ISSN:0975-9646

Intrusion Detection System on KDDCup99 Dataset: A Survey

Sonali Rathore
Research Scholar
Truba Institute of Engineering & Information Technology
Bhopal, India

Prof. Amit Saxena
Assistant Professor
Truba Institute of Engineering & Information Technology
Bhopal, India

Dr. Manish Manoria
Professor
Truba Institute of Engineering & Information Technology
Bhopal, India

Abstract — Intrusion Detection provides a technique of identifying unwanted packets so the attacks or harm made from these intrusions can be minimize. Since various techniques are implemented for the discovery and categorization of intrusions. Some of the IDS is implemented on the network based and some are implemented for Host based. Here in this paper a survey of all the techniques implemented for the discovery and categorization of intrusions on KDDCup 99 dataset is discussed, so that by identifying their various issues a new and efficient technique is implemented which can classify and detection intrusions in KDDCup 99 dataset.

1. INTRODUCTION

Security is an important issue in the web log data where the flow of packets contains a number of intruders. Intrusion detection can be detected using misuse detection or anomaly detection. It can be implemented at Host level or network level [1].

Protecting networks from computer security attacks is a vital apprehension of computer security. As network traffic may lead to variety of information exchange and sensitive data transfer. Although it is also well known that the dependency of network are also emerging rapidly. Due to this the network condition are very crucial now a days and it will become more complicated in forthcoming time. This traffic may lead to massive damage of network system and its related resources.

Anomaly detection is a way of analyzing the traffic network on the basis of traffic pattern so that the unwanted and malicious attacks can be detected [2].

Network behaviors that cannot be characterized by any model for such condition non-model based approaches are used. Non-model based approaches can be auxiliary classified based on the unambiguous implementation and accuracy constraints that have been imposed on the detector.

Misuse Detection is the way of analyzing the identity of intrusions so that the malicious activity is detected. Misuse detection approaches analyze host or network activity, looking for events that match patterns of known attacks (signatures). First a reference database of attack signatures is constructed, and then monitored events from sensors data are compared against this database for evidence of intrusions. Signature matching is the most commonly employed misuse detection technique. For instance, Snort is a well-known open source signature-based network intrusion detection system [11]. Other misuse detection approaches include rule-based systems, state transition analysis, machine learning and data mining techniques.

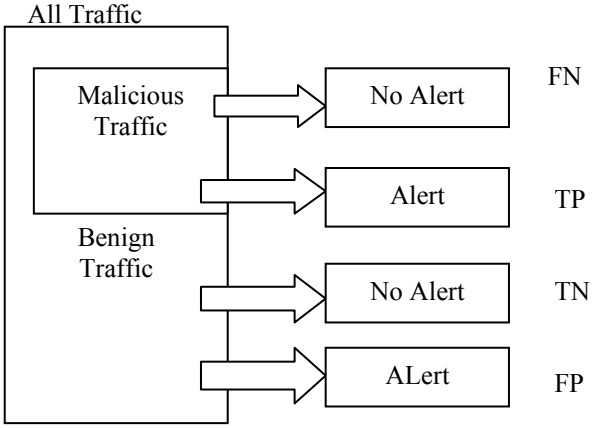

Figure 1. Anomaly Identification Process

KDD Cup99 Dataset

KDD Cup99 dataset contains about 4GB of compressed data in hwihc nearly 7 weeks of network traffic data is collected. It contains 41 feature attributes of the network traffic and contains classes as normal and attacks. The Dataset contains various attacks such as Denial of Service Attack (DOS), User to Root Attack (U2R), Remote to Local Attack (R2L) and Probing Attack.

The table shown below is the summary statistics of the KDD Train Set.

	Original Records	Distinct Records	Reduction Rate
Attacks	3,925,650	262,178	93.2%
Normal	972,781	812,814	16.44%
Total	4,898,431	1,074,992	78.05%

Table 1. Training Dataset

	Original Records	Distinct Records	Reduction Rate
Attacks	250,436	29,378	88.26%
Normal	60,591	47,911	20.92%
Total	311,027	77,289	75.15%

Table 2. Testing Dataset

2. LITERATURE REVIEW

Z. Muda et. al [1] proposed a new technique of detecting intrusions using hybrid combinatorial method of K-mean clustering and OneR Classification. The methodology includes K-mean clustering to be applied first on the KDD Cup99 dataset and then OneR classification is applied to classify the intrusions in the dataset. The methodology is compared on the basis of accuracy of detecting intrusions and detection rate and finding false alarm rate. The hybrid learning of the approach using K-means clustering and classification using OneR technique includes classification of various normal data and attacks [1].

Mohammadreza Ektefa et. al. [2] applied some data mining techniques for the detection of intrusions in KDDCup 99 dataset. The methodology includes combination of classification tree and Support Vector Machine. After implementing the proposed methodology it proofed that the classification decision tree C4.5 is better than SVM learning algorithm. Support vector machine is a learning approach which is used for the clustering and classification of values. It consists of linear kernel based and Gaussian technique. The 'X' parameter contains the data value and 'Y' contains their respective labels. The proposed methodology implemented here provides efficient detection of false alarm rate. Although the technique provides better performance for the intrusion using data mining approach, but can't be better for warehousing [2].

Anup Goyal et. al. [3] implemented intrusion detection using genetic algorithm. This technique also includes a machine learning approach called Genetic Algorithm for the identification of harmful or unwanted attacks in the network. The genetic algorithm detects intrusion on the basis of protocol used in the connection, the services used in the network and status. The genetic algorithm approach generates a set of rules where each of the rules identifies an attack type. The classification of network connection intrusions is also detected using these rules. There are six rules generated to classify various six different types of attacks. It will efficiently identify 100% accuracy for the detection and classification of intrusions [3].

R. Shanmugavadivu has implemented an efficient intrusion detection system uusing fuzzy logic [4]. Here in this paper a fuzzy logic based system is developed for the generation of set of rules and from these set of rules intrusions are detected and classified in a better way. Here frequent items are generated from a set of rules using fuzzy rules based logic system. The proposed technique implemented here proof to be better precision as compared to the other existing technique of detecting intrusions. The methodology starts with the KDDCup 99 dataset which contains 41 training data features. These training features are modified to get 34 training features, then on the basis of the features available classify the features data values. Then these classified data features values are mine to generate set of single length of item sets. Now testing of these rules is done. For this fuzzification is done and rules are generated which are used for the detection and classification of intrusions [4].

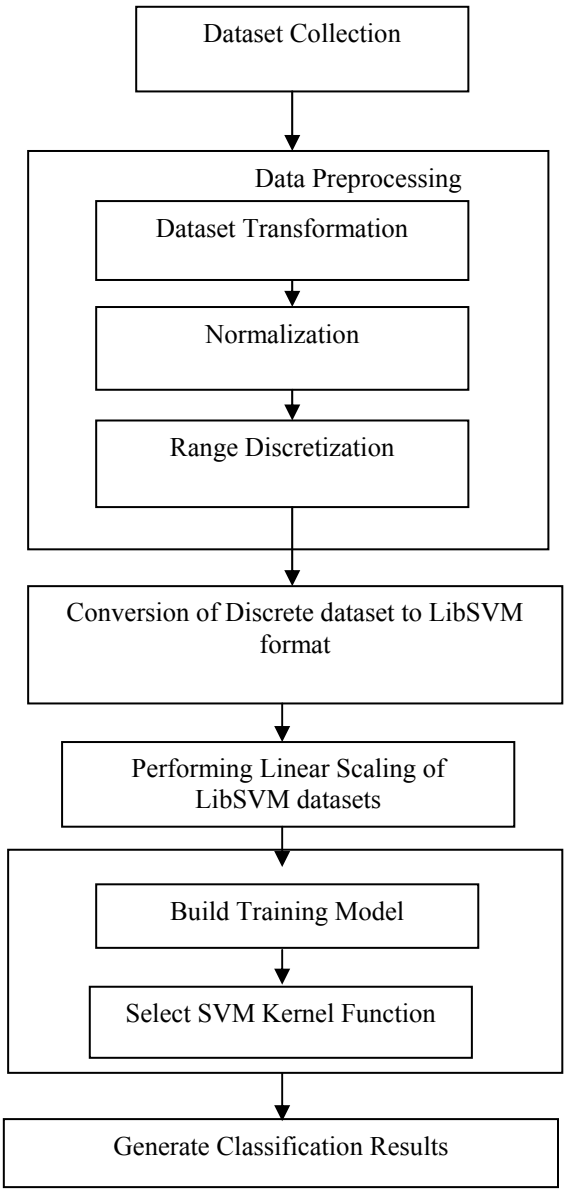

Figure Intrusion detection using SVM [5].

Yogita B. Bhavsar et. al. [5] has proposed a new and efficient way of detecting intrusions using support vector machine. Support Vector Machine is a learning algorithm which is used to classify intrusions on NSL-KDDCup 99 dataset. Although the support vector machine implemented here for the classification of intrusions but it takes a long training time. Here in this paper preprocessing of the NSL-KDDCup 99 dataset is done so that the training time taken by the SVM can be reduced. By using SVM data mining approach attack detection rate is increased and also false positive rate decreases [5].

S. Devaraju et. al. [6] proposed and discusses various intrusion detection techniques using neural network in KDDCup 99 dataset. Here in this paper various neural network based classifiers are implemented for the detection of intrusions in the network. The various Neural network based classifiers Generalized Regression Neural Network, Radial Basis Neural Network, Feed Forward Neural Network and Probabilistic Neural network. The various

techniques of classifications using Neural network is applien on all the features of the KDD Cup99 dataset and the reduced features of KDDCup 99 dataset. The result analysis shows that the KDDCup 99 dataset with reduced features performs better [6].

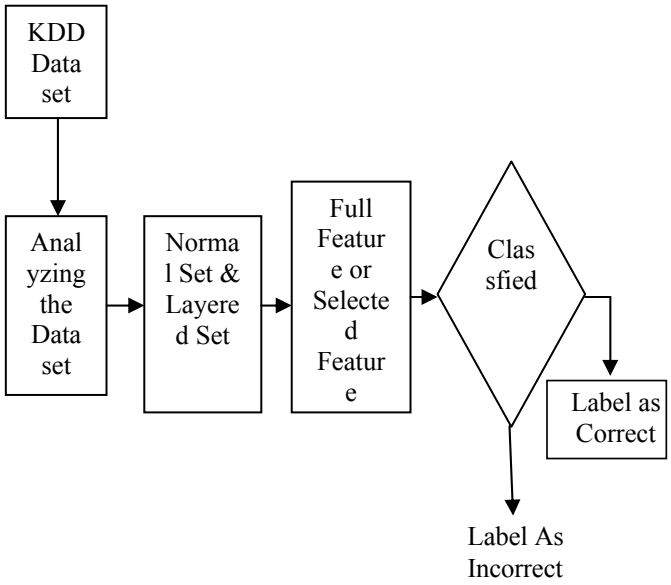

Figure Representation of Layered Approach

Jun Wang, Xu Hong et. al. [7] proposed a real time based detection of intrusions using hybrid combination of PSO-SVM. Here support vector machine is applied on the KDDion of Cup 99 dataset for the classification of intrusions and then particle swarm optimization technique is applied for the optimization of classification. The methodology implemented here provides higher detection rate as compared to the SVM algorithm [7].

H. Güneş Kayacık et. al. [8] has implemented and find the most relevant feature of the KDD Cup99 dataset. Although there are various intrusion detection techniques are implemented for KDD Cup99 dataset, but the technique only provide classification of detection of intrusions, but here in this paper the importance or relevancy of each of the feature is determined and analyze the most discriminating feature of the KDD Cup99 dataset [8].

Shaik Akbar et. al. [9] implemented and discusses study of various various attack in the network . The various Anomaly detection and misuse detection identification using Fuzzy logic and Nave Bayes network and Artificial Intelligence and neural network and genetic algorithms. This paper discusses and finds the various advantages and their issues, hence they divide all the intrusion detection techniques into three categories Rule based detection, Artificial intelligence based detection and various computational detection techniques [9].

Jonathan Palmer implemented Naïve Bayes classifier which is based on the Live Capturer of packets for the detection of intrusions [10]. This technique is used for the detection and identification of intrusion in the network. Hence it identifies the error rate and true positive rate of the algorithm [10].

Intrusion detection using Naïve Bayes

Mrutyunjaya Panda at. Al. [12] proposed a new scheme for intrusion detection using Naïve Bayes Classifier. Here in this paper a Naïve Bayes classifier is applied on the KDDCup 99 dataset. The proposed scheme is applied on the network traffic and then preprocessing is applied on the input traffic pattern which generates pattern building and detector is used for the detection of intrusions and hence generates alarm. Naïve Bayes classifier uses a probabilistic approach for the pattern generation of rules and hence can be used for the detection and classification of intrusions.

Intrusion Detection using K-Means

Gerhard M¨unz at. al. [13] proposed a new approach of detection intrusion using K-means clustering. Here an efficient flow based technique is applied using K-means clustering for intrusion detection. K-means clustering is applied on the network traffic flow and hence classify the normal and abnormal instance traffic from the whole dataset. The cluster centroids are used as patterns for the detection of anomalies in the traffic.

Issue: The technique implemented is only applied on KDD dataset while can't perform for other datasets such as DARPA.

Intrusion Detection using C4.5 Classification

Manasi Gyanchandani et. at. [14] Uses the enhancement measure of detecting intrusions using C4.5. The technique is applied on NSL-KDD dataset for the detection of IDS. The techniques efficiently increase the performance of the methodology by enhancing the classified instances of the dataset.

Issue: Although the technique is efficient and provides better classification but an efficient technique is implemented for NSL-KDD Dataset.

Intrusion Detection using SVM & Decision Tree

Snehal A. Mulay et. al. [15] uses the binary classification for the classification of dataset. Here in this paper first of all support vector machine is applied for the classification of intrusions and then decision tree is created for the effective classification of intrusions. The Decision tree created for the classification of intrusions contains a tree from root node to leaf nodes, where the intrusion to be detected is first compared with the root node of the tree and then move to the leaf node for the detection of intrusions.

Issue: Since the technique is efficient in terms of accuracy and better classification of intrusions but can't be applied on missing attributes.

Intrusion detection using Genetic Algorithm & Fuzzy Logic

Mostaque Md. Et.al. [16] proposes a combinatorial method of applying genetic algorithm and fuzzy logic for the detection of intrusions. The genetic algorithm is applied for generating intrusive rules from the network traffic and then Fuzzy Logic is applied for the optimization of classified values.

Issue: The hybrid technique implemented here provides more false alarm rata and also for the larger dataset with high dimension the technique doesn't provides sufficient

rules with dependent attributes for the detection of attributes.

Intrusion Detection using Fuzzy Logics

Prabhdeep Kaur et. al. [17] uses the fuzzy logics for the network intrusion detection. The Fuzzy Logic are used for the generation of if-then rules for the correctly identification of normal and abnormal attacks. Since intrusion detection can be classified as misuse and anomaly detection approach which can be classified as known and un-known attacks.

Issue: Detection of known and unknown attacks as well as the decrease of efficiency.

Intrusion Detection using Neural Networks

Alan Bivens et. al. [18] uses the neural network for the efficient detection of intrusion in the network traffic. The approach is applied on DARPA and Live TCPDump Data where the data is first compared with the threshold value of the current time and then Preprocessing of the data is done and then clusters and normalization is done and finally an expert system is created using neural network for the correct classification and detection of intrusions.

Issue: The technique provides efficient results for the detection of anomalies in the network traffic but further enhancements can be done for the supervised and unsupervised learning approach.

Intrusion Detection using PSO

Anazida Zainal et. al. [19] proposed an efficient supervised learning based intrusion detection using Particle Swarm Optimization. Here particle swarm optimization is used for the extraction of features for the intrusion detection. The various values stored in the dataset can be categorized and classified using particle swarm optimization techniques where a particular objective function is used for the optimization of particles in the dataset.

Issue: The technique effectively detects the intrusions in the packet but the false alarm rate is increased.

Intrusion Detection using PSO-SVM

Abdulaziz Alsadhan, Naveed Khan proposed and implement an efficient intrusion detection using the combinatorial method of particle swarm optimization and Support Vector Machine for the intrusion detection [20]. The technique is implemented for the wireless sensor network in which particle swarm optimization is applied first for the grouping of similar features of the dataset and then support vector machine is used for the classification of the datasets which contains intrusion as high, low or medium intrusions.

Issue: There are so many technique implemented so that the features can be extracted at grater rate and also detection rate is increased.

Intrusion Detection System using Honeypots

Ram Kumar Singh et. al. [21] proposed a new and efficient technique for the detection of intrusions using honeypots. Since honeypot is implemented at the network based and host based but rules are applied on the honeypot for the detection of intrusion in the packets.

Issue: The technique only identifies the intrusions or anomalies for which rules are generated.

3. CONCLUSION

Intrusions detection system is way of analyzing the traffic so that the unwanted packets that may contain virus or harm to the network can be detected and countermeasure. KDD Cup99 dataset contains a number of traffic packet features and their respective classified attacks such as Neptune, Smurf, Nmap. Here in this paper all those techniques which are used for the detection of intrusion by analyzing KDD Cup99 dataset is discussed.

4. REFERENCES

[1] Z. Muda, W. Yassin, M.N. Sulaiman," Intrusion Detection based on K-Means Clustering and OneR Classification", IEEE 2011.

[2] Mohammadreza Ektefa, Sara Memar," Intrusion Detection Using Data Mining Techniques", IEEE 2010.

[3] Anup Goyal, Chetan Kumar", GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System", 2005.

[4] R. Shanmugavadivu, Dr.N.Nagarajan," Network Intrusion Detection System Using Fuzzy Logic", IJCSE 2011.

[5] Yogita B. Bhavsar, Kalyani C.Waghmare," Intrusion Detection System Using Data Mining Technique: Support Vector Machine", IJETAE 2013.

[6] S. Devaraju and S. Ramakrishnan," Performance Comparison for Intrusion Detection System Using Neural Network With KDD Dataset", ICTACT 2014.

[7] Jun Wang, Xu Hong, Rong-rong Ren, Tai-hang Li," A Real-time Intrusion Detection System Based on PSO-SVM", IWISA 2009.

[8] H. Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood," Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets", 2005.

[9] Shaik Akbar, Dr.K.Nageswara Rao," Intrusion Detection System Methodologies Based on Data Analysis", IJCA 2010.

[10] Jonathan Palmer, Naive Bayes Classification for Intrusion Detection Using Live Packet Capture", Spring 2011.

[11] Roesch Martin, 1999. Snort–lightweight intrusion detection for networks, 13th Systems Administration Conference (LISA), pages 229–238.

[12] Mrutyunjaya Panda1 and Manas Ranjan Patra," Network Intrusion Detection Using Naïve Bayes", IJCSNS 2007.

[13] Gerhard M¨unz, Sa Li, Georg Carle," Traffic Anomaly Detection Using K-Means Clustering", 2008.

[14] Manasi Gyanchandani, R. N. Yadav, J. L. Rana," Intrusion Detection using C4.5: Performance Enhancement by Classifier Combination", ACEEE2010.

[15] Snehal A. Mulay, P.R. Devale," Intrusion Detection System using Support Vector Machine and Decision Tree", IJCA 2010.

[16] Mostaque Md. Morshedur Hassan," Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic", IJIRC2013.

[17] Prabhdeep Kaur, Sheveta Vashisht,"Mingle Intrusion Detection System Using Fuzzy Logic", IJEAT 2013.

[18] Alan Bivens, Chandrika Palagiri", Network-based Intrusion Detection using Neural Networks", ANNIE 2002.

[19] Anazida Zainal, Mohd Aizaini Maarof, and Siti Mariyam Shamsuddin," Features Selection Using Rough-PSO in Anomaly Intrusion Detection", 2005.

[20] Abdulaziz Alsadhan, Naveed Khan," A Proposed Optimized and Efficient Intrusion Detection System for Wireless Sensor Network", IJERECE 2013.

[21] Ram Kumar Singh, Prof. T. Ramanujam," Intrusion Detection System Using Advanced Honeypots", IJCSIS 2009.