RESEARCH ARTICLE OPEN ACCESS

Implementation Of Machine Learning Algorithms For Detection Of Network Intrusion

Ponthapalli . Raviteja [1], Mandapati. Satya Venkata Sarojini Devi [2],
Mukka . Gowri [3], Majji. Vamsi Sai Krishna [4],
P V S Prabhakar [5]

Department of Computer Science and Engineering
Nadimpalli Satyanarayana Raju Institute Of Technology
Andhra Pradesh – India

ABSTRACT

For the past few years, network plays a significant role in communication. The computer network allows the computing network devices to exchange information from individuals to individuals. The services of various organizations, companies, colleges, universities are accessed throughout the computer network. This leads to a massive growth in the net - working field. The accessibility of internet has acquired a lot of interest among individuals. In this context, security of information has become a great challenge in this modern area. The information or data that we would like to send be supposed to be secured in such a way that a third party should not take control over them. When we are talking about security, we have to keep three basic factors in our mind: Confidentiality, Integrity and availability. Confidentiality means privacy of information. It gives the formal users the right to access the system via internet. This can be performed suitably along with accountability services in order to identify the authorized individuals. The second key factor is integrity. The integrity service means exactness of information. It allows the users to have self-assurance that the information passed is acceptable and has not been changed by an illegal individual.

Keywords:- Network Security; Intrusion Detection System (IDS); DOS; U2R; R2L; KDD; Support Vector Machine (SVM).

I. INTRODUCTION

An Intrusion Detection System (IDS) is used to watch malicious activities over the network. It can sort the unfamiliar records as normal or attack class. First monitoring is executed, and then they order the network traffic into malicious class or regular class. It works as an alarm system that describes when an illegal activity is observed. The exactitude of the IDS depends upon detection rate. If the performance is high for the IDS, then the correctness of detection is also high. More or less of the intrusion detection schemes are marketed with the ability to stop attacks before they are successful. We are still discovering new systems marketed as intrusion detection schemes. Intrusion detection systems have survived for a long time. They are employed to shield an association from attack. It is a relative concept that tries to identify a hacker when penetration is attempted. Ideally, such a scheme will only alarm when a successful attack is reached. Intrusion detection system is not a complete solution to all attack types. A full protection plan or security tools cannot be superseded by an IDS. The genuine users who may attempt to access the information for their pleasure cannot be placed with the help of IDS. The

goals of IDS provide the necessities for the IDS policy. The potential goals include the tracing:

- IDS detect attacks.
- IDS traces user activity from point of entry to
- IDS generate alerts when required.
- Detect errors in system configuration.
- Provides security of the system without the need of non – expert staff.
- IDS can detect when the system is under attack.
- Provides evidences for attack.

II. ALGORITHMS

a) DECISION TREE

A tree has many analogies in real life, and turns out that it has influenced a broad area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the public figure survives, it uses a treelike model of decisions. Though a commonly employed instrument in data mining for deriving a strategy to achieve a special destination, it's also widely applied in machine learning.

b) RANDOM FOREST

To increase in computational power, we can now choose algorithms which do very intensive calculations. One such algorithm is “Random Forest”. Random forest is like bootstrapping algorithm with Decision tree (CART) model. Say, we have 1000 observation in the complete population with 10 variables. Random forest tries to build multiple CART model with different sample and different initial variables. For instance, it will take a random sample of 100 observation and 5 randomly chosen initial variables to build a CART model. It will replicate the process (suppose) 10 times and then realize a final prediction on each notice. Final prediction is a mapping of each prediction. This final prediction can simply be the mean of each prediction.

c) LOGISTIC REGRESSION

A popular statistical technique to predict binomial outcomes (y = 0 or 1) is Logistic Regression. Logistic regression predicts categorical outcomes (binomial / multinomial values of y). The predictions of Logistic Regression (henceforth, LogR in this article) are in the form of probabilities of an event occurring, i.e. the probability of y=1, given certain values of input variables x. Therefore, the effects of long range between 0-1.LogR models the data points using the standard logistic function, which is an S- shaped curve also called as sigmoid curve and is eaten by the equation.

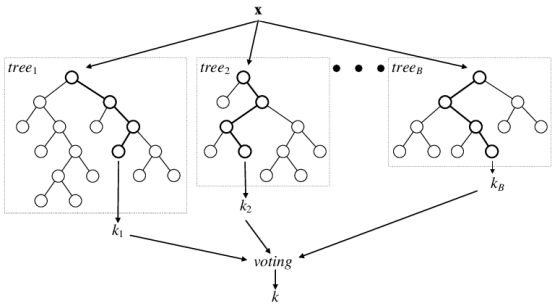
d) SUPPORT VECTOR MACHINES

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be applied for both categorization and regression challenges. Yet, it is mostly used in classification problems. In this algorithm, we plot each data point as a point in n-dimensional space (where n is the number of features you have) with the value of each feature being the value of a particular coordinate. And then, we perform an arrangement by finding the hyper-plane that

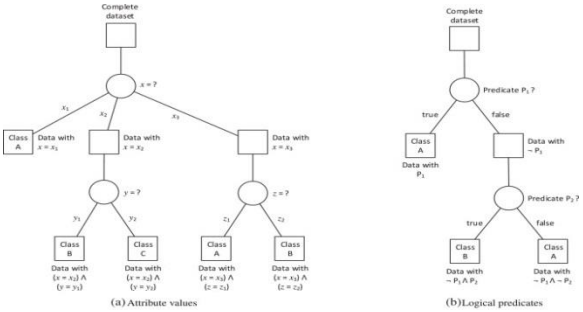
separate the two classes very well (look at the below snapshot). The SVM algorithm is implemented in follow using a substance. The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra, which is out of the scope of this introduction to the SVM. A powerful insight is that the linear SVM can be assign to different use the inner product of any two given observations, instead than the notices themselves. The inner product between two vectors is the sum of the multiplication of each couple of input values. For instance, the inner product of the vectors [2, 3] and [5, 6] is 2*5 + 3*6 or 28. The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

$$f(x) = B0 + sum(ai * (x,xi))$$

III. ARCHITECTURE OF RANDOM FOREST



IV. ARCHITECTURE OF DECISION TREE



V. SCREEN SHORTS

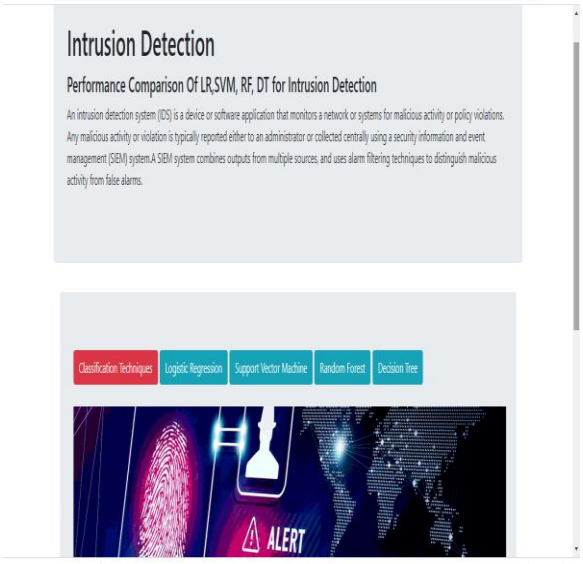


Fig 1:HOME PAGE

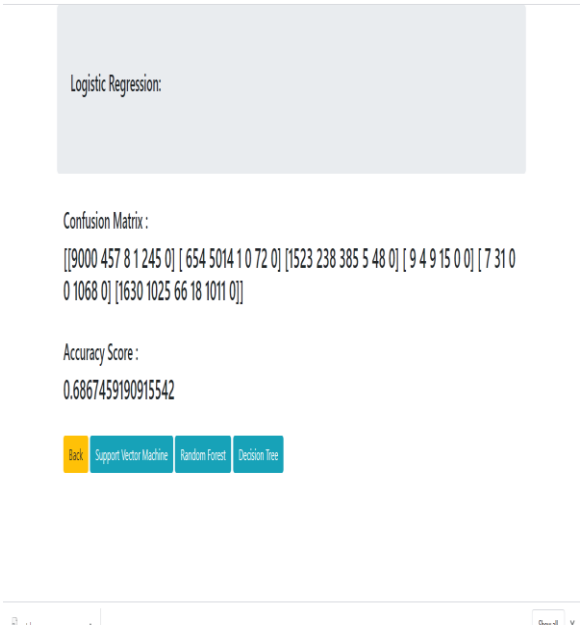


Fig 2:LOGISTIC REGRESSION

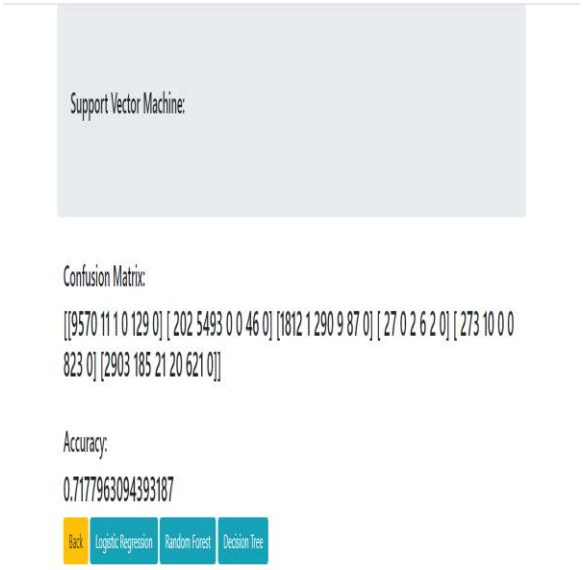


Fig 3:SUPPORT VECTOR MACHINE(SVM)

Fig 4:RANDOM FOREST

Fig 5:DECISION TREE

VI. CONCLUSION

Network Intrusion Detection System is the most used defense technology in the domain of network security. In recent years many of the techniques have been implemented for the intrusion detection system. In this paper, a detailed survey of major techniques implemented on intrusion Detection is presented. Techniques based on Random Forest algorithm Extreme learning the machine, techniques, classification algorithms such as Support Vector Machine, Logistic regression, Decision Tree have been implemented. From the experimental results, this work concludes, the Random forest classifier is considered as a best algorithm because of its highest classification accuracy. On the other hand, while comparing the execution time, the Random forest classifier taking minimum execution time from others.

References

[1] Anderson, J.P. *Computer Security Threat Monitoring and Surveillance*; Technical Report; James P. Anderson Company: Philadelphia, PA , USA, 1980.

[2] Michie, D.; Spiegelhalter, D.J.; Taylor, C. *Machine Learning, Neurall and Statistical Classification*; Ellis Horwood Series in Artificial Intelligence: New York, NY, USA, 1994; Volume 13.

[3] Buczak, A.L.; Guven, E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun. Surv. Tutor.* **2015**, *18*, 1153–1176. [CrossRef]

[4] Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381. [CrossRef]

[5] Agrawal, S.; Agrawal, J. Survey on anomaly detection using data mining techniques. *Procedia Comput. Sci.* **2015**, *60*, 708–713. [CrossRef]

[6] Denning, D.E. An intrusion-detection model. *IEEE Trans. Softw. Eng.* **1987**, 222–232. [CrossRef]

[7] Heberlein, L.T.; Dias, G.V.; Levitt, K.N.; Mukherjee, B.; Wood, J.; Wolber, D. A network security monitor. In Proceedings of the 1990 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA, USA, 7–9 May 1990; pp. 296–304.

[8] Kuang, F.; Zhang, S.; Jin, Z.; Xu, W. A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection. *Soft Comput.* **2015**, *19*, 1187–1199. [CrossRef]

[9] Syarif, A.R.; Gata, W. Intrusion detection system using hybrid binary PSO and K-nearest neighborhood algorithm. In Proceedings of the 2017 11th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 31 October 2017; pp. 181–186.

[10] Pajouh, H.H.; Dastghaibyfard, G.; Hashemi, S. Two-tier network anomaly detection model: A machine learning approach. *J. Intell. Inf. Syst.* **2017**, *48*, 61–74. [CrossRef]

[11] Mahmood, H.A. Network Intrusion Detection System (NIDS) in Cloud Environment based on Hidden Naïve Bayes Multiclass Classifier. *Al-Mustansiriyah J. Sci.* **2018**, *28*, 134–142. [CrossRef]

[12] Shah, R.; Qian, Y.; Kumar, D.; Ali, M.; Alvi, M. Network intrusion detection through discriminative feature selection by using sparse logistic regression. *Future Internet* **2017**, *9*, 81. [CrossRef]

[13] Peng, K.; Leung, V.C.; Huang, Q. Clustering approach based on mini batch kmeans for intrusion detection system over big data. *IEEE Access* **2018**, *6*, 11897–11906. [CrossRef]

[14] Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008, pp. 1096–1103.

[15] Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked denoising autoencoders: Learning useful representations in a deep network

with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.

[16] Deng, J.; Zhang, Z.; Marchi, E.; Schuller, B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2–5 September 2013; pp. 511–516.

[17] Hinton, G.E. A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*; Springer: Berlin, Germany, 2012; pp. 599–619.

[18] Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]

[19] Boureau, Y.l.; Cun, Y.L.; Ranzato, M.A. Sparse feature learning for deep belief networks. In Proceedings of the 21st Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–10 December 2008; pp. 1185–1192.

[20] Zhao, G.; Zhang, C.; Zheng, L. Intrusion detection using deep belief network and probabilistic neural network. In Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, China, 21–24 July 2017; Volume 1, pp. 639–642.

[21] Alrawashdeh, K.; Purdy, C. Toward an online anomaly intrusion detection system based on deep learning. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 195–200.

[22] Yang, Y.; Zheng, K.; Wu, C.; Niu, X.; Yang, Y. Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks. *Appl. Sci.* **2019**, *9*, 238. [CrossRef]

[23] Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.

[24] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105. [CrossRef]

[25] Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach.
IEEE Trans. Neural Netw. **1997**, *8*, 98–113. [CrossRef]

[26] Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada , 26–31 May 2013; pp. 6645–6649.

AUTHOR DETAILS

Ponthapalli . Raviteja is presently pursuing B.Tech(CSE) Department of Computer Science Engineering from N S Raju institute of technology, Visakhapatnam.

Mandapati. Satya Venkata Sarojini Devi is presently pursuing B.Tech(CSE) Department of Computer Science Engineering from N S Raju institute of technology, Visakhapatnam.

Mukka . Gowri is presently pursuing B.Tech(CSE) Department of Computer Science Engineering from N S Raju institute of technology, Visakhapatnam.

Majji . Vamsi Sai Krishna is presently pursuing B.Tech(CSE) Department of Computer Science Engineering from N S Raju institute of technology, Visakhapatnam.

P V S Prabhakar ,B.Tech,M.Tech is working as an Assistant Professor in the Department of Computer Science Engineering from N S Raju institute of technology, Visakhapatnam.