Far East Journal of Electronics and Communications
SASTRA Nat. Conf. Elec. Sci. Tech., 2015

http://dx.doi.org/10.17654/EC03010133
Special Issue (March, 2016), Pages 133-139 **ISSN: 0973-7006**

EFFICIENT ATTRIBUTES IDENTIFICATION PRACTICE ON INTRUSION DETECTION SYSTEM DATASET THROUGH PREDICTION MECHANISMS

R. Bala Krishnan and N. R. Raajan

Department of CSE
Srinivasa Ramanujan Centre (SRC)
SASTRA University
Kumbakonam 612001, India
e-mail: balakrishnan@src.sastra.edu

School of Electrical and Electronics Engineering
SASTRA University
Thanjavur 613401, India

Abstract

The growth of internet applications makes the intrusions on the networking system at a high level. In such a situation, it is essential to offer security to the network systems by an efficient intrusion detection and avoidance mechanisms (IDS and IPS). This can be accomplished by formulating an efficient intrusion detecting system that finds effective attributes of the IDS dataset and can distinguish the abnormal and normal activities in the network. Many intrusion detection system applications have been proposed in the past and the systems are having limitations in terms of detection accuracy. To overcome these limitations, the proposed research work offers a new

Received: October 23, 2015; Revised: December 11, 2015; Accepted: December 31, 2015
Keywords and phrases: intrusion detection system, prediction, SVM, Bayesian, anomaly detection, misuse detection.
Communicated by A. Srinivasan, Guest Editor

scheme of identifying the suitable attributes for the IDS by prediction mechanisms such as Bayesian classification and support vector machine. The experiments are conducted using KDD CUP'99 dataset and the observation shows that the proposed approach reduces false positive alarms, detects suitable attributes for the intrusion detection, precisely and accurately. Hence, the overall accuracy detection has been improved.

I. Introduction

The network based services and information over the networks attained a tremendous growth and the services over the network become more significant than ever before. The intrusion detection mechanisms are the preceding stripe of fortifications against computer attacks in the wake of secure network architecture design and firewalls. In spite of the embarrassment of available intrusion prevention mechanisms, attacks beside computer systems are still unbeaten. Thus, the intrusion detection systems (IDSs) participate imperatively in network security [1]. This paper attempts to address the issue of effective attribute detection for identification of false positive attacks in network systems, which is attempted from various sources [2]. The role of data mining in the intrusion detection is used to spot the user actions and extract it from the collected knowledge, so that it is easy for the analysts to focus on real attacks. The approaches of data mining such as signature based and anomaly based detection are used to identify the intrusive attacks, which are sufficient to identify the false positive attacks and it is one of the major issues. IDS incorporates the two popular methods for detection processes, namely misuse detection and anomaly detection. Anomaly detection scheme follows a strategy in its methodology which is finding out the divergence of normal and abnormal user behaviors [3-5]. Misuse detection deals with identifying the user activities on the basis of its access rights and from which it identifies the legal [6] and illegal activities. One of the most significant issues is false positives (FPs) [5]. The outcome from the IDS application is the FP, which informs about a user that it is unauthorized penetration or attack although it is authorized.

The rest of the paper is summarized as follows: Section II presents related works on intrusion detection systems, Section III states the proposed intrusion detection system (IDS) models. The experimental observations of the proposed application are stated in Section IV and the conclusion of the work in Section V.

II. Related Works

The intrusion detection system is implemented with several approaches such as machine learning and statistical data mining. These approaches proceed with the content from the existing knowledge, which is collected from various domains. This session describes some of the famous approaches which are given below that are cusps in the development of intrusion detection system applications.

Data mining (DM). It holds a group of proficiencies that use the practices of evoking formerly unknown but likely practicable data from large amount of data stores. In other words, this technique allows finding regularities and irregularities in huge input data sets. However, they are memory exhaustive and have need of double storage: one for the normal IDS data and another for the data mining [7]. The part of data mining for intrusion detection is to categorize typical user activity and extort it from the composed data so that it is painless for the analysts to spotlight on real attacks, to discover false alarm generators and terrible sensor signatures and also to locate inconsistent activity that uncovers a genuine attack [9].

Data mining often involves four courses of assignment such as association rule learning, classification, clustering and regression. This advanced approach such as machine learning allows computers to learn from data such as from sensor data or system databases [8]. A major spotlight of machine learning research is to robotically learn to identify complex patterns and make intellectual decisions on the basis of the data. A pattern matching technique principally looks for a definite attack pattern [1], which may be given in the record that is audited previously. A human knowledge is mandatory to spot and extort inconsistent elements or patterns from input

data [3]. This approach is effectual in reducing the need of maintaining large amount of audit data [4]. This approach is also unable to detect new attack signatures. Traditional IDSs with expert systems hold the statistical contents such as the application users, workstations and programs and by regular monitoring of user activities, it identifies the abnormal behavior and it results in the detection of intruders. The major drawback of the expert systems based IDS is that it needs frequent updates by a system administrator.

III. Proposed System

The workflow of the proposed system is presented in Figure 1. It works with the KDD CUP'99 dataset and it is classified into training and testing data.

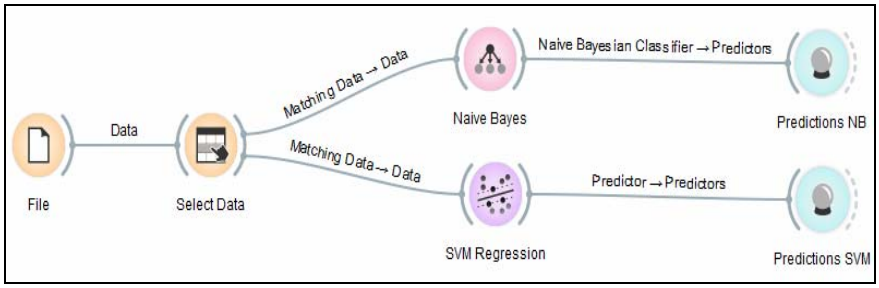

Figure 1. Abstract process of the proposed mechanism.

The training dataset is subjected to construction phase, in which the classifier is being build and is updated in the knowledge base for Bayesian classifier and support vector machine [9, 2]. The construction of the classifier is based on the attributes of the data set. The outcome from the prediction mechanism at the execution phase is either to accept or to reject the request. As per the available trained knowledge, the mechanism classifies the request as an intrusive attack or a normal request. Hence the accuracy measure of the classifiers purely depends on the attributes of the dataset. The KDD CUP'99 dataset has 42 features [5] and are stated in Figure 2. The proposed scheme offers five methodologies with different attributes of the dataset for identifying the efficiency of the attributes under the prediction measure. The methodologies and the serial number of the attributes are stated in Table 1.

S.No	Name	Type	S.No	Name	Type
1	Duration	Continuous	22	Is_guest_login	Discrete
2	Protocol_type	Discrete	23	Count	Continuous
3	Service	Discrete	24	Srv_count	Continuous
4	Flag	Discrete	25	Serror_rate	Continuous
5	Src_bytes	Continuous	26	Srv_serror_rate	Continuous
6	Dst_bytes	Continuous	27	Rerror_rate	Continuous
7	Land	Discrete	28	Srv_rerror_rate	Continuous
8	Wrong_fragment	Continuous	29	Same_srv_rate	Continuous
9	Urgent	Continuous	30	Diff_srv_rate	Continuous
10	Hot	Continuous	31	Srv_diff_host_rate	Continuous
11	Num_failed_logins	Continuous	32	Dst_host_count	Continuous
12	Logged_in	Discrete	33	Dst_host_srv_count	Continuous
13	Num_compromised	Continuous	34	Dst_host_same_srv_rate	Continuous
14	Root_shell	Continuous	35	Dst_host_diff_srv_rate	Continuous
15	Su_attempted	Continuous	36	Dst_host_same_src_port_rate	Continuous
16	Num_root	Continuous	37	Dst_host_srv_diff_host_rate	Continuous
17	Num_file_creations	Continuous	38	Dst_host_serror_rate	Continuous
18	Num_shells	Continuous	39	Dst_host_srv_serror_rate	Continuous
19	Num_access_files	Continuous	40	Dst_host_rerror_rate	Continuous
20	Num_outbound_cmds	Continuous	41	Dst_host_srv_rerror_rate	Continuous
21	Is_host_login	Discrete	42	Normal or Attack	Discrete

Figure 2. KDD CUP'99 dataset format (value, name and its type).

Table 1. Attributes of dataset with methodology

S.no. (Attributes)	**Methodology**
1, 3, 4, 11, 14, 15, 16	Methodology 1
1, 2, 28, 31, 34, 35, 36	Methodology 2
2, 3, 5, 6, 11, 29, 32	Methodology 3
1, 8, 11, 12, 25, 29, 38	Methodology 4
1, 2, 3, 4, 5, 6, 7	Methodology 5

IV. Experimental Observations

The proposed model with five methodologies has been implemented using Orange 2.6 as data miner software. The performance of the algorithm is assessed on a Pentium Core 2 Duo System of 2.6GHz with 2GB RAM on Windows 8 platform. The methodologies generate the prediction outcome for various attributes using the Bayesian and support vector machine classifier. As for the proposed system testing, 10 percent of the total data is taken as training data for the construction phase and 10 percent of total data is taken as test data (with corrected labels) for the execution phase. The outcomes of the methodologies are stated in Table 2. From the observation outcomes, it is stated that the Methodology 3 with attributes Protocol_type, Service, Src_bytes, Dst_bytes, Num_failed_logins, Same_srv_rate and Dst_host_ count (2, 3, 5, 6, 11, 29, 32) offers an efficient attribute set on the prediction measure.

Table 2. Methodologies and its prediction measure accuracy

S.no.	Methodology	Accuracy percentage	
		Bayesian classifier	Support vector machine
1	Methodology 1	77.31%	84.71%
2	Methodology 2	68.1%	89.7%
3	Methodology 3	91.32%	95.03%
4	Methodology 4	87.14%	88.12%
5	Methodology 5	93.68%	91.33%

V. Conclusion

In this work, an efficient attribute identification scheme for the intrusion detection system has been proposed. From the experimental results, it is found that the proposed model identifies the efficient attributes for the prediction measure and proves that the accuracy detection of the classifiers depends on the attributes. The proposed work could be planned to upgrade with temporal models and it could be the future direction towards the progress of this work.

Acknowledgement

The authors thank the anonymous referees for their valuable suggestions which let to the improvement of the manuscript.

References

[1] Nawal A. Elfeshawy and Osama S. Faragallah, Divided two-part adaptive intrusion detection system, Journal of Wireless Networks 19 (2013), 301-321.

[2] Symantec-Internet Security Threat Report Highlights (Symantec.com), http://www.prdomain.com/companies/Symantec/newreleases/Symantec_internet_205032.htm

[3] J. Ulvila and J. Gaffney, Evaluation of intrusion detection systems, Journal of Research of the National Institute of Standards and Technology 108(6) (2003), 453-473.

[4] G. Guofei, P. Fogla, D. Dagon, W. Lee and B. Skoric, Measuring intrusion detection capability: an information-theoretic approach, Proceedings of the Computer and Communications Security, 2006, pp. 90-101.

[5] O. Linda, T. Vollmer and M. Manic, Neural network based intrusion detection system for critical infrastructures, IJCNN'09, International Joint INNS-IEEE Conference on Neural Networks, Atlanta, Georgia, 2009, pp. 15-23.

[6] E. Hooper, An intelligent detection and response strategy to false positives and network attacks, Proceedings of the Fourth IEEE International Workshop on Information Assurance, University of London, Royal Holloway, United Kingdom, IEEE Computer Society Press, 2006, pp. 12-31.

[7] P. Georgios and K. Sokratis, Reducing false positives in intrusion detection systems, Department of Computer Science and Biomedical Informatics, University of Central Greece, 2009, Available on Science Direct Search.

[8] N. Fenton and M. Neil, Managing risk in the modern world, Applications of Bayesian Networks, Knowledge Transfer Report, London Mathematical Society, 2007.

[9] H. Wang, F. Peng, C. Zhang and A. Pietschker, Software project level estimation model framework based on Bayesian belief networks, Proceedings of the Sixth International Conference on Quality Software (QSIC'06), Beijing, China, 27-28 October, 2006, pp. 209-218.