IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Survey on Intrusion Detection System Based on Feature Classification and Selection

Madhavi Dhingra

Dept. of Computer Science and Engineering, Amity University Madhya Pradesh, Maharajpura Dang, Gwalior (MP)-474005

*Corresponding Author: madhavi.dhingra@gmail.com, Tel.: +91-9229125600

DOI: https://doi.org/10.26438/ijcse/v7i3.399403 | Available online at: www.ijcseonline.org

Accepted: 10/Mar/2019, Published: 31/Mar/2019

***Abstract*—** Wireless networks are facing variety of attacks nowadays. To prevent from such attacks, a few Intrusion Detection frameworks are being created to distinguish and evacuate the attacks. Intrusion detection frameworks need to manage huge information having duplicate and excess features that require moderate training and testing processes leading to higher resource utilization and poor discovery rate. The performance of the Intrusion detection frameworks depend on the accuracy of the predicted attacks. Various performance parameters are to be considered for determining accuracy of a framework. The whole process is highly dependent on the network features and thus, Feature Classification is a vital issue in intrusion detection process. This paper covers the importance of feature selection, the common feature selection methods and various feature classification approaches that have been used in the field of Intrusion Detection System. The paper has also revised about the different researches that had taken place in the relevant field.

***Keywords*—**Intrusion detection System, Feature Classification, Feature selection, Wireless Attacks

I. INTRODUCTION

IDS can be a mix of programming and hardware. The vast majority of intrusion detection systems play out their errand progressively. In any case, there are likewise intrusion detection systems that don't work progressively, due to performing examination for legal reviews. There are some intrusion detection systems that respond to interruptions progressively way. This response typically forces to diminishing the misfortune and harm by ending a system association and different methodologies. For those intrusion detection systems need to do reviewing information, it is troublesome even by utilizing PC's capacity since distinguishing suspicious conduct for approaching information even in little systems are confounded procedure.
Features may contain false connections, which impede the way toward identifying interruptions. Further, a few features might be excess since the data they include is contained in different features. Additional features can build calculation time, and can affect the precision of intrusion detection system. Determination of features improves classification process via finding out the subset of features, which best groups the training data. The features under thought rely upon the sort of intrusion detection system, for instance, network based intrusion detection system will break down system related data. It isn't known which of these features are repetitive or unessential for intrusion detection system and which ones are applicable or fundamental for intrusion detection system. There does not exist any model or process that catches the connection between various features. On the off chance that such a model existed, the interruption discovery procedure would be basic and direct.

From the past years, a few intrusion detection systems have been prescribed that mostly target rule-based frameworks, in light of the fact that their execution relies upon the principles distinguished by the security specialists [1]. The volume of system traffic is broad along these lines thus encoding rules is for the most part lacking just as moderate. Thus the rule based frameworks also require major changes.

These days, one genuine test for intrusion detection system is feature selection performed by using system traffic information. Researchers uses different feature selection techniques for the classification problems [2]. In any case, a few calculations are sensitive to feature selection as the original information obtained from the configuration of system is not reasonable for detection. In this manner, feature selection is a noteworthy strategy for improving characterization precision, reducing the superfluous or repetitive information and contributing a superior comprehension of important features and the basic procedure that created the datasets.

There is a need to overcome certain issues, for example, highlight repetition, high-dimensional highlights, and overfitting. Also, another issue for intrusion detection system is managing imbalanced datasets, for example, most of the

examples have a place with Probe and denial of Service (DoS) assaults, while very few examples are related with client to-root (U2R) or remote-to-client (R2L) attacks because of this issue, frequently the classifier gets the excess and one-sided tests, while in certifiable the minority assaults are normally more risky than larger part assaults. A solitary IDS can look at huge measures of data containing excess and incorrect features, while at this stage, IDS experiences a few challenges, for example, noise and expanded classifier time to handle previously mentioned issues. A successful intrusion detection system is expected to diminish false alert rates and in the meantime, it is additionally required to be powerful in recognizing the attacks and oversee high recognition rates and decrease the time.

This paper covers different feature selection mechanisms in Section II, Implemented feature selection approaches in the field of Intrusion detection system in Section III and concluded in Section IV.

II. FEATURE SELECTION MECHANISMS

The feature selection process involves four major steps that include Feature set generation, Feature set evaluation, Stopping criteria and the result [3].

Subset generation is a kind of heuristic search. Each state of the search space domain serves as the candidate subset that is evaluated. This phase is based on two major issues, the first one is to decide the starting research point, which has major impact over the results of the research. Search can start with the empty subset which successively adds the other features one by one , this is called the forward feature section approach. Another way of starting search mechanism is to start the search with full set of features and then deleting the features one by one, this is known as backward feature selection approach. A third bidirectional approach also exists which can add or remove the features at the same time.
The other issue in feature subset generation is determining the search mechanism. The searching strategy may follow complete search or sequential search or random search.
Feature selection algorithm performs search procedure on all the possible feature subsets and tries to find the most efficient, optimal subset. The process will go on until some threshold is reached. Thus, to stop the procedure, threshold is determined according to the given criteria and then a number of iterations are performed on the feature subset.

The feature selection algorithms are classified into three categories [4].:

1. Embedded Approaches - Embedded approach uses filter selection as well as learning algorithm both at the same time. The procedure identifies the useful attributes and irrelevant attributes during the learning algorithm itself. These approaches achieve higher efficiency in comparison to other feature selection approaches. Example of such approach is Decision trees etc.
2. Wrapper Approaches - Wrapper approaches first perform the learning procedures and then identifies the optimal subset features based on the result of the learning procedure. These procedures include huge amount of computations in comparison to Filter techniques.
3. Filter Approaches - Filter approaches filter the feature subset and determine the best features. These best features are applied to the learning algorithm for building a learning model.

III. IMPLEMENTED FEATURE SELECTION APPROACHES IN DEVELOPING IDS

To deal with the digital attacks, remarkable researches have been done for exploring and developing Intrusion detection frameworks deal with the malware and vulnerabilities. From the year 2000s, there have been numerous effective applications that have combined data mining and machine learning strategies for the development of Intrusion Detection System. After that, numerous Data mining and machine learning algorithms were explicitly intended for the reason. Data Mining techniques inspect the significant data inside large volume of information by logically finding fundamental real patterns, examples, and relationship from the information as revealed in [5,6]. Artificial neural network is also used for multiclass issues for the developing intrusion detection system by using a multilayer feed-forward kind of neural network. It has used back-propagation method to anticipate intrusions [7]. Number of techniques have been utilized in the field of intrusion detection, for example, decision Trees, SVM and random forests [8]. These techniques have performed the operations on a set of patterns, and the results have demonstrated that particular classification methods are explicitly proficient for a given attack class while others were not. In addition, a multiclassifier model for Intrusion Detection System [9] was also utilized by using random forests for network intrusion detection systems. Sone naive bayes feature classification approaches have also been developed by researchers[10,11]. In the area of system security, machine learning , data mining and feature selection are performing key roles because numerous specialists are using them to enhance the execution of learning algorithms in the various fields, for example, content mining, PC vision, and image processing and so on [12]. Feature selection is generally utilized for some reasons, for example, expanded effectiveness of the learning algorithm, accomplishing a high precision rate and getting effortlessness for classification issues [13]. Also, feature selection decides suitable subset from the first dataset so as to limit the effect of superfluous and excess features without reducing the precision of the classifier. Overall, a few issues are important to be considered- first, repetition in datasets

and mix process for feature selection systems through any learning algorithm and second, a excess element choice as well as incapable feature determination from datasets—these issues lead to troublesome stages for any learning algorithm. The very fast decision tree has[14], utilized information gain for feature selection and it incorporates many improved forms while preparing the model.

One of the popular and effective feature selection technique, for example, filter is used to rank all the features with no classifier and wrapper[15]. The wrapper has added great execution in little sets and the filter method is more affordable from a computational perspective [16]. In spite of the fact that, the relationship between feature selection and calculation isn't yet very much overseen but there is requirement in the area. Subsequently, the regular learning methods, for example, support vector machine [17], boosting [18], and sparse logistic regression is more powerful in the area of feature selection and feature classification problems.

Gisung Kim had invented a new hybrid technique that progressively works on both misuse identification and anomaly detection in a deteriorated structure[19]. It has used the C4.5 choice tree and one-class support vector machine to upgrade the capability of anomaly detection. C4.5 choice tree does not form clusters, which can corrupt the profiling capacity along these lines truncating the effectiveness of the framework.

Shi-Jinn Horng had given an intrusion detection framework which combines clustering process, a basic feature selection algorithm and the support vector machine[20]. The resultant classifiers indicated better execution and reduced the training time over the SVM classifiers. This methodology gives better execution regarding precision in contrast with alternate network based intrusion detection system. But its main limitation is that it just identifies Dos and Probe assaults and cannot detect U2L and R2L assaults.

A hybrid intelligent decision technique was give by Mrutyunjaya Panda[21] by including guided learning strategies with a classifier to so as to recognise network attacks. The outcomes demonstrate that there is no single best calculation in comparison to others in all circumstances. Juan Wang [22] exhibited a decision tree based intrusion detection system. It has used information gain ratio. But, in this methodology the error rate continues as before.

Hong Kuan Sok [23] had utilized the ADTree calculation for feature reduction. It provides great performance over classification. The classification process has been rearranged and the speed has been improved radically because of the decreased activities required to perform the classification.

Tavallaee had used KDD CUP 99 data set and after examining whole KDD dataset it demonstrated that there were two vital issues in the informational collection which influenced the execution of frameworks, and hence results in an exceptionally poor understanding of anomaly detection approaches[24]. NSL-KDD was proposed to overcome some of the issues. The proposed dataset has also experienced a few issues but the researchers trust that the dataset still can be utilized as a powerful benchmark tin studying various intrusion detection strategies.

In the same field, some computational knowledge approaches were used for developing effective intrusion detection. These strategies incorporate various soft computing techniques.

Previously, two sorts of agents have been proposed ie. simple and multi-agents. Simple agents can detect the environment and follow up on them. Bakar [25] had proposed a simple agent based methodology for intrusion detection which had used rough set-based classification technique but the procedure is computationally costly, particularly in its computation stage.

Xiaodong Zhu[26] had introduced an intrusion detection framework named multi-agent-based intelligent intrusion detection system. This framework has an adaptive learning agent that analyses both system based and host-based review data, and can adapt more than one method of data mining. The test results demonstrate that their framework has high self-adjusting capacity and intelligence.

An incorporated structure has also been proposed[27] to design a mobile agent based system for enhancing the security of self-governing system. This structure offered two benefits - a secure agent based administration framework and the ability of accomplishing enhanced network functionalities.

Mobile agents could encourage the usage of attack safe IDS structures [28]. Agents should move when detect threat or suspicious action and work non concurrently. Besides, agents are agreeable to hereditary decent variety, which likewise maintained a strategic distance from attacks went for bypassing the known and stable discovery systems of intrusion detection system.

Some other feature classification techniques were also implemented like neuro-fuzzy inference system standard [29], neural network based model for anomaly detection[30], Markov show identification strategy[31] and a partial swarm optimisation based intrusion detection framework(particle swarm optimization)[32]. Several other mechanisms have been given by researchers based on certain requirements, each of which has its own benefits and limitations[33,34].

IV. CONCLUSION

Wireless Networks are facing major security issues with the advancement of technology day by day. Various Intrusion detection mechanisms have been developed for avoidance, detection and removal of attacks. An intrusion detection system must be lightweight and must guarantee the detection of attacks. Wireless networks deal with huge amount of data that include various features of the network. Some features are irrelevant while some other are redundant. The unnecessary features can degrade the performance of IDS and the machine learning approach. It affects the training and testing methods of the machine learning.

The feature selection or classification algorithms play an important role in identifying the important and the necessary features that can help in developing an efficient intrusion detection system. This paper has reviewed about several IDS that have been implemented using different feature classification approaches based on which different quality of results have been achieved.

REFERENCES

1. Snort Intrusion Detection System. 2006. Available online: http://www.snort.org, 10 October 2017.
2. Li, Y.; Wang, J.-L.; Tian, Z.-H.; Lu, T.-B.; Chen, Y. Building lightweight intrusion detection system using wrapper-based feature selection mechanisms. Comput. Secur., 28, pp. 466–475, 2009.
3. M. Dash, H. Liu, "Feature Selection for Classification," Intelligent Data Analysis, Elsevier, pp. 131-156, 1997.
4. An Introduction to Feature Selection by Jason Brownlee on October 6, Machine Learning Process, 2014.
5. Hodo, E.; Bellekens, X.; Hamilton, A.; Dubouilh, P.; Iorkyase, E.; Tachtatzis, C.; Atkinson, R. Threat analysis of iot networks using artificial neural network intrusion detection system. In Proceedings of the IEEE International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, pp. 11–13 May 2016.
6. Hodo, E.; Bellekens, X.; Hamilton, A.; Tachtatzis, C.; Atkinson, R. Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey; Cornell University Library: Ithaca, NY, USA, 2017.
7. Brifcani, A.; Issa, A. Intrusion detection and attack classifier based on three techniques: A comparative study. Eng. Technol. J., 29, pp. 368–412, 2011.
8. Roopadevi, E.; Bhuvaneswari, B.; Sahaana, B. Intrusion Detection using Support Vector Machine with Feature Reduction Techniques. Indian J. Sci. 23, pp. 148–156, 2016,.
9. Zhang, J.; Zulkernine, M. A hybrid network intrusion detection technique using random forests.In Proceedings of the IEEE First International Conference on Availability Reliability and Security (ARES'06), Vienna, Austria, 20–22 April 2006.
10. Farid, D.M.; Zhang, L.; Hossain, M.A.; Strachan, R. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. Expert Syst. Appl., 41, pp.1937–1946, 2014.
11. Koc, L.; Mazzuchi, T.A.; Sarkani, S. A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. Expert Syst. Appl., 39, pp.13492–13500, 2012.
12. Farid, D.M.; Harbi, N.; Rahman, M.Z. Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection; Cornell University Library: Ithaca, NY, USA, arXiv preprint, 2010.
13. Fahad, A.; Zahir, T.; Ibrahim, K.; Ibrahim, H.; Hussein, A. Toward an efficient and scalable feature selection approach for internet traffic classification. Comput. Netw., 57, pp. 2040–2057, 2013.
14. Al-mamory, S.O.; Jassim, F.S. On the designing of two grains levels network intrusion detection system. Karbala Int. J. Mod. Sci., 1, pp. 15–25, 2015.
15. Yang, J.; Olafsson, S. Optimization-based feature selection with adaptive instance sampling. Comput. Oper. Res., 33, pp. 3088–3106, 2006.
16. Sánchez-Maroño, N.; Alonso-Betanzos, A.; Calvo-Estévez, R.M. A wrapper method for feature selection in multiple classes datasets. In International Work-Conference on Artificial Neural Networks; Springer: Berlin/Heidelberg, Germany, 2009.
17. Sani, R.A.; Ghasemi, A. Learning a new distance metric to improve an svm-clustering based intrusion detection system. In Proceedings of the IEEE International Symposium on Artificial Intelligence and Signal Processing (AISP), Mashhad, Iran, 3–5 March 2015.
18. Sarikaya, R.; Hinton, G.E.; Deoras, A. Application of deep belief networks for natural languageunderstanding. IEEE/ACM Trans. Audio Speech Lang. Process., 22, pp. 778–784, 2014.
19. Gisung Kim and Seungmin Lee, A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection With Misuse Detection, ELSEVIER, Expert Systems with Applications vol. 41 pp. 1690 – 1700, 2014.
20. Shi-Jinn Horng and Ming-Yang Su, "Novel Intrusion Detection System Based On Hierarchical Clustering and Support Vector Machines", ELSEVIER, Expert Systems with Applications. pp. 38 306-313, 2011.
21. Mrutyunjaya Panda and Manas Ranjan Patra, "A Comparative Study Of Data Mining Algorithms For Network Intrusion Detection", First International Conference on Emerging Trends in Engineering and Technology, pp 504-507, IEEE, 2008.
22. Juan Wang, Qiren Yang, Dasen Ren, "An intrusion detection algorithm based on decision tree technology", In the Proc. of IEEE Asia-Pacific Conference on Information Processing, 2009.
23. Hong Kuan Sok et.al, "Using the ADTree for Feature Reduction through Knowledge Discovery" Instrumentation and Measurement Technology Conference (I2MTC), IEEE International ,pp 1040 – 1044, 2013.
24. Avallaee M, Bagheri E, Lu W, Ghorbani A. "A detailed analysis of the KDD CUP 99 data set", 2009 IEEE Symposium on Computational intelligence for security and defense applications,,pp 1-6, 2009.
25. Bakar AA, Othman ZA, Hamdan AR, Yusof R, Ismail R: An Agent Based Rough Classifier for Data Mining. Eighth International Conference on Intelligent Systems Design and Applications , vol 1. IEEE Computer Society, Washington, pp.145-151, 2008.
26. Zhu X, Huang Z, Zhoul H: Design of a Multi-agent Based Intelligent Intrusion Detection System. IEEE International Symposium on Pervasive Computing and Applications. IEEE, Amsterdam, pp.290-295, 2006.
27. Fonk C-h, Parr GP, Morrow PJ: Security schemes for Mobile Agent based Network and System Management Framework. *J. Networks Syst. Manag. Springer*, 19: 232-256, 2011.
28. Mell P, Marks D, McLarnon M: A Denial of service resistant intrusion detection architecture. Comput Networks J Elsevier, Amsterdam, 2000.
29. Mar J, Yeh Y-C, Hsiao I-F: An ANFIS-IDS against Deauthentication DOS Attacks for a WLAN Taichung, 17–20 October 2010. IEEE, Amsterdam, pp. 548-553, 2010.
30. Linda O, Vollmer T, Manic M: Neural network based intrusion detection system for critical infrastructures. In *Proceedings of IEEE*

International Joint Conference on Neural Networks, Georgia. IEEE, Amsterdam, pp.102-109, 2009.
31. Li Y, Wang R, Xu J, Yang G, Zhao B: Intrusion detection method based on fuzzy hidden Markov model. Sixth IEEE International Conference on Fuzzy Systems and Knowledge Discovery, vol 3. IEEE, Piscataway; pp. 470-474, 2009.
32. Tian J, Gu H: Anomaly detection combining one-class SVMs and Particle swarm optimization algorithms, Nonlinear Dynamics. Volume 61. Springer, Berlin; 2010:303-310.
33. Karmore, Preetee K.; Bodkhe, Sonali T, "A Survey on Intrusion in Ad Hoc Networks and its Detection Measures," Proc. International Journal on Computer Science & Engineering, vol. 3, Issue 5, pp. 1896-1903, 2011.
34. Aakanksha Kori, Harsh Mathur,"A Performance Evaluation of Intrusion Detection system to get better detection rate using ANN Technique", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol.2, Issue 5, pp.217-223, 2017.

Authors Profile

Ms. Madhavi Dhingra pursed Master of Computer Applications from Medicaps Institute affiliated by RGPV and Masters of Technology in Computer Science and Engineering from UPTU University in 2014 and currently pursuing Ph.D in Computer Science Engineering. She is a life member of CSI and IET. She has published more than 10 research papers in reputed international journals and conferences including IEEE and it's also available online. Her main research work focuses on Network Security, Information Security and Privacy,Computational Intelligence based education. She has more than 8 years of teaching experience.

