Intrusion Detection Based On J48 Algorithm

Aswathy T[1], Misha Ravi[2]

[1]*M. Tech. Student, Computer Science and Engineering, Sree Buddha College of Engineering, Kerala, India.*
[2]*Assistant Professor, Computer Science and Engineering, Sree Buddha College of Engineering, Kerala, India.*

---***---

Abstract – *The prompt development of computer networks, Intrusion Detection System (IDS) is increasingly a key element of the system security. There are various approaches being employed in intrusion detection systems, but the main problem is to correctly detect intruder attacks in the computer system. This paper presents a decision tree based intrusion detection system with NSL –KDD dataset. As the experimental results, the proposed system is based on J48 decision tree algorithm and it efficiently detects intrusion with an accuracy of 96.50%. Experimental results show that the proposed method outperforms baselines with respect to various evaluation criteria.*

Key Words: Intrusion, NSL-KDD, Network Data, Weka, Decision tree.

1. INTRODUCTION

In computational intelligence community, research on machine learning for intrusion detection is becoming an increasingly important technology. The intrusion detection system identifies network intrusions and monitors network traffic. Intrusion detection systems are categorized as misuse based intrusion detection and anomaly based intrusion detection systems. Misuse based system [4] deals with predefined attack signatures and anomaly based intrusion detection system detects deviations from normal attack behaviours. Various techniques and algorithms are used in intrusion detection. This paper studies about intrusion detection using a decision tree algorithm. In statistical learning and data mining community, the decision tree algorithm is a well-known classification method. In this paper, J48 algorithm is used as a decision tree algorithm.

Intrusion detection systems are dealing with attacks by collecting information from a variety of system and network sources and then identifying the symptoms of security problems. An intrusion detection system is a defence measure that identifies network activity to find intrusions. The main aim is to decrease false positives generated by the decision tree algorithm. Based on network attack behaviors, Intrusion detection system is designed. Learning based intrusion detection system has a training data and a traffic states. The training data contains a set of attribute values. The traffic states include normal or attack data. From the attribute values, the prediction algorithm detects patterns of intrusions. For intrusion detection, available dataset is NSL-KDD dataset.

1.1 Intrusion Detection

An intrusion detection system is a software that identifies network traffic and vulnerabilities in the computer system. It has the ability to recognize patterns of typical attacks. The major classifications of intrusion detection are network intrusion detection systems and host-based intrusion detection system. Network intrusion detection systems are placed at a point within the network. This type of systems detects unwanted traffic at each layer, but concentrate mostly on the application layer. Most network intrusion detection systems are easy to deploy on a network and can often view traffic from many systems at once. Host intrusion detection systems are run on individual host. In this type of systems will detect only the data coming on the host machine and work on the entire network.

2. RELATED WORK

In 1980's intrusion detection research is started and then plenty of techniques have been introduced to build intrusion detection systems. Some of such systems are given below.

Juan Wang et al. [1] presents an intrusion detection algorithm based on decision tree technology. In this system, C4.5 decision tree is used to build the intrusion detection system. Firstly, to convert the decision tree into the rules and it is saved in a knowledge base. These generated rules are used to decide whether the network behavior is normal or abnormal. The decision tree is constructed on stage one. At stage two, classification rules are extracted and in the final stage, according to classification rules, network behavior is determined. In intrusion rules construction process, the information gain ratio is used in place of information gain.

Kajal Rai et al. [2] proposed a decision tree based intrusion detection. According to C4.5 decision tree approach, a decision tree algorithm is developed in the proposed system. In this system, the algorithm is designed to address two issues such as feature selection and split value. Using the information gain, the most relevant features are selected and in such way that makes the classifier unbiased towards most frequent values, the split value is selected. The proposed Decision Tree Split (DTS) algorithm can be used for signature based intrusion detection and the algorithm is compared with the Classification and Regression Tree (CART) and AD Tree. Experimentation is performed on NSL-KDD dataset.

Kai Hwang et. al [3] developed an intrusion detection with weighted signature generation over anomalous internet episodes. This intrusion detection system contains SNORT and an anomaly detection subsystem. To characterize anomalous attacks and extract their signatures by the weighted signature generation algorithm. For accurate intrusion detection, the signatures are extracted from an anomaly detection system and adds them into SNORT. From internet connections, anomalous traffic episodes are mined. The episode rule mining engine consists of training and detection phase. The normal traffic database without attacks is generated during the training phase. In the detection phase, attacks may have monitored.

3. SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

The proposed system has only admin module. The admin module has two sub modules, namely dataset management module and prediction module.

3.1.1 Dataset Management Module.

The dataset management module is handled by admin. For intrusion detection, the dataset used is NSL-KDD dataset. The dataset consists of 41 attributes including class label. It contains normal connection and attack types.

3.1.2 Prediction Module

In prediction module, the algorithm used is J48 decision tree algorithm. The prediction algorithm decides whether the network connection is normal or abnormal.

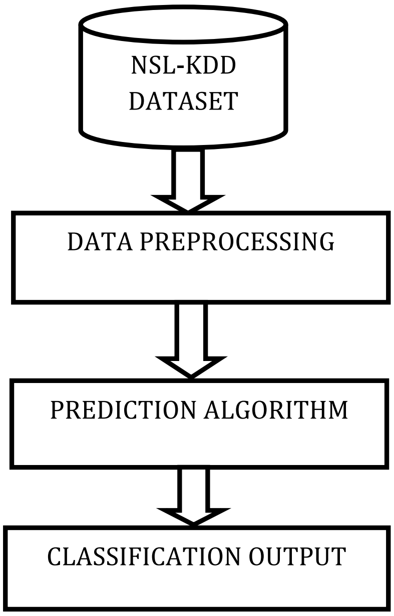

Fig 1: System Architecture

The proposed system [5] consists of mainly two steps such as data preprocessing and prediction. In data preprocessing the dataset is uploaded and it is stored in a database table. The dataset used is NSL-KDD and it contains 41 attributes and a class label. It contains normal connection and four attack categories. The four attack categories are probe, DOS, R2L and U2R. In prediction, J48 decision tree algorithm is used. The implementation of J48 algorithm is done with the help of Weka tool. Weka is the Waikato Environment for Knowledge Analysis is a data mining tool available free of cost under the GNU General Public License. It is a collection of machine learning algorithms. The version used in this system is 3.8.3 and is used for predictive modeling. This tool accepts the input file either in comma separated value (csv) or attribute-relation file format (arff) file format. The classification results produced by Weka tool are in the form of a confusion matrix which gives the actual versus predicted classification output.

For building decision trees in Weka, J48 algorithm is used.

The full name of J48 is weka. classifiers. trees. J48. To make a decision tree, a greedy and top- down approach is adopted. At every node in the decision tree, the training set is partitioned and results small partitions. The J48 classifier is a recursive algorithm for constructing C4.5 pruned or unpruned decision trees. Once a decision tree is constructed, then it can be used to classify testing data that has the same features as the training data. The classification output of the proposed system is it identifies the occurrence of intrusion and it is stored in a data table. The precision value, recall and accuracy of the J48 decision tree algorithm is computed with the help of NSL-KDD dataset.

3.1.3 Evaluation Criteria

For the purpose of this project, precision value, recall, accuracy is used as evaluation criteria. Also performance time is calculated. In this proposed system, 47 milliseconds are obtained as performance time.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Precision = TP/TP+FP

Recall = TP/TP+FN

Where TP is the True positive, which are actually an attack and classified as abnormal. TN is the True negative, it represents the number of normal connections correctly classified as normal. FP is the False positive which is actually normal, but classified as an attack and FN is the False negative, which is the number of attacks incorrectly classified as normal.

In this work, J48 decision tree algorithm is used for prediction and got the following values for the NSL-KDD data set.

- Accuracy = 96.50%
- Precision = 95.89%
- Recall = 99.05%

3. CONCLUSION

In security infrastructure, intrusion detection and prevention are very important. This paper has presented an intrusion detection system based on a decision tree algorithm. J48 algorithm to build patterns of intrusions. The proposed approaches are implemented in Java programs using the WEKA environment. It improves the detection performance of the proposed method. The intrusion detection system was able to achieve 96.5% accuracy.

REFERENCES

[1] J. Wang, Q. Yang and D. Ren, "An Intrusion Detection Algorithm Based on Decision Tree Technology," *2009 Asia-Pacific Conference on Information Processing*, Shenzhen, 2009, pp. 333-335.

[2] Rai, Kajal & Devi, Mandalika & Guleria, Ajay. (2016). Decision Tree Based Algorithm for Intrusion Detection. International Journal of Advanced Networking and Applications. 7. 2828-2834.

[3] K. Hwang, M. Cai, Y. Chen, and M. Qin, "Hybrid Intrusion Detection with Weighted Signature Generation over Anomalous Internet Episodes," Vol. 4, pp. 41 – 55, Feb. 2007.

[4] M. Kumar, M. Hanumanthappa and T. V. S. Kumar, "Intrusion Detection System using decision tree algorithm," *2012 IEEE 14th International Conference on Communication Technology*, Chengdu, 2012, pp. 629-634.

[5] L. Li, Y. Yu, S. Bai, Y. Hou and X. Chen, "An Effective Two-Step Intrusion Detection Approach Based on Binary Classification and k -NN," in *IEEE Access*, vol. 6, pp. 12060-12073, 2018.

BIOGRAPHIES

Aswathy T, she is currently pursing Master's Degree in Computer Science and Engineering in Sree Buddha College of Engineering, Elavumthitta, Kerala, India. Her research area of interest includes the field of data mining, internet security and technologies

Misha Ravi received the master's degree in Software Engineering from Cochin University of Science and Technology, Kerala. She is an Assistant Professor in Department of Computer Science and Engineering, at Sree Buddha College of Engineering.