Network Anomalous Intrusion Detection using Fuzzy-Bayes

Adetunmbi Adebayo.O[1,2]., Zhiwei Shi[1], Zhongzhi Shi[1] and

Adewale Olumide S.[2]

[1]Key Laboratory of Intelligent Information Processing

Institute of Computing Technology, CAS, Beijing 100080 China,

Tel: 86-10-62565533 ext. 5661 86-10-62565533 ext. 5688
{oluadetunmbi, shizw,shizz}@ics.ict.ac.cn

[2]Department of Computer Science,

Federal University of Technology, Ondo State, Nigeria.

adewale@ictp.it *(+234-0803-361-6386)*

Abstract: Security of networking systems has been an issue since computer networks became prevalent, most especially now that Internet is changing the facie computing. Intrusions pose significant threats to the integrity, confidentiality and availability of information for the internet users. In this paper, a new approach to real-time network anomaly intrusion detection via Fuzzy-Bayesian is proposed to detect malicious activity against computer network; the framework is described to demonstrate the effectiveness of the technique. The combination of fuzzy with Bayesian classifier will improve the overall performance of Bayes based intrusion detection system (IDS). Also, the feasibility of our method is demonstrated by the experiment performed on KDD 1999 IDS data set.

Key words: intrusion detection, fuzzy, naïve-Bayes

1. INTRODUCTION

Developing an efficient and effective intrusion detection system to preserve data integrity and system availability has been the aim of researchers in computer security for almost three decades. Intrusion

detection is a process of detecting security breaches by examining events occurring in a computer system.

Basically, there are two approaches to intrusion detection model as described in [3]: Misuse detection model refers to detection of intrusions that follow well-defined intrusion patterns. It is very useful in detecting known attack patterns. Anomaly detection Model refers to detection performed by detecting changes in the patterns of utilization or behavior of the system. It can be used to detect known and unknown attack.

Intrusion detection can also be classified as Network-based (NIDS) or host-based (HIDS) based on source of data used for analyses. The former collect raw network packets as the data source from the network and analyze for signs of intrusions [1, 4]. Host-based IDS operates on information collected from within an individual computer system such as operating system audit trails, C2 audit logs, and System logs [4].

Fuzzy is a novel classification technique that has been widely successfully applied in many applications, and it has been reported to perform well in detecting different attacks due to various reasons spelt out in [2 ,5, 7 and 8]. Ajith, *et al* [2] exploited fuzzy for intrusion detection and the results show an outstanding performance in terms of accuracy. Also, Naïve Bayes has been successfully applied in solving various problems [9].

Fuzzy is introduced to strengthen the detection ability of naïve bayes due to uncertainty nature of intrusions by recognizing anomalous events, consequently leading to reduction in false alarm rate. Fuzzy had been recognized to posses the following quality among others that makes it suitable for the subject matter: ability to readily combine inputs from widely varying sources, degree of alert that can occur with intrusions is often fuzzy because there is no clear distinction between normal and anomaly behavior in a networked computer.

We demonstrate the feasibility of our approach by carrying out experiments on KDD-cup 1999 intrusion detection dataset.

2. THE FUZZYBAYESIAN CLASSIFIER

In naïve Bayes classifier, instances to be classified are described by attribute vectors $\vec{x} = (x_1,...,x_n)$. Bayes classifier assigns to instances most probable or maximum a posterior (MAP), classification from a finite set of c

classes. Bayes classifier is given as:

$$c = \arg\max_{c_j \in C} P(c_j) \prod_{i=1}^{n} P(x_i \mid c_j)$$

Naïve Bayes classifier is trained by a set of labeled training data presented to it in relational form with desirable features because the strength of this model lies very much on the feature set used. In this work, the use of fuzzy is employed during the examination of network connection states to assign weight to various quantifiable variables in the selected features based on predefined fuzzy rules before presentation to the naïve bayes classifier. Weights are assigned with 0.0 representing absolute falseness and 1.0 representing absolute truth. Weights assigned to each features are then used to multiply the prior probability of each class during testing, and with a threshold set, normal and attack traffic can be classified.

3. PROPOSED SYSTEM ARCHITECTURE

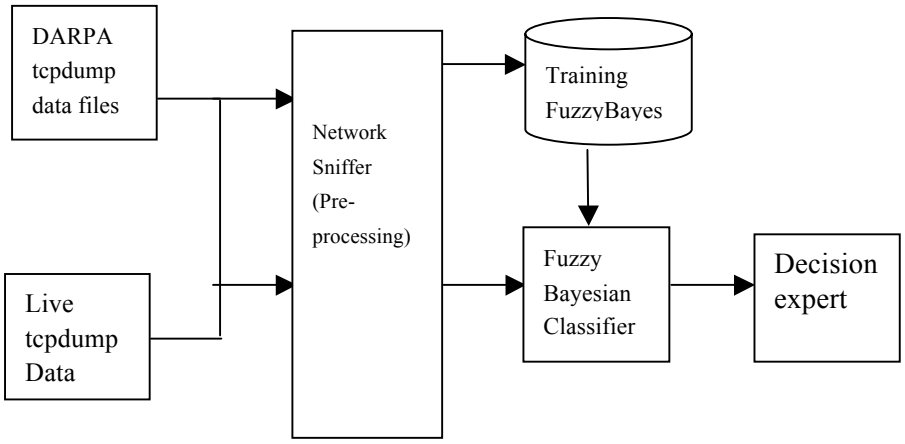

Fig. 1: proposed system architecture

Figure 1 shows the structure of our proposed architecture for real time intrusion detection system via fuzzy-bayes which are divided into two main phases: learning and testing. The network sniffer processes the tcpdump binary into a standard format putting into consideration both the temporal and spatial information of network connections. During the learning phase, two major parameters affecting machine learning are considered: the imbalance of data sets and identification of important features.

Class imbalance problem occurs when there are many more instances of some classes than others in a classification problem which results into suboptimal classification performance, which can have a detrimental effect on learner's behavior [10]. In order to balance the original training data with the utmost aim of improving the system performance, we adopt selective sampling method.

Often, output Y is only determined by the subsets of the input features X. Removing irrelevant features in learning process leads to reduction in computational cost, over fitting, model size and leads to increase in accuracy. In selecting important features, leave-one-out technique of deleting one feature at a time to rank input features and identify the most important features for intrusion detection is adopted [6].

The fuzzybayes technique is then used to obtain the optimal detection model for our system after the learning phase to classify new pattern samples in testing phase.

4. EXPERIMENTAL SETUP

KDD cup 1999 dataset [11] was used for the experiment. The data set has five different classes namely Normal, Dos, R2L, U2R and Probes. In this work, the last four were combined into a class called Abnormal. The training and testing data comprised of 5,924 and 12,130 records respectively.

Frequency table for all the 41 variables were generated given class (normal and abnormal). From the analysis of the frequency table, it showed that some of the extracted attributes did not have any significance in detections of attacks. Attributes on columns 14, 18, 19 and 20 are good examples, while attribute on columns 0, 1, 8, 15, 16, 17, 21 and 36 made little or no impact.

Each variable (attribute) was partitioned into maximum of 20 membership functions except those with variations less than five which forms the basis for fuzzy rules and assigning of weights.

In cases where clear distinction could not be established amongst the variations of an attribute; two or more variables were combined to differentiate close cases and consequently in assigning of weights. With all the 41 attributes, result obtained is shown in Table 1

Table. 1: Percentage accuracy of classified data

Class	No. of records	Correctly Classified.	Wrongly classified	Percentage Of accuracy
Normal	6514	6331	183	97.19
Abnormal	5616	5397	219	96.10
Total	12130	11728	402	96.67

Result obtained was the same when only 29 attributes were used (i.e, removing columns 0,1,8,14,15,16,17,18,19,20,21, and 36 from the experiment).

5 CONCLUSION

This paper proposed a light weight anomaly detection framework based on fusions of fuzzy and Bayes with the utmost aim of addressing large number of false alarms caused by incorrect classification of events in current system. We have demonstrated the effectiveness of our method on KDD-cup 99 intrusion detection datasets and accuracy is over 96%. And our method reveals redundant features, thereby minimizing the number of features the FuzzyBayes classifier should process and consequently increase IDS detection rate. We plan to conduct more experiments with real-life data using our proposed system.

ACKNOWLEDGMENT

This work is supported jointly by Third World Academy of Sciences (TWAS) and Chinese Academy of Sciences (CAS)

REFERENCES

1. Alan Bivens, Chandrika Palagiri, Raheda smith, Boleslaw Szymanski, Mark Embrechts, "Network-Based Intrusion detection using Neural Networks", www.cs.rpi.edu/~szymansk/paper/anie02.pdf
2. Ajith Abraham, Ravi Jainb, Johnson Thomas, and Sang Yong Han, "D-SCIDS: Distributed soft computing intrusion detection system", Journal of Network and Computer Applications, Elsevier, 2005.
3. Biswanath Mukherjee, Todd L. Heberlin, and Karl N. Levitt. Network intrusion detection. IEEE Network, 8(3):26-41, 1994.
4. Byunghae Cha, kyung Woo Park and Jaittyun Seo, "Neural Networks Techniques for Host Anomaly Intrusion Detection using Fixed Pattern Transformation. ICCSA 2005, LNCS 3481 pp. 254-263, 2005.
5. Susan M. Bridges and Rayford B. vaughnn, "Intrusion detection via fuzzy data mining", Twelfth annual Canadian Information Technology Security Symposium June 19-23. The Ottawa Congress Centre.
6. Andrew H. Sung and Srinivas Mukkamala, "Identifying important features for intrusion detection using support vector machines and neural networks". IEEE Proceedings of the 2003 Symposium on Application and the Internet (SAINT ' 03).
7. Jonatan Gomez and Dipankar Dasgupa "Evolving Classifiers for Intrusion Detection", Proceedings of the 2002 IEEE Workshop on

Information Assurance, United States Military Academy, West Point, NY, June 2001

8. John E. Dickerson, Jukka Juslin, Uuramia Koukousoula, and Julie A. Dickerson, "Fuzzy intrusion detection (FIRE)", Electrical and Computer Engineering Department, IOWA state University, Ames, IA, USA. www.eng.iastate.edu/~julied/research.html
9. Christopher Krungel, Darren Mutz, William Robertson and Fredik Valuer, "Bayesian Event classification for intrusion detection", Proceedings of the 19th Annual Computer Security Applications Conference (ACSAC'03), 2003
10. Miroslav Kubat and Stan Matwin, "Addressing the curse of Imbalanced Training Sets: One sided selection" Proc. 14th International Conference on Machine Learning, 1997
11. KDD Cup 1999 Data: http://kdd.ics.uci.edu/databases/kddcup99/