IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Detailed Survey on Intrusion Detection System based on NSL-KDD Dataset using Various Approach

Ms.Shivangi Soni, Prof. Chetan Gupta, Prof. Shivendra Dubey

M.Tech Scholar,Dept. of CSE, Asst. Prof.,Dept. of CSE,Asst. Prof., Dept. of CSE

SIRTS, Bhopal, SIRTS, Bhopal, SIRTS, Bhopal

**Abstract** — Focusing on the deficiencies of conventional intrusion detection model, several researchers propose the intrusion detection system based on data mining technology. It solves the problem that self-adaptability of the system is poor, and the conditions of misreport or omission are also further improved. However, as far as mass data are concerned, more and more resources need to be consumed in the intrusion detection system based on data mining technology, and the detection speed gets slower and slower. One more thing arises when we think of the above approach when we create the clusters of same thing for easy generation of patterns in future we not always find the good result. Because each time the support is different and according to the support pruning value is also changed. So we need to process the rule generation dynamic, which solves the above problem and based on that we achieve different results on different conditions.

**Keywords: KDD, IDS, Dos, Probe, R2L, U2R.**

## I. Introduction

The process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusions is known as Intrusion detection. Intrusion detection is classified into two types: misuse intrusion detection and anomaly intrusion detection.

1. Misuse intrusion detection uses well-defined patterns of the attack that exploit weaknesses in system and application software to identify the intrusions. These patterns are encoded in advance and used to match against the user behavior to detect intrusion.

2. Anomaly intrusion detection uses the normal usage behavior patterns to identify the intrusion. The normal usage patterns are constructed from the statistical measures of the system features, for example, the CPU and I/O activities by a particular user or program. The behavior of the user is observed and any deviation from the constructed normal behavior is detected as intrusion. We have two options to secure the system completely, either prevent the threats and vulnerabilities which come from flaws in the operating system as well as in the application programs or detect them and take some action to prevent them in future and also repair the damage. It is impossible in practice, and even if possible, extremely difficult and expensive, to write a completely secure system. Intrusion Detection (ID) is to collect and analyze information from several key points of computer networks or computer systems. It can check whether there is security policy violation behavior or signs of attack in the network or system. Intrusion Detection System (IDS) [2] is consisted of intrusion detection software and hardware. Current IDS use misuse detection and anomaly detection

methods to analyze intrusion detection. It can be achieved online testing and after testing.

Characteristics of ID are dynamic, active, real-time and complex. How to extract user interest information from massive data? Many scholars put forward a lot of new ideas and algorithms. Columbia University, USA, Wenke Lee [3] first proposed the application of data mining techniques to intrusion detection system. ID based on data mining algorithms has become hot research and has many theoretical results [4, 5, 6]. Mainly by using data classification, association analysis and sequential pattern mining to handle vast amounts of security audit data and extract security related Behavior, generate intrusion detection rules and establish anomaly detection model. The application of data mining techniques to IDS may avoid the poor real-time and low efficiency problem. It can adapt to the rapid development of network intrusion detection technology and improve the detection efficiency and reduce human intervention.

An intrusion detection system normally consists of three functional components. The first component of an intrusion detection system, also known as the event generator, is a data source. Data sources can be categorized into four categories namely Host-based monitors, Network-based monitors, Application-based monitors and Target-based monitors. The second component of an intrusion detection system is known as the analysis engine. This component takes information from the data source and examines the data for symptoms of attacks or other policy violations. The analysis engine can use one or both of the following analysis approaches:

**Misuse/Signature-Based Detection:** This type of detection engine detects intrusions that follow well-known patterns of attacks (or signatures) that exploit known software vulnerabilities. The main limitation of this approach is that it only looks for the known weaknesses and may not care about detecting unknown future intrusions.

**Anomaly/Statistical Detection:** An anomaly based detection engine will search for something rare or unusual. We analyses system event streams, using statistical techniques to find patterns of activity that appear to be abnormal. The primary disadvantages of this system are that they are highly expensive and they can recognize an intrusive behavior as normal behavior because of insufficient data. The third component of an intrusion detection system is the response manager. In basic terms, the response manager will only act when inaccuracies (possible intrusion attacks) are found on the system, by informing someone or something in the form of a response.

## II. Literature Survey

In this paper [1] an Intrusion Detection is developed based on Crow SearchOptimization algorithm with Adaptive Neuro-FuzzyInference System. For detecting the abnormalities present in the network or system, the intrusion detection system (IDS) is used. The ANFIS is the combination of fuzzy interference system and artificial neural network, and to enhance the performance of the ANFIS model the crow search optimization algorithm is used to optimize the ANFIS. The NSLKDD data set was used to validate the performance of intrusion detection of the proposed model. The results of the intrusion detection based on the NSL-KDD dataset was better the previous approaches.

Liang et al. [9] used a hybrid positioning strategy based on a multi-agent system for the intrusion detection system. This system includes a data acquisition module, a data management module, an analysis module and a response module. In this study, an algorithm for a deep neural network for intruder detection is used to implement the analysis module. The results show the effectiveness of deep learning algorithms for detecting transport-layer attacks.

A UNSW-NB15 dataset was generated by Moustafa et al. [10] for intruder detection. Nine forms of current attack modes and new standard traffic patterns are found in this dataset. To distinguish between normal and abnormal observations, it contains 49 attributes that include host-flow and network packet control. We demonstrate the complexity of the UNSW-NB15 dataset in three ways in this article. Next, it describes the mathematical study of findings and attributes. Second, it provides the analysis of characteristic associations. Third, in order to test the difficulty in terms of accuracy and false alarm rate (FAR), five existing classifiers are used. The results of experiments indicate that UNSW- NB15 is more complex than KDD99 and a new dataset is considered for the assessment of NIDS.

Zhao et al. [11] proposed an intrusion detection method using a deep belief network (DBN) and a probabilistic neural network (PNN). First, the raw data is converted into small data, while the essential attributes of the raw data are preserved using DBN's nonlinear learning ability. Second, a swarm of particle optimization algorithms are used to improve learning performance to optimize the number of nodes with hidden levels per level. Subsequently, PNN is used to classify low dimensional data.

Chuanlong Yin et al. [12] examined the model's performance in binary classification and the classification of different classes. The result analysis is very suitable for modeling a classification model with great precision and that its performance is superior to conventional classification methods for machine learning in binary and multi-class classification.

Yuan et al. [13] proposed a DDoS attack detection approach based on deep learning (Deep Defense). The deep learning approach can automatically extract high-level functions from low-level functions and achieve powerful representation and conclusion. A recurring deep neural network project is proposed to learn patterns from network traffic sequences and to track network attack activity. Experimental results show that the model works better than traditional machine learning models, as it reduces the error rate from 7.517% to 2.103% compared to conventional machine learning methods in the larger dataset.

Ma T et al. [14] proposed a new approach called KDSVM, which took advantage of k- mean techniques and the advantage of learning functionality with a deep neural network (DNN) model and a support vector machine (SVM) classifier to detect intrusion networks. KDSVM consists of two phases. In the first step, the data set is divided into k subsets as a function of each sampling distance from the cluster centers of the k-means approach and, in the second step, the test data set is far from the same cluster center and entered in the DNN model with SVM.

Feng et al. [15] implemented intrusion detection in the network by using a machine- learning data classification algorithm. The basic task is to identify network activities as regular or abnormal connection logs, reducing classification errors within a network protocol. Each classification model has its own strengths and shortcomings, including vector machine methods, although different classification models have been developed for the network intrusion detector.

Ali et al. [16] proposed a hybrid machine learning technique for detecting network intrusions, which is based on a combination of K-medium clusters and Sequential Minimal Optimization (SMO) classification.

Laftah et al. [17] proposed a modified K-mean algorithm to create a high-quality training dataset that greatly improves the performance of classifiers. The modified K-mean is used to create new small training datasets that represent the complete set of original training data, significantly reduce the time spent training classifiers and improve the performance of the intrusion detection system.

## III. Problem Domain

Some behaviors in intrusion instances are similar to normal and other intrusion instances as well. In addition, a lot of algorithms including K-Means are unable to correctly distinguish intrusion instances and normal instances. In 2012 LI Yin–huan [7] proposed an Intrusion detection model which is based on FP-Growth tree. Their proposed algorithm provides good experimental result. But in the data analysis phase the rule is fixed and if the rule generation is fixed the data clustering which will be generated for pattern finding is fixed. But in today’s environment we notice drastic change in s user behavior or customer behavior day by day. So data analysis is based on random selection or on the basis of dynamic behavior. By which the pattern recognizes is change and more accurate for analysis. This provides us to determine abnormal or unknown patterns and take appropriate response measures. A record is judged as known attacks if it is matched in rule database. The alarm unit of Administrator Processing Module (APM) will be triggered to alarm. New intrusion detection rule is added to Rule Database (RD) dynamically. We also use alike patterns for making clusters which reduces the time also. This also help us to Eliminate redundant data and noise data which can reduce the false positive data which will be input to association analysis module. This provides a better association with homogeneous element. Our Dynamic approach also provides a chance when the item set is frequent and it may change its cluster. So the accuracy of finding is improved. In 2012 Z. Muda et al. [8] suggest that Anomaly detection is one of intrusion detection system. Current anomaly detection is often associated with high false alarm with moderate accuracy and detection rates when it's unable to detect all types of attacks correctly. So to overcome this problem, they propose a hybrid learning approach through combination of K-Means clustering and Naïve Bayes classification. Their proposed approach will be clustering all data into the corresponding group before applying a classifier for classification purpose. But the patter data will be change according the sample inserted in the day by day basis, so the above approach behaves static function.

## IV. Propose Methodology

In our approach we used data mining with clustering Algorithm for gathering or cluster the same set of items. When a dataset is clustered, every point is assigned to some cluster, and every cluster can be characterized by a single reference point, usually an average of the points in the cluster. Any particular division of all points in a dataset into clusters is called a partitioning. One of the most familiar applications of clustering is the classification of plants or animals into distinct groups or species. However, the main purpose of clustering Landsat data is to reduce the size and complexity of the dataset. Data reduction is accomplished by replacing the coordinates of each point in a cluster with the coordinates of that cluster's reference point. Clustered data require considerably less storage space and can be manipulated more quickly than the original data. The value of a particular clustering method will depend on how closely the reference points represent the data as well as how fast the program runs. After the clustering approach we find the outcomes according to the set, the same set of data is characterize in the same place and other are placed in different points. We then categorized according to the behavior of the IDS. We apply dynamic rule generation so that each time it is categorized differently so that we achieve better and accurate result.

## V. Conclusion

The integration comprehensive research fields like data mining on the one hand and the younger and still evolving with Intrusion detection system. The application of data mining techniques to IDS is one important direction for future development of intrusion detection. Selecting appropriate data mining algorithms and designing IDS model are effective measures in order to improve system detection performance.

## VI. References

[1] S Manimurugan , Al-qdah Majdi , Mustaffa Mohmmed, C Narmatha , R Varatharajan "Intrusion Detection in Networks using Crow Search Optimization algorithm with Adaptive Neuro-Fuzzy Inference System", "Microprocessors and Microsystems" Elsevier 6 September 2020

[2] QING Si-han, JIANG Jian-chun, MA Heng-tai, etc. "Research on Intrusion Detection Techniques: A Survey". Communications Journal, pp.19-29, July 2004.

[3]Lee W, Stolfo S, Mok k. "Mining Audit Data to build Intrusion Detection Models". In Proc. of the International Conference on Knowledge and Data Mining, Aug. 1998.

[4] WU Yu-gang, QIN Yong, SONG Ji-guang, etc. "Research Overview of Intrusion Detection Algorithms based on Association Rules". Computer Engineering and Design, Vol.32.No.3. Mar. 2011.

[5]Eleazar Eskin, Andrew Amold, Michael Prerau, etc. "A Geometric framework for unsupervised Anomaly Detection: Detecting Intrusion in Unlabeled Data". Kluwer: Data mining for Security Application (DMSA-2002), 2002.

[6]LI Yang. "Application of K-means Clustering Algorithm in Intrusion Detection". Computer Engineering, Vol.33No. 14, pp.l54-156. 2007.

[7]LI Yin–huan, "Design of Intrusion Detection Model Based on Data Mining Technology", 2012 International Conference on Industrial Control and Electronics Engineering.

[8]Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification", 2011 7th International Conference on IT in Asia (CITA).

[9] Chao Liang, Bharanidharan Shanmugam, Sami Azam, Mirjam Jonkman, Friso De Boer, Ganthan Narayansamy, "Intrusion Detection System for Internet of Things based on a Machine Learning approach", International Conference on Vision Towards Emerging Trends in Communication and Networking, IEEE, 2019.

[10] 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, 2015.

[11] Zhao, G.; Zhang, C.; Zheng, L. "Intrusion Detection Using Deep Belief Network and Probabilistic NeuralNetwork", In Proceedings of the 2017 IEEE International Conference on Computational Science andEngineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC),Guangzhou, China, 21– 24 July 2017; Volume 1, pp. 639–642.

[12] Yin C et al, "A Deep Learning Approach for Intrusion Detection using Recurrent Neural Networks", IEEE Access 5:21954–2196, 2017.

[13] Yuan X, Li C, Li X, "Deep Defense: Identifying D-DoS Attack via Deep Learning", IEEE international conference on smart computing (SMARTCOMP), 2017.

[14] Ma T et al, "A Hybrid Methodologies for Intrusion Detection Based Deep Neural Network with Support Vector Machine and Clustering Technique", International conference on frontier computing. Springer, 2016.

[15] Feng, W., "Mining Network data for Intrusion Cetection through Combining SVMs with Ant Colony Networks", Future Gener. Comput. Syst., 2014, 37, 127–140.

[16] 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.

[17] 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.