International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 7, July-2018

Improved Approach A2DE based Intrusion detection System

Arvind Kumar, Rahul Gupta (Assistant Professor)

Department of Computer Science & Engineering
S.R. Institute of management and technology
Lucknow, India

***Abstract*—With the advancement in information and communication technology (ICT), it has become a vital component of human's life. But this technology has brought a lot of threats in cyber world. These threats increase the chances of network vulnerabilities to attack the system in the network. Intrusion detection is a procedure of checking the unapproved or undesirable activities happening in a PC framework which abuse network security rehearses. It is an essential technique in the network security domain which tries to identify whether the security of computer system has been compromised or not. In IDS, there are various methods used in data mining and existing technique is not strong enough to detect the attack proficiently. Weighted Average 2-Dependence Estimator (WA2DE) is an enhanced version of A2DE and in this technique; we have to assign weights to each attribute. The dependent attributes having lesser weights by defining the degree of the dependencies. This paper manages a novel ensemble classifier (WA2DE+ Random Tree) for IDS. Proposed ensemble classifier is built using two well-known algorithms WA2DE and Random Tree. This tree improves accuracy and reduces the error rate. The performance of proposed ensemble classifier (WA2DE+Random Tree) is analyzed on Kyoto data set. Proposed ensemble classifier outperforms WA2DE and Random Tree algorithms and efficiently classifies the network traffic as normal or malicious.**

Keywords:- IDS, WEKA, DM, WMN, etc.

I. INTRODUCTION

Intrusion detection is the procedure of shrewdly observing the occasions happening in computer system or organize and investigating them for indications of infringement of the security strategy, Parker has characterized six security issues to be considered while planning IDS: Confidentiality, Integrity, Availability, Utility, Authenticity, and Possession of a computer or network. Intrusions are caused by aggressors getting to the frameworks from the Internet, approved clients of the frameworks who endeavor to increase extra benefits for which they are not approved, and approved clients who abuse the benefits given them. IDSs are s/w or h/w items that mechanize this observing and examination process. [1] Intrusion detection is a procedure of observing the unapproved or undesirable activities happening in a PC framework which abuse organize security rehearses. It is an essential technique in the network security domain which tries to identify whether the security of computer system has been compromised or not. IDSs are named system, host, or application construct frameworks depending with respect to their method of arrangement and information utilized for examination. A number of variants of IDS have been proposed in the literature but majority of them seem to be unsuitable for detecting real-time malicious activities. This can be attributed to the various challenges like data handling, computational cost, storage complexity, etc. [2]

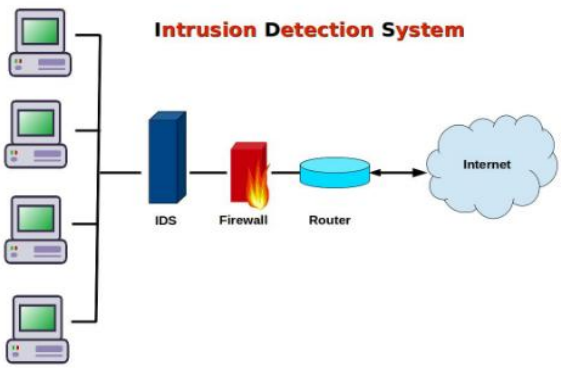

Fig.1 Intrusion Detection System

1.1 Signature Based IDS

In SBIDS, otherwise called abuse detection, signatures of known assaults are put away and the occasions are coordinated against the put away marks. It will signal an intrusion if a match is found. The primary downside with this strategy is that it can't recognize new assaults whose marks are unknown. This implies that an IDS utilizing abuse identification will just distinguish known assaults or assaults that are sufficiently comparative to a known assault to coordinate its signature.[1] Signature-based detection is the way toward looking at marks/examples of known assault with the watched occasions to recognize conceivable occurrences. The most common form of signature based IDS used commercially specifies each pattern of events that corresponds to an attack as a separate signature.[3]

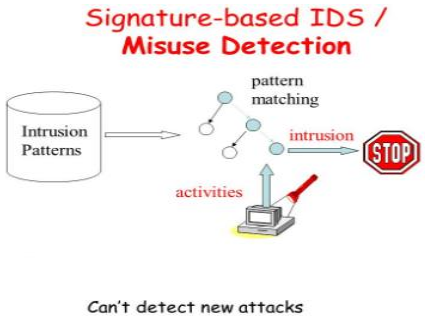

Fig.2 Signature Based IDS

1.2 Anomaly Based IDS

ABIDS has pulled in numerous scholastic analysts because of its potential for tending to novel assaults. Oddity detection is the distinguishing proof of new or obscure information that a machine learning framework doesn't know about amid preparing. ABIDS have two noteworthy favorable circumstances over signature based IDS. The main preferred standpoint is the capacity to recognize obscure assaults and additionally "zero day" assaults. This is a direct result of the capacity of anomaly detection systems to display the ordinary task of a framework/arrange and recognize deviations from them. A second advantage is that the previously mentioned profiles of ordinary action are tweaked for each framework, application and additionally arrange, and thusly making it extremely troublesome for an aggressor to know with sureness what exercises it can complete without getting distinguished.[1]

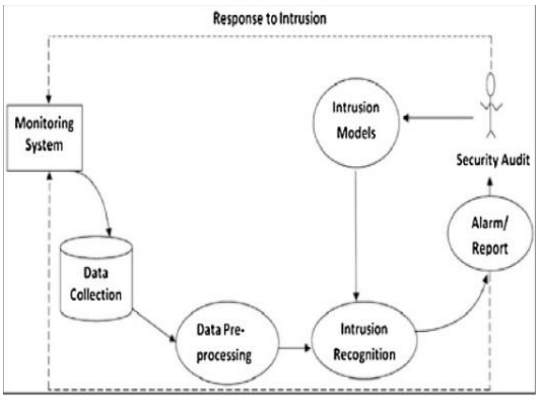

Fig.3 Anomaly Based IDS

II. Literature Survey

Md Reazul Kabir et al [2017] Our research study presents a wrapper approach for intrusion detection. In this system Feature determination method dispense with the superfluous highlights to decrease the time complexity and manufacture a superior model to foresee the outcome with a more noteworthy precision and Bayesian system functions as a base classifier to anticipate the kinds of attack.[4]

Hajisalem et al [2018] In this paper, we propose another half and half characterization technique in light of Artificial Bee Colony (ABC) and Artificial Fish Swarm (AFS) calculations. The Fuzzy C-Means Clus- tering (FCM) and Correlation-based Feature Selection (CFS) techniques are applied to divide the training dataset and remove the irrelevant features, respectively. In accumulation, If-Then standards are created through the CART procedure as per the chose includes keeping in mind the end goal to recognize the typical and anomaly records. In like manner, the anticipated hybrid strategy is prepared by means of the produced rules.[5]

Partopour et al [2018] we present the utilization of an outfit learning strategy known as Random Forests to microkinetics demonstrating and the computationally proficient reconciliation of microkinetics into response designing models. Initially, we indicate how Random Forests can be utilized for mapping precomputed microkinetics information. Arbitrary Forests can be utilized to anticipate new datasets while keeping the forecast precision high and the computational load low. The strategy is likewise used to distinguish the critical factors in the instrument concerning general response rate and selectivity.[6]

Divya Gupta et al [2017]This paper presents a novel IDS based on probabilistic data structures named as ProIDS. In the proposed ProIDS, a popular probabilistic data structure (PDS), Bloom filter has been used to store the information about the suspicious nodes. Using Bloom filter, the number of hits on suspicious nodes per unit time has been computed using the modified version of Count min sketch, i.e., MCMS, a PDS. The work also presents a detailed theoretical analysis backed by relevant technical description.[7]

R. Vijayanand et al [2018] In this work, a hybrid genetic algorithm (GA) and mutual information (MI) based feature selection system is proposed for IDS. The execution of IDS with the proposed highlight determination system is examined with IDS having mutual information, genetic algorithm and GA+MI based feature selection procedures using ADFA-LD dataset.[8]

III. PROPOSED WORK

Naïve Bayes is the arithmetical algorithm for the learning method for the detection of intrusion. This algorithm used attributes and they all are independent but this independency can influence the accuracy. A2DE (Averaged 2-Dependence Estimator) algorithm used to overcome the problem and it is further accurate than Naïve Bayes.

Proposed Algorithm:

Step:1 Input csv file and transform it into text file
Step:2 Convert data into binary form
Step:3 Select Normal Labels as 1 and Attack Labels as 1 and 2
Step:4 Performed preprocessing over the file
Step:5 Now eliminate additional attributes from data
Step:6 Separate the data into test and training file for further processing
Step:7 Select Weighted A2DE and Random Tree in vote
Step:8 Select test file for further calculation
Step:9 Average of chances evaluate
Step:10 Get the output in the form of normal or malicious in network
Step:11 Stop

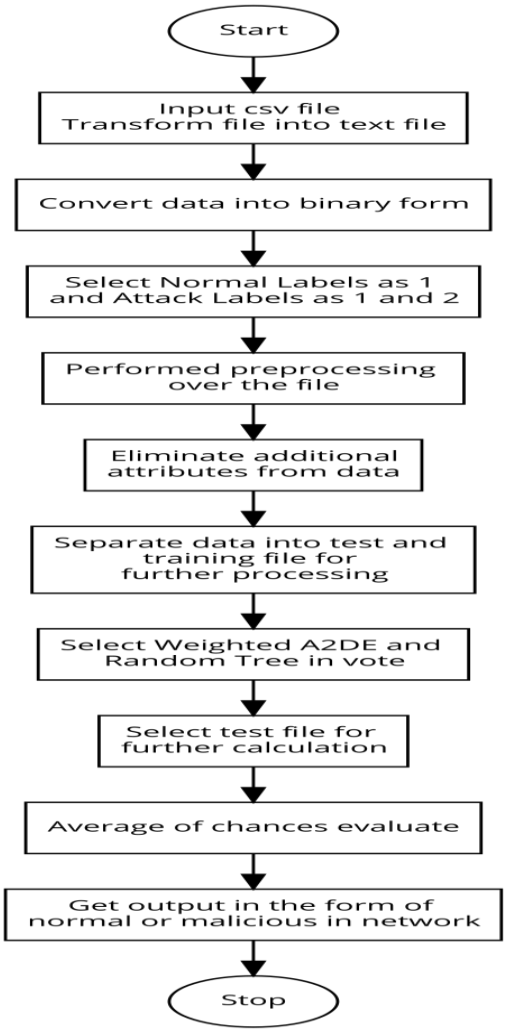

Fig.4 Flowchart of Proposed Work

IV. RESULT ANAYSIS

In the result analysis, the experiment of proposed work (WA2DE-RT) performed by using ensemble classifier. Kyoto dataset 2006 used for the investigational study of the traffic data. This dataset contains 24 features and we used only 15 features and excluded the features which are related to security analysis.

Instances: 85346

Attributes: 15

Duration_binarized
Service
SourceByte_binarized
DestinationByte_binarized
Count_binarized
Same srv rate_binarized
Serror rate_binarized
Srv serror rate_binarized
Dst host count_binarized
Dst host srv count_binarized
Dst host same src port rate_binarized
Dst host serror rate_binarized

Dst host srv serror rate_binarized

Flag

Label

Test mode: 10-fold cross-validation

1. Training

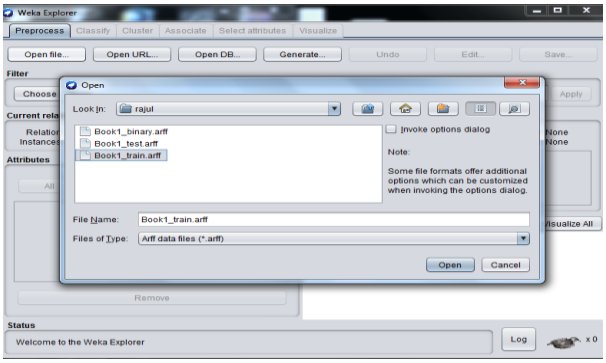

Fig.5 Select training file

Base Work:

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.996	0.865	0.955	0.996	0.975	0.285	0.935	0.996	Attack
0.135	0.004	0.667	0.135	0.224	0.285	0.935	0.439	Normal
Weighted Avg.	0.952	0.821	0.940	0.952	0.937	0.285	0.935	0.968

=== Confusion Matrix ===

a	b	<-- classified as
80681	294	a = Attack
3783	588	b = Normal

Propose Work:

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.998	0.734	0.962	0.998	0.979	0.463	0.941	0.997	Attack
0.266	0.002	0.853	0.266	0.406	0.463	0.941	0.514	Normal
Weighted Avg.	0.960	0.696	0.956	0.960	0.950	0.463	0.941	0.972

=== Confusion Matrix ===

a	b	<-- classified as
80774	201	a = Attack
3208	1163	b = Normal

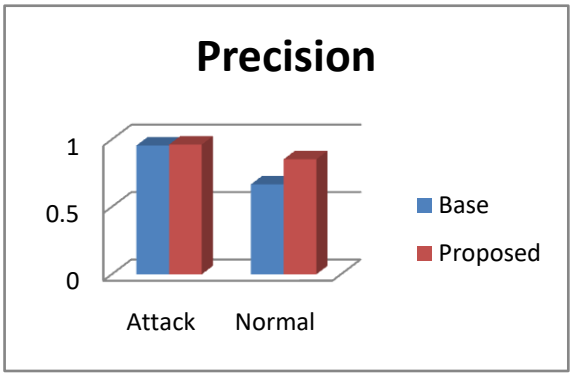

Fig.6 Precision Comparison

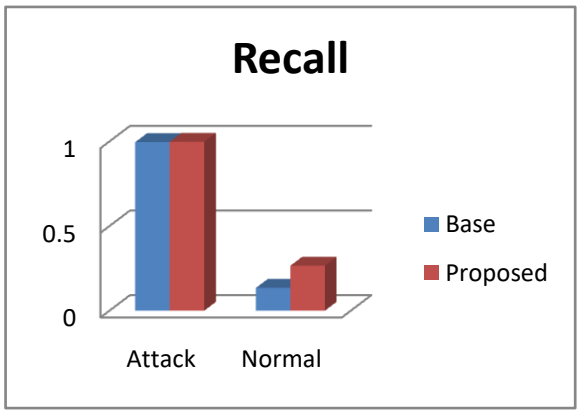

Fig.7 Recall Comparison

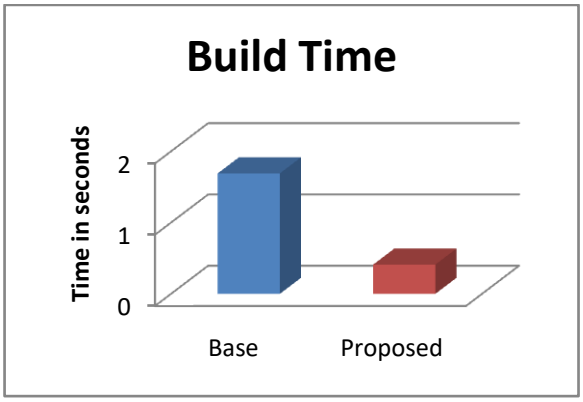

Fig.8 Time taken to build model

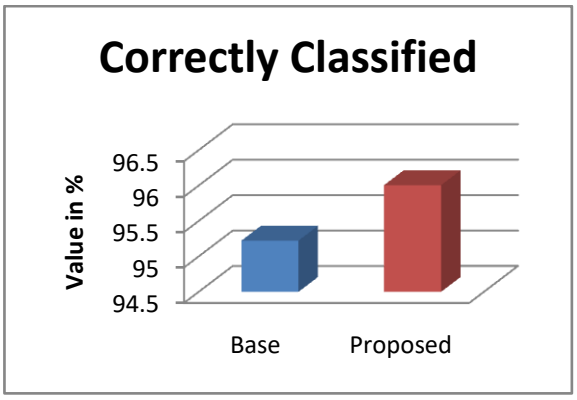

Fig.9 Correctly Classified Instances

V. CONCLUSION

Intrusion detection is the process of identifying the occurrence of different attacks in a network by analyzing the features of network traffic for providing security to a network. WMN has advanced routing protocols for providing guaranteed communication with end devices. The attacks on routing protocols could damage the entire communication network. Subsequently, the routing based pantomime assaults like black hole, grey hole and so forth are much of the time utilized by assailants against WMN. Traffic based IDS proficiently recognizes these assaults by dissecting the activity data and routing information. In this thesis, we proposed a novel ensemble classifier (WA2DE Random Tree) for intrusion detection system. The proposed method effectively classifies n/w traffic as typical or malignant. The outcomes demonstrate that proposed classifier is exact than RF and AODE classifier. We considered Kyoto data set for experimental analysis. As Base classifiers are not capable of detecting the attacks accurately, proposed Ensemble classifier outperforms base classifiers WA2DE and Random Tree. The results presented in this paper show that integration of WA2DE, Random Tree and pre-processing technique will yield the good result for IDS.

References

[1] Manasi Gyanchandani, J.L.Rana, R.N.Yadav, "Taxonomy of Anomaly Based Intrusion Detection System: A Review", International Journal of Scientific and Research Publications, Volume 2, Issue 12, December 2012.

[2] Divya Gupta, Sahil Garg, Amritpal Singh, Shalini Batra, Neeraj Kumar, M.S. Obaidat, "ProIDS: Probabilistic Data Structures based Intrusion Detection System for Network Traffic Monitoring", 978-1-5090-5019-2/17/$31.00 ©2017IEEE.

[3] Jaina Patel, Mr. Krunal Panchal, "Effective Intrusion Detection System using Data Mining Technique", Journal of Emerging Technologies and Innovative Research (JETIR), June 2015, Volume 2, Issue 6.

[4] Md Reazul Kabir, Abdur Rahman Onik,Tanvir Samad, "A Network Intrusion Detection Framework based on Bayesian Network using Wrapper Approach", International Journal of Computer Applications (0975 – 8887) Volume 166 – No.4, May 2017.

[5] Vajiheh Hajisalem, Shahram Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection", Computer Networks, 2018.

[6] Behnam Partopour, Randy C. Paffenroth, Anthony G. Dixon, "Random Forests for mapping and analysis of microkinetics models", Computers and Chemical Engineering (2018).

[7] Divya Gupta, Sahil Garg, Amritpal Singh, Shalini Batra, Neeraj Kumar, M.S. Obaidat, "ProIDS: Probabilistic Data Structures based Intrusion Detection System for Network Traffic Monitoring", 978-1-5090-5019-2/17/$31.00 ©2017IEEE

[8] R. Vijayanand, D. Devaraj and B. Kannapiran, "A novel intrusion detection system for wireless mesh network with hybrid feature selection technique based on GA and MI", Journal of Intelligent & Fuzzy Systems,2018.