ISSN No. 0976-5697

Volume 3, No. 2, March-April 2012

International Journal of Advanced Research in Computer Science

TECHNICAL NOTE

Available Online at www.ijarcs.info

Emerging Trends in Data Mining for Intrusion Detection

Alok Ranjan*
Research Scholar, Department of Computer Science,
Rayalaseema University, Kurnool, INDIA
mibalok07@gmail.com

Dr. Ravindra S. Hegadi
Assistant Professor, Department of Computer Science,
Karnataka University, Dharwad, Karnataka, INDIA
ravindrahegadi@rediffmail.com

Prasanna Kumara
Research Scholar, Department of Computer Science,
Rayalaseema University, Kurnool, INDIA
prasanna.s_kumara@yahoo.com

Abstract: IDS is the technique to detect the Intrusion to the database over the network. The IDS analyzes the data resources using the data mining technique like Association Analysis and Clustering Analysis to extract the useful patterns and rules. The Intrusion Detection System is facing great challenges like unauthorized use of data, data modification from intruders etc. The applications of data mining technique in computer security field improve the development of IDS. It is necessary to classify the degree of attacks in IDS and use it in IDS by data mining. IDS can detect the attack activities in network, which result in uncertainty. This paper will help network Security Analyst to determine intrusions

Keywords: Data mining, Intrusion Detection System, Attacks, Config file.

I. INTRODUCTION

Nowadays internet usages are providing more convenient and easier life. However, there are some serious security problems like hackers for remote attacking, computer viruses etc, which spread very quickly across the network. According to statistical data, there are many attacks to invade some important commercial websites. Nowadays, we rely so much on the information system, but the network invasion is quite a threat to the safety of it.

Data mining technology is useful for some fields, such as commerce, finance medical etc. The data, information message of internet and the audit information resides in the server which remain in server and work station become analytical object of data mining technology, and the count is very huge because the count of them are very large. Data mining applied into intrusion detection can generate many concise and exact detection modes automatically from a great deal of audit data [1].

II. INTRUSION DETECTION

Intrusion Detection is a process of monitoring and analyzing the activities in the computer system. The main objective is to identify the threats to the system, and then to protect and safeguard the system from those threats. Some of the technologies have has been verified and applied [2]. The strategies of intrusion detection are being divided into two categories:

(i) The Misapplication Detection Strategy
(ii) The Abnormal Detection Strategy.

Both strategies have their own advantages and disadvantages. In Misapplication Detection Model, every entity in the data collection is marked as normal or invasive. We construct invasion model through machine study or the analytical technology of data mining from these data. This invasive model will bring rules and classify the given data according to the characteristics of the sample data. The advantage of adopting misapplication detection model is that the detection of known attack behavior is very effective. At the same time, invasive model has a favorable openness and it can use the data, which has new mark as training collection [3, 4]. Then add it to invasion model after learning, so that it includes new styles of attack. The disadvantage of invasion model is that the accuracy of data log influences the rationalization of the model very much, and incorrect log data will lead to the increased rate of misreport. Besides, the collect behavior of these log data also is quite a difficult task. The major limitation of misapplication detection model is that it cannot distinguish the attack behavior of unknown style because it just includes the description of known attack actions. Before the invasion behavior added to detection model, the system is totally open to this kind of attacks, and to some extent it is unsafe. At present, the abnormal detection model is the most widely used intrusion detection measure, and still there are lots of researches going on in this field.

III. DATA MINING TECHNIQUE IN IDS

Data Mining is also called Knowledge Discovery in Database. These data are usually numerous, incomplete, indistinct and random. Data as original information can also be called knowledge. In general, knowledge refers to concepts, rules and restraints. The method of finding knowledge can be mathematical or non-mathematical. It can be deductive or inductive. The known knowledge can be used to optimize enquiry, manage information, control progress and make intellectual decision. Data mining is a criss-cross subject. It helps people to apply data from low and simple inquiry to discover knowledge in data and support decision [5].

The main techniques of data mining are association analysis, clustering analysis, classification, prediction, time-Series Patterns and bias analysis. Just in a decade, data mining technology has become popular and active research measures. In the analysis of intrusion detection system, the data circulating in network has the following characteristics:

a. Mass data, even if a small commercial website,
b. The number of data message sent and received are quite impressive and incomplete whose transportation is busy.

Data message that overweigh network carry will be discarded when network is unstable. Data information may get changed in the transportation of message. We can see that these data is in accordance with the feature of object of data mining [6]. We want to apply data mining technology in the proposed Intrusion Detection System. Figure-1 describes application of IDS with data mining. Both the Abnormal Detection Model and Misapplication Detection Model need to log data as training data whose accuracy will largely influence Intrusion Detection System. Because of density and accuracy of visit network, it is difficult to acquire completely the attack actions. In addition, it is also uneasy to log the attack behaviour. Data mining techniques can reveal this problem.

The analysis of general network visit and isolated point is an invasive behaviour which reduces the difficulty of acquisition of training data. Normally, we need to decide which feature is effective to express network and system action when we build intrusion detection system. Some characteristics can be extracted directly from audit data; others need to be taken from the statistics and are computed using the audit data, such as the quantity in TCP message in certain period of time [7]. It is difficult to make these decisions, with the use of data mining techniques.

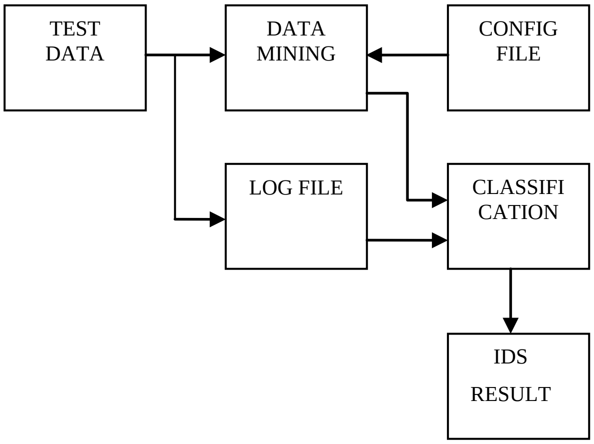

Figure 1: IDS with Data Mining

Intrusion Detection System is a passive method. It monitors information system and sends out warning when it detects intrusion, but data mining techniques can analyse these data when network information is acquired, it can forecast the visits on its own initiative, reduce the frequency of matching, and thus achieving the function of active defence. Data Mining techniques, for instance, Clustering, Classification, Feature Summary, and Association Rules can be applied in the Intrusion Detection System [8, 9]. It has been proved that data mining techniques improves the property of intrusion detection system, the processing rate and reduces the rate of misreporting.

IV. THE DISTINCTION OF ATTACKS

Traditional Abnormal Detection Model uses the normal data, which are used to extract the information to build the model during the training process [10]. The Abnormal Detection Model builds the model based on the normal visiting behavior of intruders. The main advantage of the Abnormal Detection Model is to detect any possible attacks, which are beyond the detecting capacity of the Abnormal Detection Architecture, and it supports the function of Invasion Detection System, but it has its own limitations:

i. The Abnormal Detection Model needs well-defined collection of data sets for training the system.
ii. It is difficult to include all normal behaviors. When invade behavior does not match the normal behavior described by the model. The abnormal detection model will alarm because this action has the feature of invasion. However, this action is normal visit just probably because customers have adjusted to the orders.

It is difficult to describe one’s behavior i.e. normal or invasive in an Intrusion Detection System, and most of situations are uncertain [11]. The majority of the network behaviors may be either normal or invasive. Under these circumstances, the Intrusion Detection System cannot make sound judgment. Thus, the rate of uncertainty attack is high. In order to resolve this problem, the Intrusion Detection System needs to deal with this kind of uncertainty. Thus, the classification of any types of attacks always depends on log file information and the techniques of data mining [12]. This model is always the part of Intrusion Detection System. Intrusion Detection System is a passive method. IDS monitor the information system (data) and when there is an intrusion in an manner, it it will send the signal.

V. CONCLUSION

The application of Data Mining in Intrusion Detection System is emerging trend in the recent years. The Data Mining techniques can extract characteristics of sample data, thus reduces the difficulties involved in the collection of training data. Thereby achieving the active defence for Intrusion Detection System. The traditional Intrusion Detection System cannot do all of these. It is necessary to describe this indeterminacy because the data of network traffic and host audit and the detective process of Intrusion Detection System are indeterminable. This paper describes the distinction of attack degree due to above reason.

VI. FUTURE SCOPE

The Data Mining plays a major role in wide variety of its application areas. The sequence representation of data in network traffic is uncertain. There is limitation in the application of intrusion detection technology. The flexibility of system is not good to analyze the huge amount of data based upon proposed method. Still there is scope for research in this area.

VII. REFERENCES

[1] Christine Dartigue, Hyun Ik Jang, and Wenjun Zeng: "A New Data-Mining Based Approach for Network Intrusion Detection", Seventh Annual Communication Networks and Services Research Conference. pp. 372–377, May 2009.

[2] Liu Wenjun: "An Security Model: Data Mining and Intrusion Detection",2nd International Conference on Industrial and Information Systems .PP. 448-450, July 2010

[3] Lee W, Salvatore "Data Mining Approaches for Intrusion Detection." Department. New York, NY, Computer Science Department, Columbia University. DARPA & NSF Reasearch

[4] Bhavani Thuraisingham, Latifur Khan, Mohammad M. Masud, Kevin W. Hamlen: "Data Mining for Security Applications"- 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing,

[5] ZHAN Jiuhua: "Intrusion Detection System Based on Data Mining"-Proceedings of the First International Workshop on Knowledge Discovery and Data Mining IEEE Computer Society Washington, DC, USA 2008

[6] Daniel Barbara, NingningWu, and Sushil Jajodia. "Detecting novel network intrusions using bayes estimators. In Proceedings of First SIAM"- Conference on Data Mining, Chicago, IL, 2001.

[7] Eskin, E., Arnold, A. et. al. A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabelled Data, http://www.cs.columbia.edu/ ~eeskin/.

[8] Xinzhou Qin et.al. "Using MIB II Variables for Network Anomaly Detection- A Feasibility Study" International Symposia on Computer Network and Multimedia Technology (CNMT), 2009.

[9] Noel, S., Wijesekera, D., and Youman, C. "Modern intrusion detection, data mining, and degrees of attack guilt. Applications of Data Mining in Computer Security," edited by D. Barbar'a and Sajodia, Kluwer Academic Publishers, 2002.

[10] Dain O, Cunningham R. "Fusing a heterogeneous alert stream into scenarios." In: Proc. of ACM Workshop on Data Mining for Security implications, 2001.

[11] Klaus Julish: "Data mining for intrusion detection" a critical review, Switzerland: IBM Research, Zurich Research Laboratory, 2001.

[12] CHEN Xun-xun, FANG Bin-xing: "Optimizing of large-number-patterns string matching algorithms based on definite-state automata." Journal of Harbin Institute of Technology, issue 2, 2007.