International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

Binary Multi Class Detection for Network Intrusion Detection using C 5.0 and ANN

Meghana Solanki[1], Priya Raut[2]
[1,2]*Assistant Professor, Department of Computer Engineering, DY Patil College of Engineering, Pune, India*

***Abstract*: Intrusion Detection Systems (IDSs) are authoritative structures. It not only audits but also evaluates events to catch signs of security problems. It also takes action to bar incursion. In this paper, the Binary Multi-class Detection (BMD) method operated together with the C5.0 method as well as the Artificial Neural Network is proposed for flexible network intrusion detection. It promotes the detection rate including the false alarm rate. There are some difficulties in data mining situations i.e. handling imbalance datasets, dealing with continuous attributes, and reducing noise in training dataset, which are cough by the proposed BMD algorithm. We correlated the work of the proposed BMD method with extant algorithms in case of the detection rate, accuracy along with false alarm rate. We adopt the NSLKDD benchmark intrusion detection dataset. The experimental results prove that the proposed BMD method has a diminished false alarm rate. It also has good detection rate hinge on the imbalanced dataset.**

***Keywords*: Intrusion Detection Systems, C5.0, Imbalance Data, Detection Rate, False Alarm Rate.**

1. Introduction

Network security is a headmost concern these days as the network usage is cultivating in multi-dimensions due to increased use of handheld gadgets. Intrusion Detection Systems can cooperate to find out baneful purpose of network users. There exist many machine learning algorithms which can grasp from the training data as well as can hypothesize in case of new untrained data. There exist two types of intrusion detection technique, the first one is Misuse Detection and other is Anomaly Detection. First one can hook the known attacks. It works on the offline data. The other can find out any abnormal behavior. It can work strong on online data. The NSLKDD data set is a classic data set. It is adopted for the research in case of intrusion detection systems. is a The process of monitoring, detecting, analyzing unauthorized use, misuse, and abuse of computer systems is done by Network attack detection. Network intrusion detection systems play a vital act in case of network information security. A large NIDS server can be set up on a backbone network, to monitor all traffic; can also scan system files looking for unauthorized activity and to maintain data and file integrity can be scanned by a NIDS server. It can also examine server log files as well as count on suspicious traffic. The main purpose behind our work is to frame a new multiclass intrusion detection method by using C5.0 algorithm as well as Artificial Neural Network method. It revamps the detection performance. We run various simulations to assess our proposed method. The impact of imbalance dataset to the final results is reduced by extracting the ten subsets data from the training data. It also enhances the training efficiency. We make the use of NSLKDD dataset in case of our experiments. It's regarded as a benchmark to assess the work of IDSs. Our proposed BMD algorithm achieves a very low false alarm rate as compared with other detection methods. It still sustains a high detection rate.

2. Literature survey

In these papers [1], [2], [3] and [4], an author proposed different approaches to solve network intrusion problems. In this paper [5], an author proposed an investigation on latest research literatures which designed intelligent intrusion detection model by using data mining as well as machine learning techniques in IDSs. In these papers [6], [7] and [8], an author studied decision tree algorithms to identify the attacks in IDSs. In these papers [9] and [10], an author introduced ID3 and C4.5; two methods which are classical ways to build a decision tree. In this paper [11], an author introduced C5.0 method which is a new updated decision tree based on C4.5 method [12] along with many recent functions. In this paper [13], an author presented not only a simple application of decision tree in IDS but also gave the normal as well as abnormal detection results. In this paper [14], an author proposed a new decision tree classifier by using the different features of raw activity data as well as different sizes of observation windows in computer network system. In this paper [15], an author described the process generating the decision tree step by step, and the decision tree was evaluated to get the multi-class detection results. In these papers [16], [17], an author proposed an effective combined classifier approach by using tree algorithm.

3. System architecture

A. Decision tree

C5.0 method is evolved by Quinlan based on C4.5 algorithm [11]. It has many recent technologies as well as the most important application is "boosting" technology. The most prominent methods in intrusion detection system are neural network method, method, as well as random forest method.

Why do we prefer C5.0 method in the proposed BMD algorithm? The C5.0 method has good detection accuracy as well as a short detection time. We also find that C5.0 method operate on both continuous as well as categorical features. Furthermore, they are potent against redundant as well as correlated variables. They are crucial to handle the 42 features of NSLKDD dataset.

B. Artificial neural network

ANNs are constituted of multiple nodes, which act like biological neurons of human brain. The neurons are hooked up by links as well as they collaborate with each other. The nodes can accept input data and do simple exercise on the data. The output of these activities is hand hovered to other neurons. The output at each node is called its activation or node value.

C. Implementation of BMD

The universal structure of proposed BMD method is interpreted in figure 1. The proposed BMD method can be divided into the following steps: at the beginning, as per the predefined rules the training dataset draw ten subsets. In next step, the given subsets train the ten predictors. Then, Artificial Neural Network method further handles the final output results. At last, the BMD method will judge the new sample from the testing dataset to obtain its predicted label. We have a set of possible labels. We start with training ten predictors by using ten subsets. The final label figure outing is not only difficult but also extremely computationally extravagant. For this reason, we recommend Artificial Neural Network method. The final Label can be picked up by Artificial Neural Network method. Finally, we will use this BMD algorithm to predict new sample if they come.

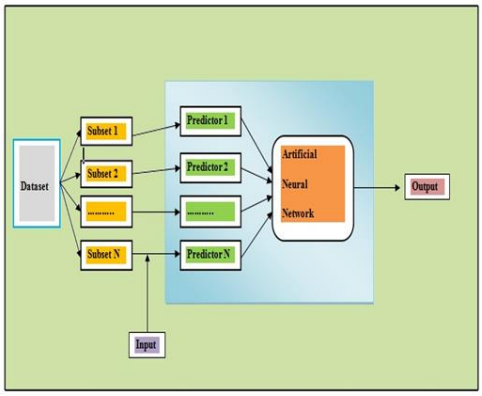

Fig. 1. Architecture for intrusion detection system

4. Experimentation and Result Analysis

A. Datasets

The given dataset is taken from the annual Knowledge Discovery and Dataset Mining contest (NSLKDD). The NSL-KDD data set with 42 attributes is used in this empirical study. This data set is an improvement over KDD'99 data set. There are four different categories of attack: Denial of Service Attack (DoS): e.g. SYN Flood, Remote to Local Attack (R2L): e.g. buffer overflow to gain root privileges, User to Root Attack (U2R): e.g. guessing passwords, Probing Attack: e.g. port scanning. Table 3 present the different amounts of each attack type in the training dataset. Table 4 gives the same information for the testing data.

B. Preprocessing

We adopt decision trees to recognize the most important features. Trees elect features with the highest information gain as well as position the most important features at the elite levels of the tree. Therefore, we not only create ten C5.0 methods but also average the usage of the used features to get an idea of the significance for each feature.

C. Handling imbalance dataset

Table 1 gives the distribution of each attack training dataset, where the magnitude of U2R and R2L is only 0.04% and 0.032%, respectively. Table 2 shows the bulk of each attack in testing dataset, where U2R and R2L have many recent attack types and they are not in training dataset. We agree to use subsets to train the proposed BMD method instead. We excerpt ten subsets with replacement.

Table 1
Attack in training dataset

Attack	Percentage (%)
Probe	0.85
Dos	81.20
U2R	0.04

Table 2
Attack in testing dataset

Attack	Percentage (%)
Probe	2.56
Dos	76.57
U2R	0.03
R2L	7.10

D. Simulations

Proposed BMD method uncovers four types of network attacks including DoS, Probe, R2L and U2R as well as one normal type. The performance of the BMD method is assessed by the detection accuracy, detection false alarm rate and detection rate. We run 25 times of the proposed BMD method to obtain average results. We adopted R Confusion matrix for all types is shown in table 3. Table 4 shows the performance of proposed BMD method. We correlate the proposed BMD method with SVM algorithm as well as C5.0 method. Table 5 shows the overall accuracies of these three methods. The proposed BMD method surpasses both SVM algorithm as well as C5.0 method. The work of the proposed BMD method is substantial and marvelous in case of the detection capability of all types. It is clear that the proposed BMD method carry the minimal false alarm value at all types including normal, DoS, Probe, R2L and U2R.

Table 3
Confusion Matrix for All Attack Type

		Actual				
	Features	**DoS**	**Normal**	**Probe**	**R2L**	**U2R**
Pred	**DoS**	254817	91	88	0	0
	Normal	6203	70221	210	15078	5
	Probe	361	280	4122	151	0
	R2L	612	90	250	2502	40
	U2R	2	38	3	36	38

Table 4
Performance (%) Of Proposed BMD Method

Metrics & Feature	**DoS**	**Normal**	**Probe**	**R2L**	**U2R**
Accuracy	98.60	95.81	94.00	57.21	76.02
False Alarm Rate	3.012	7.336	0.465	5.996	0.045
Detection Rate	98.01	99.26	88.01	13.03	52.32

Table 5
Accuracy (%) of Comparison Results

Method	**Accuracy**
SVM	84.01
C5.0	92.88
Proposed BMD method	94.10

5. Conclusion

In this paper, we have constructed a new Binary multiclass detection method for enhancing the performance of network intrusion detection. C5.0 algorithm along with Artificial Neural Network method is used to set up the model. At the first stage the dataset will be preprocessed into ten subsets. Then, the given subsets will train all the predictors. The output of all predictors is handled by ANN. The reliable benchmark dataset for comparing the performance of network intrusion detection is NSLKDD dataset. The time efficiency of training phase is improved using ten subsets with replacement. It also reduces the tenderness of imbalance distributions of different types in training dataset. The proposed BMD method compared with SVM and C5.0 based on several crucial assessment metrics. With the help of experiments, we demonstrate that the proposed multiclass classifier has superior detection accuracy capability. It not only achieves a very low false alarm rate but also high detection rate. We run our algorithm for 25 times to obtain the average results. The experiment results depict that our proposed algorithm is reproducible. It is persistent over a different number of runs. Our proposed algorithm gains a competitive performance, as compared with the other detection algorithms, based on the benchmark dataset.

References

[1] W. Hu, W. Hu, and S. Maybank. Adaboost-based algorithm for network intrusion detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 38(2):577–583, 2008.
[2] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maci´a-Fern´andez, and E. V´azquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. computers & security, 28(1):18–28, 2009.
[3] R. M. Elbasiony, E. A. Sallam, T. E. Eltobely, and M. M. Fahmy. A hybrid network intrusion detection framework based on random forests and weighted k-means. Ain Shams Engineering Journal, 4(4):753–762, 2013.
[4] R. A. R. Ashfaq, X.Z. Wang, J. Z. Huang, H. Abbas, and Y.L. He. Fuzziness based semi-supervised learning approach for intrusion detection system. Information Sciences, 378:484–497, 2017.
[5] M. Ahmed, A. N. Mahmood, and J. Hu. A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60:19–31, 2016.
[6] X. Li and N. Ye. Decision tree classifiers for computer intrusion detection. Journal of Parallel and Distributed Computing Practices, 4(2):179–190, 2001.
[7] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P.-N. Tan. Data mining for network intrusion detection. In Proc. NSF Workshop on Next Generation Data Mining, pages 21–30, 2002.
[8] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan. Hybrid decision tree and na¨ıve bayes classifiers for multiclass classification tasks. Expert Systems with Applications, 41(4):1937– 1946, 2014.
[9] J. Quilnan and D. Michie. Discovering rules by induction from large collection of examples. Expert Systems in the Micro Electronic Age. Edinburgh: Edinburgh UniversityPress, pages 168–201, 1979.
[10] J. R. Quinlan. C4. 5: Programming for machine learning. Morgan Kauffmann, 38, 1993.
[11] Q. J. R. “C5”, http://rulequest.com, 2007.
[12] R. Pandya and J. Pandya. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. International Journal of Computer Applications, 117(16), 2015.
[13] C. Sinclair, L. Pierce, and S. Matzner. An application of machine learning to network intrusion detection. In Computer Security Applications Conference, 1999. (ACSAC’99) Proceedings. 15th Annual, pp. 371–377. IEEE, 1999.
[14] X. Li and N. Ye. Decision tree classifiers for computer intrusion detection. Journal of Parallel and Distributed Computing Practices, 4(2):179–190, 2001.
[15] J. H. Lee, J. H. Lee, S. G. Sohn, J. H. Ryu, and T. M. Chung. Effective value of decision tree with NSLKDD 99 intrusion detection datasets for intrusion detection system. In Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on, volume 2, pages 1170–1175. IEEE, 2008.
[16] S. Stolfo et. al. The third international knowledge discovery and data mining tools competition,
http://NSLKDD. ics. uci.
edu/databases/NSLKDDcup99/NSLKDDcup99. html
[17] J. R. Quinlan et al. Bagging, boosting, and c4. 5. In AAAI/IAAI, Vol. 1, pages 725–730, 1996.