ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

Hybrid Approach for IDS using FGA and Machine Learning

Ashwini B. Bhopale, Prof. C. M. Jadhav

PG Student, Dept. of Comp. Engg, BIGCOE, Solapur, India

HOD, Dept. of Comp. Engg, BIGCOE, Solapur, India

ABSTRACT: System security is of essential part now days for huge organizations. The Intrusion Detection frameworks (IDS) are getting to be irreplaceable for successful assurance against assaults that are continually changing in size and intricacy. With information honesty, privacy and accessibility, they must be solid, simple to oversee and with low upkeep cost. Different adjustments are being connected to IDS consistently to recognize new assaults and handle them. This paper proposes a semi-supervised model based on combination of ensemble classification for network traffic anomaly detection. Ids is try to perform in real time, but they can not improved due to the network connections. In this research, we are trying to implement intrusion detection system (IDS) using anomaly intrusion detection method for misuse as well anomaly detection. We have used various individual classification methods and its ensemble model on KDD99 and NSL-KDD data set to check the performance of model. Due to irrelevant features in data set, also applied Information Gain (IG) feature selection technique on best model. Finally our proposed model produces highest detection rate and lower false positive ratio compare to others.

KEYWORDS: anomaly detection, Information Gain, intrusion detection, network security, Network traffic anomaly, semi supervised model.

I. INTRODUCTION

Computer networks security is important domain of research for years. For protecting important information or data the network security technology has become very useful. Any fruitful endeavor or unsuccessful endeavor to trade off the honesty, privacy, and accessibility of any data asset or the data
itself is viewed as a security assault or an interruption. Every day industries has to deal with the variety of attacks. Avoiding this problems with the help of Intrusion Detection System (IDS). The wide use of computer networks and the increase in web based business has made security of the host and network an important issue as these are vulnerable to attacks. These attacks can be passive that just reads confidential data or it can be active attack that also modifies or fabricates the data [8].
Since it is not possible to avoid these vulnerabilities and design a completely secure system. Intrusion detection has become a major challenge. The key objective of Intrusion detection system is to recognize the attack and in some matter examine it. Several techniques and methods have been developed. But with the progression of new attacks more robust systems need to be designed.
Basically Intrusion Detection System (IDS) ordered into two distinctive arranged Host Base Intrusion Detection System (HIDS) and Network base Intrusion Detection System (NIDS). Today's system security foundation promisingly relies on Network intrusion detection Framework (NIDS). NIDS gives
security from known interruption assaults. It is unrealistic to stop interruption assaults, so associations should be prepared to handle them. ID is a cautious component whose main role is to keep work continuing considering every conceivable assault on a framework. Interruption recognition is a procedure used to distinguish suspicious movement both at system and host level. Two principle ID methods accessible are abnormality identification and abuse location.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

The oddity identification model depicts the typical conduct of a client to recognize this tering they observed the importance of ultra-metric distance.
They also used ultra-metric distance in proposed method from the aggregated distance matrices. Finally, they generated hierarchical clustered structure with separation of the cluster. This model ensemble both framework as partitioned clustering and hierarchical clustering.

Sandro Vega-Pons, Jos Ruiez-Shulcloper[3], proposed A Survey of Clustering Ensemble Alogorithms which provides alternative ensemble method when occurring cluster analysis problem. In that generating a number of clusters from the same dataset and finally ensemble all together. The main aim of this model is to improve the quality of individual clusters. This paper presents an overview of clustering ensemble methods that can be very important for the clustering practitioners for the purpose of community.
In Kaur, P.[4], present Adaptive Intrusion Detection
Based on K-SVMeans Algorithm, which overcome the problem of one individual algorithms for improving the efficiency. To improve the efficiency of clustering algorithm it is a challenging work without overcoming the generalization performance
of Support Vector Machine(SVM). Have gone through this difficulties , so this model developed new hybrid method based on combination of SVM and k-means clustering. SVM is used to build classifiers which can help users to take business decision very well. Response time of support vector machine is concern in real-time task network traffic analysis. The KSVM algorithm combine the k-means clustering technique with SVM.
Basant Subba, Santosh Biswas, Sushanta Karmakar[5], proposed A Neural Network Based System for Intrusion Detection and Attack Classification. This paper used a simple Artificial Neural Network (ANN) based IDS model. Using different optimization techniques presents intrusion detection system with the help of feed forward and the back propagation algorithms. This method used to reduce the overall computational overhead and also maintain performance level. The proposed model results shows that high accuracy and higher detection rate on NSL-KDD dataset. This model is very useful for analyzing the detection of intrusions in real time environment.
In Wagh SK, Kolhe SR.[6], present Effective semisupervised approach towards intrusion detection system using machine learning techniques. This system proposed new semisupervised method using machine learning approach, which effectively increased detection rate and by default accuracy also improved. Semi-supervised learning method contains both unlabeled data and labeled data. This paper present a new method called as novel self-learning construction. In that input as a train data given to the supervised classifier with unlabeled dataset for testing purpose. Entropy is calculated for correct predicated label. Threshold calculation is done by adding confident data into the training data. For data selection number of statistical methods used. Alert mechanism is improved with this proposed model
According to Naila Belhadj Aissa, Mohamed Guerroumi [7], present A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems, which focus on identifying the anomaly by using genetic clustering algorithm. The proposed model is Genetic Clustering for Anomaly-based Detection (GC-AD). GC-AD utilizes a divergence measure to frame k cluster and find centroid of the k groups by applying on hereditary procedure. It worked on KDD99 dataset for gaining the exactness of frameworks system. The output is calculated with kmeans grouping The main goal for CG-AD (Clustering Genetic for Anomaly-based Detection) is to get an ideal homogenous apportioning of typical and oddity cases.

II. SYSTEM OVERVIEW

Existing system has three phase. The first phase is choosing a proper dataset and applied on preprocessing phase for minimizing or elimination of the noise forced on the data. Second is building the hybrid model which consisting number of classifiers which produced accuracy of each individual classifier. Finally in Analysis phase, result is generated by comparing accuracy of each individual model and selects the best model as ensemble.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

A. Problem Statement

To develop the Intrusion Detection System for distributed architecture for the detection and correctly classification of the incoming network packet attacks using ensemble method.

B. Proposed System

The goal of proposed anomaly network intrusion detection system is to maximize the detection accuracy ,to minimize false positive rate and detector generation time. Basically there are two phase in the proposed system, we have taken NSLKDD dataset for system training as well testing purpose.

Training Phase:

1) Upload training data for feature extraction.
2) Apply Genetic algorithm for for rule creation
3) Create rules set as normal pool as well as intrusion pools set.

Testing Phase:

1) Upload Testing data or any packet which is collected from network environment.
2) Extract all feature using attribute selection.
3) Apply Normalization approach on dataset.
4) Apply ensemble approach on all train as test features.
5) Show results with classification accuracy.
6) classify all attacks.
7) Show detection results.

current client's irregular or unaccustomed activity [10]. Identification is the procedure of observing the activities happening in a network framework or organizes and breaking down them for indications of likely occurrence, which are infringement or looming dangers of infringement of network security arrangements, adequate use strategies, or normal security hones. Fundamentally when an interloper endeavor to break into a data framework or perform an activity not authoritatively permitted, we imply to this activity as an interruption. Interruption system may incorporate abusing programming bugs and plan mis-configurations, secret word incensed, sniffing unsecured exchange, or misusing the outline defect of express conventions. An Interruption Location
Framework [12] is a plan for distinguishing interruptions and reporting them definitely to the best possible power.

III. REVIEW OF LITERATURE

In Kagan Tumer a, Adrian K. Agogino[1], proposed Ensemble Clustering with Voting Active Clusters (VACs). This paper consist various clustering models into one cluster that is ensemble clustering. In ensemble clustering does not require all collected data in one central location. The contribution of this proposed model is providing an adaptive voting method to maximize the quality measure by clustering update their votes. This method achieved better performance than traditional cluster ensemble method. But this model works only on noisefree
condition.
Li Zheng Tao Li, Chris Ding[2], presents Hierarchical Ensemble Clustering model which works on both partitional clustering and hierarchical clustering. For hierarchical clus_ Objectives of Proposed System:

1) To study existing Network Intrusion Detection Systems (NIDSs) and types of NIDSs.
2) Execute the same system on HIDS as well NIDS
base environment.
3) To propose a new integrated approach for network
anomaly detection using ensemble approach, compare the result with individual algorithmic results.
4) To compare the experimental results of existing methodology with proposed system for network anomaly detection.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

IV. SYSTEM ARCHITECTURE

The proposed system worked with ensemble model. When two or more models are ensemble together to form a new model called as ensemble model. This ensemble model combines the output of sever al classifiers and produced a single composite classification. Our proposed model consisting number of classifiers. System first collect data from different online as well as offline sources. Once data has collected by system it will apply some data mining strategies with different classification approaches.

SYSTEM ARCHITECTURE

The proposed system worked with ensemble model. When two or more models are ensemble together to form a new model called as ensemble model. This ensemble model combines the output of several classifiers and produced a single composite classification. Our proposed model consisting number of classifiers. System first collect data from different online as well as offline sources. Once data has collected by system it will apply some data mining strategies with different classification pproaches.

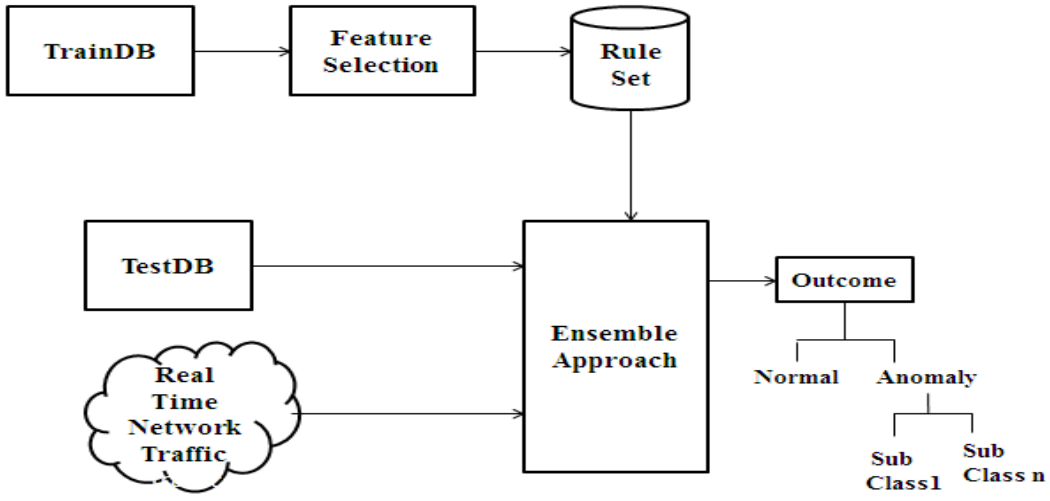

Fig. 1. System Architecture

A. System Modules

1) Data Preprocessing:

Data preprocessing done by Weka tool. This is the offline method. Data preprocessing includes following three main task:

a) Converting non-numerial features of NSL-KDD dataset into numerical values.

b) At the end transfering attack types into numerial values.

c) Finally preparing proper dataset.

2) Feature Selection:

Normalization done in this phase. Min-Max normalization used to normalizing the features. Information Gain (IG) is use to reduce the features. Information Gain(IG) is apply on Feature selection phase. Information Gain is nothing but attribute selection mechanism in both training and testing dataset.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

3) Ensemble Model:
Building the hybrid model consisting number of classifiers. Mapping TestDB with Rule set and apply on ensemble method. Developing a model that exhibits the best performance and accuracy. Compare the accuracy of each classifiers and select best model.
4) Result Generation:
Finally results are generated whether incoming packet is normal or anomaly. If it is anomaly then it also finds subclasses of that anomaly.

Algorithms

1) Algorithm 1:GA(Genetic Algorithm) For Rules Creation
Input: Set of network packet which consist 41 attributes with class label
Output: Set of normal as well intrusion rules Step 1: Initialize randomly population with 41 chromosomes.
Step 2: Initialize N (In the training set total number of records).
Step 3: The new population for each chromosome.
Step 4: Apply Crossover to best selected chromosome.
Step 5: Apply Mutation for each chromosome to generate new population.
Step 6: Calculate fitness= F(x) /sum (F(x)).
Step 7: Select best fit chromosome as 50 % and delete worse fit chromosome.
Step 8: End for

2) Algorithm 2: NB(Naive Bayes)
Input: Feature of BK rules TrainF[], features if test
record TestF[]
Output: highest Similarity weight for class label
Step 1: Read all training rules from DB for each (Rec
R into Train[])!=Null
Step 2: items [] split(R)
Step 3: items1 [] split(TestF)
Step 4: CalculateWeight(DB [i], items1)
Step 5: Return w;
3) Algorithm 3: ANN(Artificial Neural Network)
Step 1: for all (T in HidenLayer [] !=null) do
Step 2: items[] split(T)
Step 3: items1[] split(InputNueron)
Step4:w=CalculateWeight(HidenLayer[i], InputNueron)
Step 5: Return w;

4) Algorithm 4: J48
Input : Feature of BK rules TrainF[], features if test
record TestF[]
Output : highest Similarity weight for class label
Step 1: for all (T in TrainF [] !=null) do
Step 2: items [] split(T)
Step 3: items1 [] split(TestF)
Step 4: w = classifyToAll(Train,TestF[], Label)
Step 5: Return w;

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

C. Dataset Details

The KDD cup 99 dataset [9] contains several statistical analyses which affects to detect the accuracy of many IDS model. The KDD datset has training and testing dataset. The total number of training dataset is approximately 4, 900, 000 .It contains 41 features with labeled as normal or specific attack type. 300, 000 instances contains in testing dataset with twenty four training attack types and extra fourteen attack types in the test set only. NSL-KDD data set is a modified version of its precursor. It consist, important records of the KDD data set. There are 4 attack classes of anomaly which are further classified as DoS, Probe, R2L and U2R. In that there are subtypes of each anomaly attack types.

D. Experimental Results

The existing survey basically focus on soft computing and classification based detection approach, basically both methods having the good detection rate but at times it generates more false positive ratio. Some systems are also not applicable in real time environment and some cant be focus on mis-classified anomalies. As observed, most applications still miss the mark as there is no system that at present gives a 100% discovery rate and the sky is the limit. In proposed system we have two tests. In the first investigation, we utilized our fluffy hereditary calculation to group typical system information and assault. At that point, we indicate identification rate acquired for NSL-KDD dataset. I characterize them into two classes which are ordinary and assault. In the second analysis, we utilized the fluffy hereditary calculation to arrange sorts of assaults in the online continuous sniffer dataset.

V. CONCLUSION

Intrusion detection system mainly used to protect systems and to achieve the better outcome. In last past ten years there are variety of intrusion detection system to the purpose of security. Each method has its own advantages and disadvantages. There is no perfect method to obtain the goal. Researchers are trying to invent new method to reduce the restrictions of the past methods. Lots of complexities are present in this modern era, so the perfect solution and the perfect detection is not an attainable goal. In this research work we proposed ensemble method for network traffic anomaly detection. Our approach concentrated on building normal traffic profile of the anomaly detection model. For normal profile experimental result showed that the features of NSL-KDD is efficient. The experimental result showed that system has excellent performance with small training dataset and detection accuracy. We also proposed a new model integrates anomaly detection system with signature-based detection system along with some enhancements of building quality normal profile. Proposed research work also perform the better detection, On the basis ensemble approach implementation we got a ids system can achieve better detection rate for all attacks as well as unknown attacks. In future work we can minimized the computation time consuming by the different algorithm.

REFERENCES

[1] Kagan Tumer a, Adrian K. Agogino(2008), Ensemble clustering with voting active clusters, Elsevier

[2] Li Zheng Tao Li, Chris Ding(2010),Hierarchical Ensemble Clustering, IEEE International Conference on Data Mining

[3] SANDRO VEGA-PONS, JOS RUIZ-SHULCLOPERy(2011), A Survey of clustering Ensemble Algorithms, International Journal of Pattern Recognition and Artificial Intelligence Vol. 25

[4] Kaur, P.(2013),Adaptive Intrusion Detection Based on K-SVMeans Algorithm (Doctoral dissertation, THAPAR UNIVERSITY PATIALA).

[5] Wagh SK, Kolhe SR., Effective semi-supervised approach towards intrusion detection system using machine learning techniques, International Journal of Electronic Security and Digital Forensics, 7(3):290-304, 2015.

[6] Naila Belhadj Aissa, Mohamed Guerroumi , A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems, IEEE, 2015.

[7] Basant Subba , Santosh Biswas, Sushanta Karmakar , A Neural Network Based System for Intrusion Detection and Attack Classification, IEEE, 2016.

[8] Wagh SK, Pachghare VK, Kolhe SR, Survey on intrusion detection system using machine learning techniques, International Journal of Computer Applications. 1;78(16). Jan 2013.

[9] Mohammed A. Ambusaidi et. al., Building an intrusion detection system using a filter-based feature selection algorithm , IEEE TRANSACTIONS ON COMPUTERS, VOL., NO , NOVEMBER 2014.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 11, November 2018

[10] Fatemeh Barani , A Hybrid Approach for Dynamic Intrusion Detection in Ad Hoc Networks Using Genetic Algorithm and Artificial Immune System, IEEE , 2014.
[11] Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis, Intrusion Detection System Using Genetic Algorithm, Science and Information Conference , 2014.
[12] Alka Chaudhary, Vivekananda Tiwari, Anil Kumar , A Novel Intrusion Detection System for Ad Hoc Flooding Attack(Using Fuzzy Logic in Mobile AdHoc Networks), IEEE, 2014.
[13] Shadi Aljawarneha, Monther Aldwairi, Muneer Bani Yassein, Anomalybased intrusion detection system through feature selection analysis and building hybrid efficient model, Elsevier ,2017.
[14] Eduardo K. Viegas, Altair O. Santin , Luiz S. Oliveira , Toward a reliable anomaly-based intrusion detection in real-world environments, Elsevier ,2017.
[15] Amira SayedA. Aziz, Sanaa EL-OlaHanafi, Aboul EllaHassanien, Comparison of classification techniques applied for network intrusion detection and classification, Elsevier , 2016.
[16] Weiwei Chen, Fangang Kong, A Novel Unsupervised Anomaly Detection Approach for Intrusion Detection System, IEEE, 2017.
[17] Kumari VV, Varma PR., A semi-supervised intrusion detection system using active learning SVM and fuzzy c-means clustering, InI-SMAC (IoTin Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, International Conference on 2017 Feb 10 (pp. 481-485), 2017.
[18] Zhang X, Zhu P, Tian J, Zhang J., An effective semi-supervised model for intrusion detection using feature selection based LapSVM., InComputer, Information and Telecommunication Systems (CITS), IEEE, International Conference on 2017 Jul 21 (pp. 283-286), 2017.
[19] N. H. Duong, H. D. Hai, A semi-supervised model for network traffic anomaly detection[C]. IEEE 2015 17th International Conference on Advanced Communication Technology (ICACT), Phoenix Park, 2015:
70-75, 2015.
[20] Gao, G., Miao, G., Sun, J. and Han, Y., Improved semi-supervised fuzzy clustering algorithm and application in effective intrusion detection system, International Journal of Advancements in Computing Technology (IJACT), Vol. 15, No. 4, pp.689696, 2013.
[21] Meng, Y. and Kwok, L-F. , Intrusion detection using disagreementbased semi-supervised learning: detection enhancement and false alarm reduction, 4th International Symposium, Lecture Notes in Computer Science, Vol. 7672, pp.483497, Springer, Berlin Heidelberg, 2012.