ISSN 2278-3091
Volume 7, No.5, September - October 2018
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse02752018.pdf
https://doi.org/10.30534/ijatcse/2018/02752018

Multilevel Ensemble Classifier using Normalized Feature based Intrusion Detection System

Apoorva Deshpande[1], Ramnaresh Sharma[2]
P.G. Student, Department of Computer Science and Engineering, MPCT, Gwalior, India[1]
Associate Professor, Department of Computer Science and Engineering, MPCT, Gwalior, India[2]

ABSTRACT: As network applications grow rapidly, network security mechanisms require more attention to improve speed and accuracy. The evolving nature of new types of intrusion poses a serious threat to network security: although many network security tools have been developed, the rapid growth of intrusive activities is still a serious problem. Intrusion detection systems (IDS) are used to detect intrusive network activity. Machine learning and data mining techniques have been widely used in recent years to improve intrusion detection in networks. These techniques allow the automatic detection of network traffic anomalies. One of the main problems encountered by researchers is the lack of data published for research purposes. In this research work the proposed model for intrusion detection is based on normalized feature and multilevel ensemble classifier. The work is performed in divided into four stages. In the first stage data is normalized using statistical normalization. In second stage multilevel ensemble classifier is used.

KEYWORDS: Intrusion Detection, Ensemble Classifier, Machine Learning, Classification, Accuracy, Detection Rate.

1.INTRODUCTION

Intrusion Detection Systems (IDS) are security tools that detect attacks on a network or host computer. An IDS is based on the host or network. A host-based IDS detects attacks on a host computer, while a network-based IDS, also known as a network intrusion detection system (NIDS), detects intruders in a network by analyzing network traffic and typically installed in the gateway network or server, host-based intrusion detection systems can be divided into four types: (a) file system monitor, (b) log file scanners, (c) link analyzers, (d) IDs based on kernels [1, 2].

Based on the data analysis technique, there are two broad categories of IDS titles, which are mainly based on signatures and anomalies. A signature-based system detects attacks by analyzing network data for attack signatures stored in its database. This type of IDS detects previously known attacks whose signatures are stored in their database. On the other hand, an IDS anomaly appearance - deviations from the traditional behavior of the subjects. The anomaly-based systems are able to detect new attacks [3-7].

Here are some very common methods used by intruders to take control of computers: Trojan horses, backdoors, denial of service, viruses transmitted via email, package tracking, identity theft and so on. a network package has 42 features and four simulated attacks like [8-12]:

DoS (Denial of Service): excessive use of bandwidth or unavailability of system resources resulting from denial of service attacks. Examples: tear and smurf.

User root (U2R) Attack: Initially, access to malicious users on a normal user account, obtained after logging in to root exploiting system vulnerabilities. Examples: Perl, Load Module and Eject attacks.

Probe attack: access to all network information before launching an attack. Examples: ipsweep, nmap attacks.

Root to Local Attack (R2L): exploiting some of the vulnerabilities of the network, the attacker gets local access by sending packets to a remote machine.

Machine learning techniques can be effective in detecting intruders. Many intrusion detection systems are based on machine learning techniques [13,14,15]. Learning algorithms are created in the offline data set or in real data from academic or organizational networks.

Typically, machine learning techniques are divided into two classes: i.e. Supervised learning and unsupervised learning. In supervised learning, the set of learning data is immediately accessible with its destination vector. The learner learns from available data taking guidance of the output vector [16,17,18].

2.RELATED WORK

Taeshik Shon [26] designed a framework consists of two Main components: Genetic algorithm (GA) for the characteristic selection machine and vector carrier (SVM) for the packet behavior classification.

Yadigar Imamverdiyev [27] discussed that intrusion detection systems are one of the most relevant security features against network attacks. Machine learning methods are used to analyze network traffic parameters in the presence of attack signs. This article discusses the extreme machine learning method for detecting intrusions in network traffic. The experimental results lead to the conclusion of the practical significance of the proposed approach to detect attacks in network traffic.

Athanasios Tsiligkaridis [28] developed a method to detect atypical bottlenecks in traffic City of Boston Our motivation is to detect these traffic jams which are often caused by an event (for example, an accident, a lane closure, etc.) and allow the city to intervene before congestion roads and adjacent roads are negatively affected. Using a data set on the traffic jams provided by the city of Boston presents a new detection system for the identification of anomalous jams. We demonstrate its effectiveness by using it to identify traffic jams that cannot be explained with typical traffic patterns.

Bhanu Vrat et al. [29] discussed that detection of anomalies is important requirement to protect a network against the strikers. Detects attacks on a network the analysis of the behavioral model was a important field of study for many researchers application systems in IPv4 and IPv6 networks. For accurate detection of anomalies, it is essential implement and use effective data mining methodology such as machine learning. In this article we considered a model of anomaly detection that uses machine learning algorithms for data mining in a network to detect anomalies present at any time. This the proposed model is evaluated against denial of service Attacks (DOS) in IPv4 and IPv6 networks selecting the most common and obvious features of IPv6 and IPv4 networks to optimize detection. The results also show that the proposed system can detects most IPv4 and IPv6 attacks effectively way.

A. Khan et al. [30] presented an experimental analysis to demonstrate the performance analysis of some existing techniques in order that they will be used further in developing Hybrid Classifier for real data packets classification. The given result analysis shows that KNN, RF and SVM performs best for NSL-KDD dataset.

Khadija Hanifi et al. [31] discussed that network attacks are exceptional cases they are not observed in the normal behavior of the traffic. In this work, to detect network attacks, using the k-means algorithm a new semi-supervised anomaly detection system was designed and implemented. During the training phase, normal samples were split into clusters by applying the k-means algorithm. So in To be able to distinguish between normal and abnormal samples, based on their distance from cluster centers and using a validation data set, a threshold value has been calculated. New samples that are far from cluster centers more than the threshold value is detected as anomalies. We used NSL-KDD- a data set labeled network connection traces - to test ours the effectiveness of the method. The experiments result in NSL-KDD dataset, shows that we have reached an accuracy of 80.119%.

Wathiq Laftah Al-Yaseen et al. [35] proposed a hybrid multilevel intrusion detection model that uses a carrier vector machine and an extremely powerful machine to improve the efficiency of detection of known and unknown attacks. A modified K-Means algorithm is also proposed to create a high quality training data set that greatly enhances the classifier performance. Modified K-Means are used to create new small training records that represent all initial training data, significantly reduce classifier training time and improve intrusion detection system performance. The famous KDD Cup 1999 dataset is used to evaluate the proposed model. Compared to other methods based on the same data set, the proposed model shows a high detection efficiency of the attacks and its accuracy (95.75%) is the best performance ever achieved.

I. METHODOLOGY

The biggest challenge for today is to protect users from intruders because the Internet is often used. Intrusion Detection Systems (IDS) are one of the security tools available to detect potential intrusions in a network or host. Research has shown that the use of machine learning techniques in intrusion detection can provide a high level of accuracy and a low rate of false alarms. Precise predictive models can be created for large amounts of data with supervised machine learning techniques, which is not possible with traditional methods.

IDS learns models from training data so that only the known attack can be detected, new attacks can not be identified. This research is based on the design of a feature-based optimized classifier and on the analysis of three different data sets.

This section describes the proposed hybrid model for intrusion detection. The KDD-99 dataset is used as a benchmark to evaluate the performance of the proposed model. The algorithm flow of the proposed method is described as follows:

Following steps will be used to build the proposed model for intrusion detection:

Step 1: Convert the symbolic attributes protocol, service, and flag to numerical.

Step 2: Normalize data to [0,1].

Step 3: Separate the instances of dataset into two categories: Normal, DOS, R2L, U2R and Probe.

Step 4: The data set is divided as training data and testing data.

Step 5: Train classifier with these new training datasets.

Step 6: Test model with dataset.

Step 7: Finally computing and comparing Accuracy and Detection rate for classifiers.

A. *Proposed Methodology*

The proposed algorithm flow diagram of intrusion detection model is illustrated in figure 1. The proposed framework consists of three phases i.e. Preprocessing, Post Processing Phase and Intrusion Detection Phase. Below each stage is described individually in details.

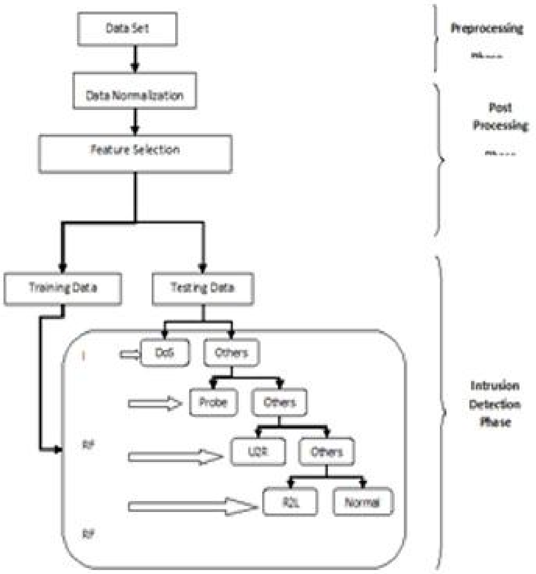

Figure 1: Proposed Flow Diagram of Intrusion Detection System

Preprocessing Phase

This stage purpose is to preprocess the database file in which there is conversion of symbolic attributes protocol, service, and flag in numerical is done. Further data is normalized. Table I illustrates the normalization of dataset instances using statistical normalization and concluded that statistical normalization illustrates best data normalization.

Table 1: Number of Instances after Normalization

Category	No. of Instances	Statistical Normalization
Normal	97278	25360
Dos	391458	3784
Probe	4107	720
U2R	52	46
R2L	1126	422
Total	494021	30332

Post-Processing Phase

Once pre-processing is applied, the pre-processing Module creates the Feature Vector matrix of dataset that represents in which each row i represents the instances and j represents the packet attributes.

Intrusion Detection Phase

For intrusion detection or classification dataset multilevel classifier is used. In this research work multilevel SVM-ELM classifier is used. Hybrid multilevel classifier is illustrated in Figure 1.

3.SIMULATION RESULTS

To evaluate the proposed algorithm, it is concentrated on three indications of performance: detection rate and accuracy.

If one sample is an anomaly and the predicted label also stands anomaly, then it is called as true positive (TP).

If one sample is an anomaly, but the predicted label stands normal, then it is called as false negative (FN).

If one sample is a normal and the predicted label also stands normal, then it is true negative (TN).

If one sample is normal, but the predicted label stands anomaly, then it is termed as false positive (FP).

TP stands the number of true positive samples, FN stands the number of false negative samples, FP stands the number of false positive samples, and TN stands the number of true negatives.

From equation (i) and (ii), the accuracy and detection rate are calculated.

Accuracy = (TP+TN)/(TP+TN+FP+FN)*100 (i)

Detection Rate =TP/(TP+FN)*100 (ii)

4.RESULT ANALYSIS

For performance evaluation, multilevel hybrid ensemble classifier is used. The performance evaluation are performed using normalized feature based multilevel ensemble classifier. By applying normalization technique over KDD-99 dataset it has been observed that best result is obtained by using ensemble classifier. Table II shows the performance evaluation of multilevel classification algorithm over datasets. From the result analysis it has been analyzed that accuracy and detection rate of hybrid multilevel RF classification achieved best result.

Table 2: Performance Evaluation of Proposed Algorithm

Performance	Multi-Level RF-ELM	Multi-Level RF	Multi-Level ELM
Accuracy	95.3906	99.4162	90.0912
Detection rate	48.3603	91.2254	17.7866

5.CONCLUSION

This research work proposes a multi-level hybrid ensemble classification intrusion detection system. The proposed model illustrates better performance than multilevel ELM and Multilevel RF-ELM models. The normalization technique is used to pre-process training dataset and provides high accuracy and detection rate as compared to existing work. According to simulation on KDD-99 dataset, the proposed algorithm achieved approx.. 99% accuracy and approx. 91% detection rate. The proposed system is implemented with the entire training and testing dataset. In future work the system will be designed for classification by using reduced features with enhanced performance with respect to accuracy, detection rate and false alarm rate.

REFERENCES

1. Garcia-Teodoro, P., “Anomaly-based network intrusion detection: techniques”, systems and challenges. Comput. Security vol. 28. issue, pp. 18–28, 2009.
2. Sufyan T Faraj Al-Janabi, Hadeel Amjed Saeed, “A neural network-based anomaly intrusion detection system”, IEEE, 2011.
3. J. Ryan, M. Lin, and R. Miikkulainen, “Intrusion Detection with Neural Networks,” Conference in Neural Information Processing Systems, 943–949.
4. A. K. Ghosh and A. Schwartzbard, “A Study in Using Neural Networks for Anomaly and Misuse Detection,” Conference on USENIX Security Symposium, Volume 8, pp. 12–12, 1999.
5. P. L. Nur, A. N. Zincir-heywood, and M. I. Heywood, “Host-Based Intrusion Detection Using Self-Organizing Maps,” in Proceedings of the IEEE International Joint Conference on Neural Networks, pp. 1714–1719, 2002.
6. K. Labib and R. Vemuri, “NSOM: A Real-Time Network-Based Intrusion Detection System Using Self-Organizing Maps,” 2000.
7. Sharma, R.K., Kalita, H.K., Issac, B., “Different firewall techniques: a survey”, International

Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2014. https://doi.org/10.1109/ICCCNT.2014.6963102

8. Meng, Y.-X., "The practice on using machine learning for network anomaly intrusion detection", International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, IEEE, 2011. https://doi.org/10.1109/ICMLC.2011.6016798
9. Sumaiya Thaseen Ikram, Aswani Kumar Cherukuri, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM", Journal of King Saud University –Computer and Information Sciences, 2016.
10. Manjula C. Belavagi and Balachandra Muniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection, Procedia Computer Science", Elsevier, 2016.
11. Saad Mohamed Ali Mohamed Gadal and Rania A. Mokhtar, "Anomaly Detection Approach using Hybrid Algorithm of Data Mining Technique", International Conference on Communication, Control, Computing and Electronics Engineering, IEEE, 2017. https://doi.org/10.1109/ICCCCEE.2017.7867661
12. Ibrahim, H. E., Badr, S. M., & Shaheen, M. A., "Adaptive layered approach using machine learning techniques with gain ratio for intrusion detection systems", International Journal of Computer Applications, vol. 56, issue 7, pp. 10–16, 2012.
13. Wen Feng, Qinglei Zhang, Gongzhu Hu, Jimmy Xiang Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks", Elsevier, Vol 37, pp 127-140, 2014.
14. Shi-JinnHorng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, Citra Dwi Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines" Expert Systems with Applications, Elsevier, vol. 38, pp. 306–313, 2011. https://doi.org/10.1016/j.eswa.2010.06.066
15. O.Y.Al-Jarrah, O. Alhussein, P.D.Yoo, S. Muhaidat, K.Taha and K. Kim, " Data Randomization and Cluster-based Partitioning for botnet intrusion detection", IEEE Transactions on Cybernetics, vol. 46, no. 8, pp. 1796-1806, 2016. https://doi.org/10.1109/TCYB.2015.2490802
16. Solane Duque, Dr. Mohd. Nizam Bin Omar, "Using Data Mining Algorithm for Developing a Model for Intrusion Detection System(IDS)", procedia Computer Science 61 (2015) 46-51.
17. Paul Dokas, Levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, Pang-Ning Tan , " Data Mining for Network Intrusion Detection ", University of Minnesota, Minneapolis, MN 55455, USA.
18. Wenke Lee, Savatore J. Stolfo, Kui W. Mok, " A Data Mining Framework for Building Intrusion Detection Models", Computer Science Department , Columbia University 500 west 120th street, New York 10027.
19. Mathew G. Schultz and Eleazar Eskin and Erez Zadok, "Data Mining Methods for Detection Of New Malicious Executables", Department of Computer Science Columbia University.
20. Aasia Abdullah and Khaleda Afroaz," Data Mining Approaches on Network Data: Intrusion Detection System", International Journal of Advanced Research in Computer Science Volume 8, No. 1, Jan-Feb 2017.
21. Ashok Chalak, Naresh D Harale and Rohini Bhosale, "Data Mining Techniques for Intrusion Detection and Prevention", IJCSNS International Journal of Computer Science and Network Security, Vol. 11 No. 8, August 2011.
22. G. V. Nadiammai and M. Hemalatha, "Effective Approach Towards Intrusion Detection System Using Data Mining Techniques", Department of Computer Science, Karpagam University, Coimbatore 641021, Tamilnadu, India.
23. Nutan Farah Haq, Musharrat Rafni, Abdur Rahman Onik, Faisal Muhammad Shah, Md. Avishek Khan Hridoy and Dewan Md. Farid, " Application of machine Learning Approaches in Intrusion Detection System : A Survey", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 3, 2015.
24. Kuang, F., Xu, W., & Zhang, S., "A novel hybrid KPCA and SVM with GA model for intrusion detection", Applied Soft Computing Journal, vol. 18, pp. 178–184, 2014.
25. Prasanta Gogoi, D.K. Bhattacharyya, B. Borah1 and Juga, K. Kalita, "MLH-IDS: A Multi-Level Hybrid Intrusion Detection Method", The Computer Journal, Vol. 57 issue 4, pp. 602-623, 2014.
26. Taeshik Shon "A Machine Learning Framework for Network Anomaly Detection using SVM and GA", IEEE, 2005.
27. Yadigar Imamverdiyev "Anomaly detection in network traffic using extreme learning machine", IEEE, 2016.
28. Athanasios Tsiligkaridis "Anomaly Detection In Transportation Networks Using Machine Learning Techniques", IEEE, 2017.
29. Bhanu Vrat et al "Anomaly Detection in IPv4 and IPv6 Networks Using Machine Learning", IEEE, 2015.
30. Khan, A., & Nigam, A., "Analysis of Intrusion Detection and Classification using Machine Learning Approaches", International Journal Online of Science, 3(9), 2017. Retrieved from http://ijoscience.com/ojsscience/index.php/ojsscience/article/view/13.
31. Khadija Hanifi ve Hasan Bank "Network Intrusion Detection Using Machine Learning Anomaly Detection Algorithms" , IEEE, 2016.
32. He, L., "An improved intrusion detection based on neural network and fuzzy algorithm. Journal of Networks, vol. 9, issue 5, pp. 1274–1280, 2014.
33. Hoque, M. S., Mukit, M. A. & Bikas, M. A. N., "An implementation of intrusion detection system using genetic algorithm", International Journal of Network Security & Its Applications, vol 4, issue 2, pp. 109–120, 2012.
34. Nour Moustafa & Jill Slay, "The evaluation of Network Anomaly Detection Systems: Statistical

analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set", Information Security Journal: A Global Perspective, 2016.

35. Wathiq Laftah Al-Yaseen, Zulaiha Ali Othman, Mohd Zakree Ahmad Nazri, "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System", International Journal in Expert Systems With Applications, Elsevier, 2017.

36. Hebatallah Mostafa Anwer et al., "A Framework for Efficient Network Anomaly Intrusion Detection with Features Selection", IEEE, 2018.

37. Nour Moustafa & Jill Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 dataset", Information Security Journal: A Global Perspective, 2015.