. Ishiguro, M., Tanaka, H., Matsuura, K., Murase, I. (2006), “The effect of information security incidents on corporate values in the Japanese stock market”, in International Workshop on the Economics of Securing the Information Infrastructure (WESII).
10. Kim, J., Bentley, P. J., Aickelin, U., Greensmith, J., Tedesco, G., Twycross, J. (2007), “Immune system approaches to intrusion detection - a review”, Natural computing, Vol. 6 No. 4, pp. 413-466.
11. Kim, Y., Chang, H. (2014), “The industrial security management model for SMBs in smart work”, Journal of Intelligent Manufacturing, Vol. 25 No. 2, pp. 319-327.
12. Matatov, N., Rokach, L., Maimon, O. (2010), “Privacy-preserving data mining: A feature set partitioning approach”, Information Sciences, Vol. 180 No. 14, pp. 2696-2720.
13. Mlitwa, N. B. W., Birch, D. (2011), “The role of intrusion detection systems in electronic information security: From the activity theory perspective”, Journal of Engineering, Design and Technology, Vol. 9 No. 3, pp. 296-312.
14. Mohammad, M. N., Sulaiman, N., Abdulkarim Muhsin, O. (2011), “A Novel Intrusion Detection System by using Intelligent Data Mining in Weka Environment“, Procedia Computer Science, Vol. 3 No. 5, pp. 1237-1242.
15. Pejic-Bach, M. (2010), “Profiling Intelligent Systems Applications in Fraud Detection and Prevention: Survey of Research Articles“, 2010 International Conference on Intelligent Systems, Modelling and Simulation (ISMS), IEEE, pp. 80-85.
16. Sayed, M., Jradi, F. (2014), “Biometrics: Effectiveness and Applications within the Blended Learning Environment”, Computer Engineering and Intelligent Systems, Vol. 5 No. 5, pp. 1-8.
17. Stoel, M. D., Muhanna, W. A. (2011), “IT internal control weaknesses and firm performance: An organizational liability lens“, International Journal of Accounting Information Systems, Vol. 12 No. 4, pp. 280-304.
18. Thamaraiselvi, G., Kaliammal, A. (2004), “Data mining: concepts and techniques”, SRELS Journal of Information Management, Vol. 41 No. 4, pp. 339-348.

19. Theoharidou, M., Kokolakis, S., Karyda, M., Kiountouzis, E. (2005), "The insider threat to information systems and the effectiveness of ISO17799", Computers & Security, Vol. 24 No. 6, pp. 472-484.
20. Toval, A., Nicolás, J., Moros, B., García, F. (2002), "Requirements reuse for improving information systems security: a practitioner's approach", Requirements Engineering, Vol. 6 No. 4, pp. 205-219.
21. Trompeter, C. M., Eloff, J. H. P. (2001), "A framework for the implementation of socio-ethical controls in information security", Computers & Security, Vol. 20 No. 5, pp. 384-391.
22. Viega, J. (2009), "Cloud computing and the common man", Computer, Vol. 42 No. 8, pp. 106-108.
23. Vigna, G., Kemmerer, R. A. (1999), "NetSTAT: A network-based intrusion detection system", Journal of computer security, Vol. 7 No. 1, pp. 37-71.
24. Von Solms, B., Von Solms, R. (2004), "The 10 deadly sins of information security management", Computers & Security, Vol. 23 No. 5, pp. 371-376.
25. Whitman, M. E., Mattord, H. J. (2010). Principles of information security, Cengage Learning.
26. Willison, R. (2006), "Understanding the perpetration of employee computer crime in the organisational context", Information and organization, Vol. 16 No. 4, pp. 304-324.
27. Workman, M. (2007), "Gaining access with social engineering: An empirical study of the threat", Information Systems Security, Vol. 16 No. 6, pp. 315-331.
28. Yeh, Q. J., Chang, A. J. T. (2007), "Threats and countermeasures for information system security: A cross-industry study", Information & Management, Vol. 44 No. 5, pp. 480-491.
29. Yen, E. C. (2007), "Warning signals for potential accounting frauds in blue chip companies - An application of adaptive resonance theory", Information Sciences, Vol. 177 No. 20, pp. 4515-4525.
30. Yu, Z., Tsai, J. J. P. (2007), "An Automatically Tuning Intrusion Detection System", IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol. 37 No. 2, pp. 373-384.
31. Zhu, D., Premkumar, G., Zhang, X., Chu, C-H. (2001), "Data Mining for Network Intrusion Detection: A Comparison of Alternative Methods", Decision Sciences, Vol. 32 No. 4, pp. 635-660.

About the author

Masoud G. Alquhtani graduated from the Faculty of Business King Saud University in Riyadh, Kingdom of Saudi Arabia (KSA) with master degree in public administration. He is currently working in Kingdom of Saudi Arabia embassy in Sarajevo as Ambassador Assistance, and is currently a doctoral student at Faculty of Economic and Business, Sarajevo, Bosnia and Herzegovina. His main research is corporate information security in developing countries. Author can be contacted at **smjm2007@hotmail.com**