PROCEEDINGS OF THE NATIONAL SEMINAR

'NETWORK SECURITY'

February 24th and 25th 2016

Organised by

DEPARTMENT OF COMPUTER SCIENCE
EMEA COLLEGE OF ARTS AND SCIENCE
(Affiliated to University of Calicut)
Kondotti, Kerala-673638

Sponsored

University Grants Commission (UGC)
New Delhi

CONTENTS

Application of Data Mining Techniques on Network Security

O.Jamsheela, Department of Computer Science, EMEA College of Arts and Science Kondotty, Malappuram, Email: ojamshi@gmail.com

Abstract—Data mining techniques are extensively used in every fields such as business, marketing, bioinformatics, science and so on. The major application of data mining is on education, scientific and engineering, health-care, business, network security and many more. Network security is an important research area. Each day hackers are using new techniques to unlock the security systems. In this paper a short survey about the application of data mining on the network security is conducted. Many research papers have been discussed the topic elaborately. So in a single paper all the contributions cannot be included. Here only selected papers have been analyzed to get an overall idea about the application of data mining to control the threats on network security.

I. INTRODUCTION

The focus of this paper is to analyze the efficiency of the application of the Data Mining algorithms on the net- work security. The remaining of the paper is organized as follows. Section 1.A introduces the Network Security. Section 1.B presents a brief introduction of Data Mining. Section II contains a literature survey and Section lll concludes the paper. A. Network Security Almost all network offers high efficiency and public service for people. But anxieties about network security is a major issue. Recently more and more network security threats are arose on every kind of networks and became an emerging topic in research field. Some examples of harmful network threats are spreading computer viruses, spams on mailboxes, hacking passwords and other information, harmful pages are sending through mails etc. Although network security is a crucial aspect and many contributions are introduced to prevent it, still new security threats are harming the network.

B. Data Mining

Data mining is the process of discovering previously unknown and useful information from large databases. The most widely used data mining technologies include associa- tion rules discovery, clustering, classification, and sequential pattern mining. Among them, the most popular technology is association rules discovery, which is mining the possibility of simultaneous occurrence of items, and then building rela- tionships among them in databases. Association rules mining can be divided into two parts: Find all frequent itemsets, and generate reliable association rules directly from all frequent itemsets. Because frequent itemsets mining is the most time- consuming procedure, it plays an essential role in data mining and knowledge discovery techniques, such as association rules, classification and clustering. In 1993, Agrawal et al. [2] First proposed the problem of finding frequent itemsets in their association rule mining model. A large number of studies have been published introducing new algorithms or improvements on existing algorithms to solve the frequent pattern mining problem more eficiently. Data mining methods are widely used in diverse areas. Here the application of data mining methods on network security is discussed.

II. LITERATURE REVIEW

Many papers have been published based on the application of data mining technique on network security. Patil et a.l [18] presented a paper to find new evasion techniques on network intrusion detection system. In the paper the authors states that the objective of an attacker is to find out new evasion techniques to stay unseen. Unfortunately, majority of the existing techniques are based on the ambiguities of the network protocols. The idea of the paper is to develop a network based intrusion detection system based on Apriori algorithm, a popular data mining technique. The work is focused on misuse detection. Normally attack signatures are collected and stored in a database in the same way as virus protection software does in order to detect the related attacks. The authors states that signature based Network Intrusion Detection Systems (NIDS) are effective at detecting attacks for what they are prepared. Firewalls do not normally block packets, but make aware about the intrusion alarm. This situation causes attackers to focus their efforts in finding evasions over the signatures of these systems. The overall idea of intruder is to perform some changes to cause evasions that the Signature based NIDS does not process the entire attack packet, which remains undetected. An evasion succeeds if the processing of the packets generates a different representation of the raw data in the Signature based Network Intrusion Detection Systems and in the end systems. Data contained in TCP segments can encapsulate some attacks, but in some situations, it will not be able to detect those attacks.

The authors is applied a method to look for new evasive techniques by analyzing NIDS behavior. In this method first build NIDS using C4.5 algorithm. Publicly available dataset KDD-99 is used. AdaBoost algorithm for supervised learning is applied where labeling of dataset is done as normal or attack. Modified Apriori algorithm generates rules which are checked on snort for evasion. To compare the results other methods like Genetic Algorithm are used.

Intrusion detection systems using data mining approachesmake it possible to search patterns and rules in large amount of audit data. Kamini Nalavade and B.B. Meshram [14] have presented a model to integrate association rules to intrusion de- tection to design and implement a network intrusion detection system. Their technique is used to generate attack rules that will detect the attacks in network audit data using anomaly detection. They have proved that the modified association rules algorithm is capable of detecting network intrusions. The authors proposed a network intrusion detection and prevention system model that analyses the various item set generated, specifically on attribute relation. In this model they have applied association rule mining to generate attack signatures from the network trafic data.

Abdelzaher et al. [1] have presented a survey paper on the application of Data mining methods to diagnose sensor network bugs. A Survey of Outlier Detection Methods in Network Anomaly Identification is presented by Gogoi et al. [11]. Outliers arise due to various reasons such as mechanicalfaults, changes in system behavior, fraudulent behavior, humanerror and instrument error [11]. The outlier detection leadsto more interesting and useful results such as identificationof system faults, network hackers etc. The administrators can take preventive actions before the abnormalities actually attackthe network. Although, outliers are considered noise or errors,they may have important information [11]. The authors haveanalyzed different methods used to detect outlines. The paper is concluded by stating that the notion of outlier is different for different application domains. The authors also states that the development of an effective outlier detection technique for mixed-type and evolving network traffic data, especially in the presence of noise, is a challenging task. They suggested that outlier detection method should be tested on real network data collected using tools such as flow-tools [19] and dataset like the MITRE [3] data set.

Finding the root-cause of a network security anomaly is essential for network operators [17].

Paredes et al. [16] introduced a technique that uses frequent itemset mining to automatically extract and summarize the traffic rows causing an anomaly in networks. Paredes et al. [17] presented a demonstration, by introducing an open-source anomaly-extraction system based on their technique to extract the trafic rows on networks. The authors have integrated the their method with a commercial anomaly detector and applied in a real network. They have reported a number of detected security anomalies and illustrated how an operator can use the system to automatically extract and summarize anomalous rows.

The article "Data Mining Algorithms for Communication Networks Control: Concepts, Survey and Guidelines" pre- sented by De Sanctis et al. [7] identifies the concepts behind the idea of using data mining for communication network control, provides a structured survey of the results in this area, and discusses the guidelines for future applications. The authors states that data mining algorithms are eficient for optimized network control compared with other methods when the network behavior is complex and changes frequently over time.

III. CONCLUSION

With the rapid development of Internet, network security issues also have been increased. Intrusion and anomalies detection can prevent network threats and can greatly improve the network security. Based on the extensive literature survey done in previous section, various data mining algorithms have been suggested to control network threats. Various authors have discussed various application of data mining in the field of network anomalies. We can conclude that data mining techniques have been extensively applied in the area of net- work security and an important application of data mining in network security is the intrusion detection. It is hoped that the survey done in this paper be helpful to researchers working in the area of application of data mining methods in network security.

REFERENCES

[1] Tarek Abdelzaher and Jiawei Han. A survey of datamin- ing methods for sensor network bug diagnosis. In Man- aging and Mining Sensor Data, pages 429–458. Springer, 2013.

[2] Rakesh Agrawal, Tomasz Imieli ´ nski, and Arun Swami. Mining association rules between sets of items in large databases. In ACM SIGMOD Record, volume 22, pages 207–216. ACM, 1993.

[3] SJ Aguirre and WH Hill. Intrusion detection Why- off: Implications for the united states navy, september 1997. Technical report, MITRE Technical Report MTR 97W096, McLean, Virginia.

[4] M Ali Aydin, A Halim Zaim, and K G¨ okhan Ceylan. A hybrid intrusion detection system design for computer network security. Computers & Electrical Engineering, 35(3):517–526, 2009.

[5] Tim Bass. Intrusion detection systems and multisensor data fusion. Communications of the ACM, 43(4):99–105, 2000.

[6] Eric Bloedorn, Alan D Christiansen, William Hill, Clement Skorupka, Lisa M Talbot, and Jonathan Tivel. Data mining for network intrusion detection: How to get started. Technical report, Citeseer, 2001.

[7] Mauro De Sanctis, Igor Bisio, and Giuseppe Araniti. Data mining algorithms for communication networks control: concepts, survey and guidelines. IEEE Network, 30(1):24–29, 2016.

[8] John E Dickerson and Julie A Dickerson. Fuzzy network pro?ling for intrusion detection. In Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American, pages 301–306. IEEE, 2000.

[9] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, and Lilly Suriani Affendey. Intrusion detection using data mining techniques. In Infor-

mation Retrieval & Knowl- edge Management, (CAMP), 2010 International Confer- ence on, pages 200–203. IEEE, 2010.

[10] Pedro Garcia-Teodoro, J Diaz-Verdejo, Gabriel Maci´ a- Fern´ andez, and Enrique V´ azquez. Anomaly-based net- work intrusion detection: Techniques, systems and chal- lenges. computers & security, 28(1):18–28, 2009.

[11] Prasanta Gogoi, DK Bhattacharyya, Bhogeswar Borah, and Jugal K Kalita. A survey of outlier detection meth- ods in network anomaly identification. The Computer Journal, page bxr026, 2011.

[12] Dina Hadiosmanovic, Damiano Bolzoni, Pieter Hartel, and Sandro Etalle. Melissa: Towards automated detectionof undesirable user actions in critical infrastructures. In Computer Network Defense (EC2ND), 2011 Seventh European Conference on, pages 41–48. IEEE, 2011.

[13] Wenke Lee, Salvatore J Stolfo, Philip K Chan, Eleazar Eskin, Wei Fan, Matthew Miller, Shlomo Hershkop, and Junxin Zhang. Real time data mining-based intrusion de- tection. In DARPA Information Survivability Conference & Exposition II, 2001. DISCEX'01. Proceedings, volume 1, pages 89–100. IEEE, 2001.

[14] Kamini Nalavade and BB Meshram. Mining association rules to evade network intrusion in network audit data. International Journal of Advanced Computer Research, 4(2):560, 2014.

[15] Mrutyunjaya Panda and Manas Ranjan Patra. A compara- tive study of data mining algorithms for network intrusion detection. In 2008 First International Conference on Emerging Trends in Engineering and Technology, pages 504–507. IEEE, 2008.

[16] I Paredes-Oliva, P Barlet-Ros, and M Molina. Automatic validation and evidence collection of security related network anomalies. Proc. of PAM (Poster Session), 2010. [17] Ignasi Paredes-Oliva, Xenofontas Dimitropoulos, Maur- izio Molina, Pere Barlet-Ros, and Daniela Brauckhoff. Automating root-cause analysis of network anomalies using frequent itemset mining. In ACM SIGCOMM Computer Communication Review, volume 40, pages 467–468. ACM, 2010.

[18] RUTUJA R PATIL and PR DEVALE. To find new eva- sion techniques on network intrusion detection system.

[19] Stuart Staniford, James A Hoagland, and Joseph M McAlerney. Practical automated detection of stealthy portscans. Journal of Computer Security, 10(1-2):105– 136, 2002.

[20] Rebecca Wright and Zhiqiang Yang. Privacy-reserving bayesian network structure computation on distributed heterogeneous data. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discov- ery and data mining, pages 713–718. ACM, 2004.

[21] Nong Ye et al. The handbook of data mining, vol- ume 24. Lawrence Erlbaum Associates, Publishers Mah- wah, NJ/London, 2003.