ISSN : 2455-9679
Impact Factor : 3.805

IJRTSM

INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

"ANOMALY DETECTION USING SUPERVISED TECHNIQUES"

Chandan Kumar Ray[1], Prof. Ashish Nema[2], Prof. Jitendra Mishra[3]

[1]*M. Tech Scholar, Department of EC, PCST, Bhopal (India)*

[2]*Assistant Professor, Department of EC, PCST, Bhopal (India)*

[3]*Head & Professor, Department of EC, PCST, Bhopal (India)*

ABSTRACT

From the last two decades the use of computers increasing day by day, and also increasing demand in future. The use of computers play a vital role in every industry and an organization, the more usage of computers need more secure environment for the run smoothly and in a secured manner. The aim of the intrusion detection system is the controlling state and dynamic behaviour of the computer system. This detection system checks all the activities of inspected packets on a network. In security infrastructure, IDSs is one of the essential elements, which allow the networks administrators to identify the policy variations. In intrusion detection system, the different types of alarm rates are increased based on that anomaly and misuse based attacks are classified. Here we presented new model for the analysis and detection of malware variants.

Key Words: *Wireless Communication, Malware detection, Intrusion detection system, Unsupervised techniques, Supervised Techniques.*

I. INTRODUCTION

From the last two decades the use of computers increasing day by day, and also increasing demand in future. The use of computers play a vital role in every industry and an organization, the more usage of computers need more secure environment for the run smoothly and in a secured manner. An intrusion-detection system (IDS) can be defined as software or hardware tools that monitoring network to detect internal or external cyber attacks. An Intrusion Detection System can observe and investigate system and user activities, recognize patterns of known attacks, identify abnormal network activity. General definition of IDS is about intrusions to network but for WSN it can be added that physical damages to sensor devices. Identifying sensor damage is important in order to serve fault tolerance and reliability [13]. With the high usage of Internet in our day today life, security of network has become the key foundation to all web applications, like online auctions, online retail sales, etc. Detection of Intrusion, attempts to detect the attacks of computer by examining different information records observed in network processes. This can be considered as one of the significant ways to effectively deal with the problems in network security.

In the presents scenario it is very challenging to maintain and used the classification of data and detection of network intrusion detection due to large and unknown number of attacker. Day to day come into new format and dynamic nature based attack pattern for the system i.e. host based and network based. To enhance the performance of intrusion detection system we used various types of techniques on the basis of their functionality and their behavior such as data mining, machine learning, evolutionary approach and swarm intelligence etc., where we adopted such types of techniques on the basis of requirement for the system and the types of attacker. Some authors also used soft computing approach with data mining and machine learning technique.

ISSN : 2455-9679
Impact Factor : 3.805

An Intrusion Detection System (IDS) inspects the activities in a system for suspicious behavior or patterns that may indicate system attack or misuse. There are two main categories of intrusion detection techniques; Anomaly detection and Misuse detection. The former analyses the information gathered and compares it to a defined baseline of what is seen as "normal" service behavior, so it has the ability to learn how to detect network attacks that are currently unknown. Misuse Detection is based on signatures for known attacks, so it is only as good as the database of attack signatures that it uses for comparison. Misuse detection has low false positive rate, but cannot detect novel attacks. However, anomaly detection can detect unknown attacks, but has high false positive rate. An intrusion detection system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches. A signature detection system identifies patterns of traffic or application data presumed to be malicious while anomaly detection systems compare activities against a "normal" baseline. On the other hand, a hybrid intrusion detection system combines the techniques of the two approaches. Both signature detection and anomaly detection systems have their share of advantages and drawbacks. The primary advantage of signature detection is that known attacks can be detected fairly reliably with a low false positive rate.

II. PROPOSED WORK

K-means is perhaps the most popular clustering method in metric spaces. Initially k cluster centroids are selected at random; k-means then reassigns all the points to their nearest centroids and recomputed centroids of the newly assembled groups. The iterative relocation continues until the criterion function, e.g. square-error converges. Despite its wide popularity, k-means is very sensitive to noise and outliers since a small number of such data can substantially influence the centroids. Other weaknesses are sensitivity to initialization, entrapments into local optima, poor cluster descriptors, and inability to deal with clusters of arbitrary shape, size and density, reliance on user to specify the number of clusters.

Nearest neighbor algorithms are among the "simplest" supervised machine learning algorithms and have been well studied in the field of pattern recognition over the last century. While nearest neighbor algorithms are not as popular as they once were, they are still widely used in practice, and I highly recommend that you are at least considering the k-Nearest Neighbor algorithm in classification projects as a predictive performance benchmark when you are trying to develop more sophisticated models.

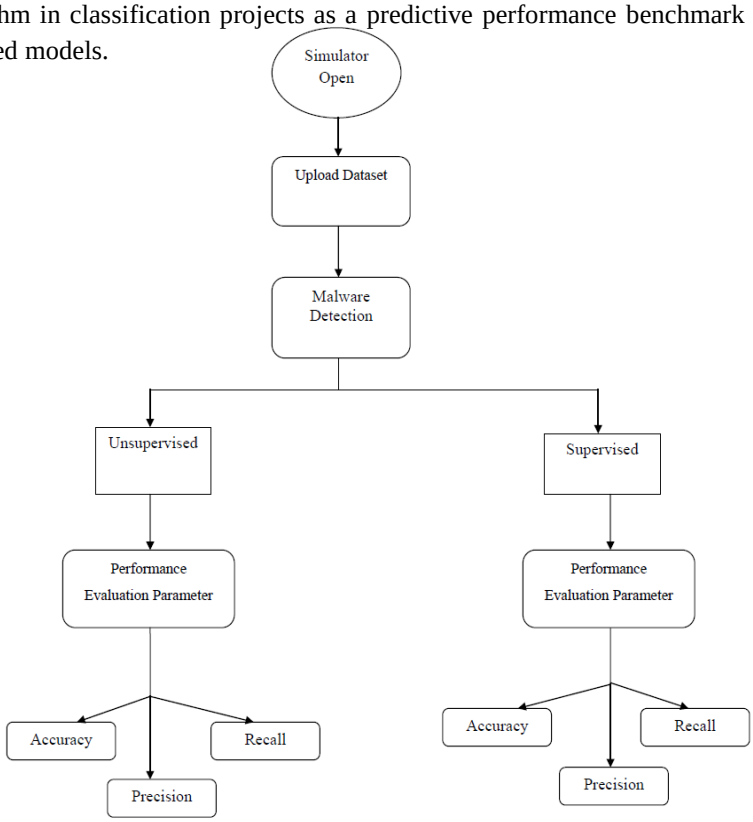

Fig 1: Shows that proposed model of malware detection using unsupervised and supervised techniques.

ISSN : 2455-9679
Impact Factor : 3.805

There are some steps we have to follow to implement this system are following:-

Step 1- begin the simulator for the experimental work.

Step 2- Upload the KDDCUP 99 dataset which is combination of normal and abnormal dataset.

Step 3- Apply the previous and proposed method with the dataset.

Step 4- After the applied unsupervised and supervised techniques we get the some performance evaluation parameter.

Step 5- we compare the all performance evaluation parameter values.

Step 6- Finally we compare the all performance parameters value with applied unsupervised and supervised techniques and choose the best optimal value.

Steps 7- If the performance parameter value is not optimal then we reject it and again go to step no. 3.

Step 8- stop the simulator and exit from the experimental environment.

III. EXPERIMENTAL WORK

In this dissertation we perform experimental process of proposed classification algorithm for anomaly detection system. The proposed method implements in mat lab 7.14.0 and tested with very reputed data set malware dataset. In this dissertation we measured detection accuracy, precision, recall and f-score for the unsupervised and supervised techniques, with using kdd cup dataset for the evaluation of given performance parameters.

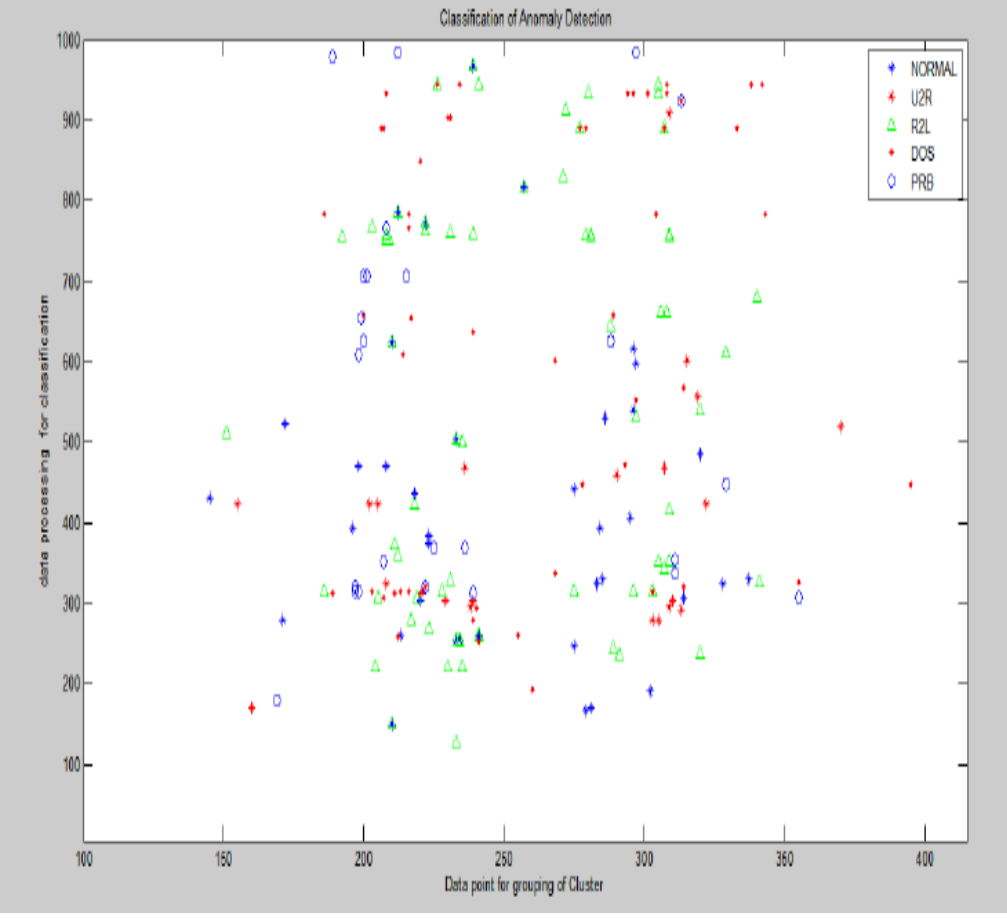

Fig 2: The above picture represents that the experimental result window with unsupervised techniques.

ISSN : 2455-9679
Impact Factor : 3.805

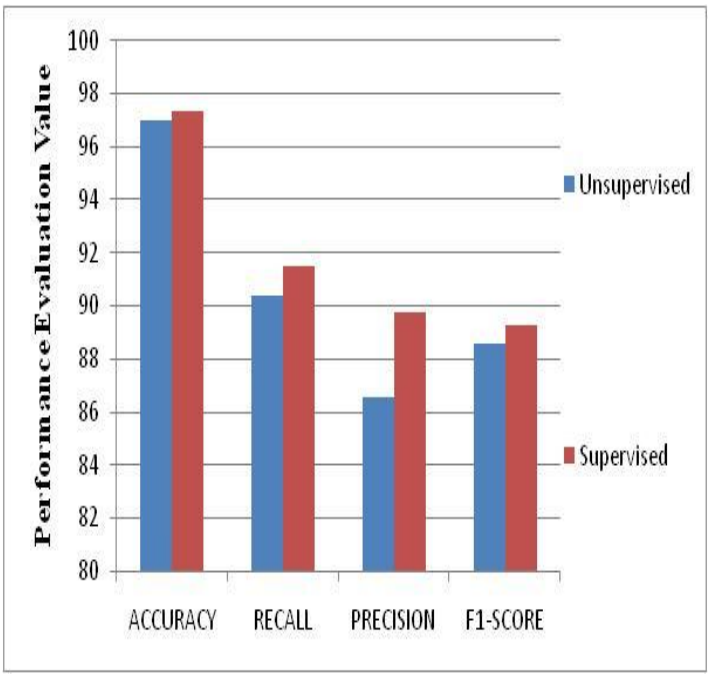

Fig 3: The above picture represents the performance parameter evaluation of given input value such as 0.25 for the unsupervised and supervised method.

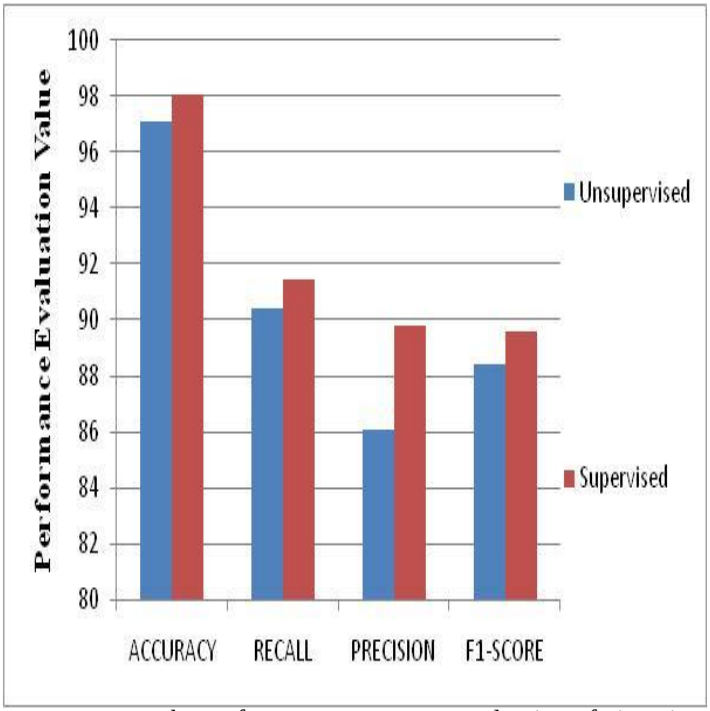

Fig 4: The above picture represents the performance parameter evaluation of given input value such as 0.45 for the unsupervised and supervised method.

IV. CONCLUSION

In this dissertation, proposed a model of anomaly based data categorization using k-means algorithm and supervised algorithm. From our experiments, the Proposed can detect known attack types with high accuracy and low false positive rate which is less than 1%. Here we presented new model for the analysis and detection of malware variants. The main idea of our approach is that the semantics of malware code, preserved across successive variants of the malware, can be used as a cluster algorithm for variants. As a step towards tackling the malware variant detection

ISSN : 2455-9679
Impact Factor : 3.805

problem using this idea. The proposed model categorized all data of KDDCUP99 is very accurately. The proposed method work in process of making group of attack very accurately, the selection process of genetic algorithm is very high for the generation of cluster. Our empirical result shows better performance in comparison with unsupervised techniques for anomaly detection.

REFERENCES

1. Sreeraj Rajendran , Wannes Meert, Vincent Lenders, Sofie Pollin, "Unsupervised Wireless Spectrum Anomaly Detection With Interpretable Features", IEEE Transactions on Cognitive Communications and Networking,, 2019, PP. 637-647.
2. Aniqua Baset, Christopher Becker, Kurt Derr, Samuel Ramirez, Sneha Kasera, Aditya Bhaskara, "Towards Wireless Environment Cognizance Through Incremental Learning", IEEE 2018, pp. 1-9.
3. Andrea Toma, Ali Krayani, Lucio Marcenaro, Yue Gao, Carlo S.Regazzoni, "Deep Learning for Spectrum Anomaly Detection in Cognitive mmWave Radios", IEEE 2020, pp. 1-7.
4. Sreeraj Rajendran, Vincent Lenders, Wannes Meert, Sofie Pollin, "Crowdsourced wireless spectrum anomaly detection", IEEE 2019, pp. 1-10.
5. Sreeraj Rajendran, Wannes Meert, Vincent Lenders, Sofie Pollin, "SAIFE: Unsupervised Wireless Spectrum Anomaly Detection with Interpretable Features" IEEE 2018, pp. 1-9.
6. Chieh-Yu Chen, Shi-Chung Chang, Da-Yin Liao, "Equipment Anomaly Detection for Semiconductor Manufacturing by Exploiting Unsupervised Learning from Sensory Data", Sensors 2020, pp. 1-27.
7. FuTao Ni, Jian Zhang, Mohammad N. Noori, "Deep learning for data anomaly detection and data compression of a long-span suspension bridge", Wiley 2019, pp. 1-17.
8. Luis Basora, Xavier Olive, Thomas Dubot, "Recent Advances in Anomaly Detection Methods Applied to Aviation", Aerospace 2019, pp. 1-27.
9. Andrea Toma, Ali Krayani, Muhammad Farrukh, Haoran Qi, Lucio Marcenaro, Yue Gao, Carlo Regazzoni, "AI-based Abnormality Detection at the PHY-layer of Cognitive Radio by Learning Generative Models", IEEE 2018, pp. 1-13.
10. Yijing Zeng, Varun Chandrasekaran, Suman Banerjee, Domenico Giustiniano UW-Madison, "A Framework for Analyzing Spectrum Characteristics in Large Spatio-temporal Scales', Association for Computing Machinery, 2019, pp. 1-16.
11. Raghavendra Chalapathy, Sanjay Chawla, "Deep Learning for Anomaly Detection: A Survey", IEEE 2019, pp. 1-50.
12. Sandamal Weerasinghe, Sarah M. Erfani, Tansu Alpcan, Christopher Leckie, Jack Riddle, "Detection of Anomalous Communications with SDRs and Unsupervised Adversarial Learning ", IEEE 2017, pp. 1-5.
13. Nistha Tandiya, Ahmad Jauhar, Vuk Marojevic, Jeffrey H. Reed, "Deep Predictive Coding Neural Network for RF Anomaly Detection in Wireless Networks", IEEE 2018, pp 1-7.
14. Changkun Liu, Xinrong Wu, Lei Zhu, Changhua Yao, Lu Yu, Lei Wang, Wei Tong, And Ting Pan, "The Communication Relationship Discovery Based on the Spectrum Monitoring Data by Improved DBSCAN", IEEE Access 2019, pp. 121793-121804.
15. Raman Singh, Harish Kumar, R.K. Singla "Review of Soft Computing in Malware Detection" IJCA, 2013. Pp 55-60.
16. Sarnsuwan N.; Wattanapongsakorn N.; and Charnsripinyo Ch."A New Approach for Internet Worm Detection and Classification" etworked Computing (INC), 6th International Conference, 2010. Pp 546-552.
17. Wang X.; Yu W.; Champion A.; Fu X.; and Xuan D "Detecting Worms via Mining Dynamic Program Execution" Authorized licensed use limited to: The Ohio State University, 2008. Pp 696-702.
18. Z. Gao, T. Li, J. Zhang, C. Zhao, and Z. Wang "A parallel method for unpacking original high speed rail data based on map reduce" Springer Berlin Heidelberg, vol. 124, 2012. Pp 59–68.

ISSN : 2455-9679
Impact Factor : 3.805

19. W. Zhu, N. Zeng, and N. Wang "Sensitivity, specificity, accuracy associated confidence interval and roc analysis with practical SAS implementations" in In Proceedings of the NorthEast SAS Users Group Conference NESUG10, 2010.

20. I. Aljarah and S. A. Ludwig "Parallel particle swarm optimization clustering algorithm based on map reduce methodology" in Proceedings of the Fourth World Congress on Nature and Biologically Inspired Computing (NaBIC'12), Mexico City, Mexico, November 2012, Pp 104–111.

21. J. Mazel, P. Casas, Y. Labit, and P. Owezarski "Subspace clustering, inter-clustering results association & anomaly correlation for unsupervised network anomaly detection" in Proceedings of the 7th International Conference on Network and Services Management, Paris, France, 2011, Pp 73–80.

22. Z. Li, Y. Li, and L. Xu "Anomaly intrusion detection method based on k-means clustering algorithm with particle swarm optimization" in Proceedings of the 2011 International Conference of Information Technology, Computer Engineering and Management Sciences. Washington, DC, USA: IEEE Computer Society, 2011, Pp 157–161.

23. Y. Ye, D.Wang, T. Li, and D. Ye "IMDS: Intelligent malware detection system" In Proccedings of ACM International conference on Knowledge Discovery and Data Mining, 2007, Pp 1043-1047.

24. Y. Ye, D.Wang, T. Li, D. Ye and Q. Jiang "An intelligent PE malware detection system based on association mining" Journal in Computer Viorology, 2008. Pp323-334.

25. L.Jing, M.K.Ng, J.Z.Huang "An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data " IEEE Transactions on Knowledge and Data Engineering, 2007, Pp 1-16.

26. P. Ferguson "Observations on emerging threats" in USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET), Apr. 2012.

27. K. Thomas and D. Nicol, "The koobface botnet and the rise of social malware" in IEEE Int. Conf. Malicious and Unwanted Software (Malware 10), 2010, Pp 63–70.