International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 6 | ISSN : 2456-3307

Anomaly Based Intrusion Detection System Using Soft Computing and Classification Approach

Anuradha S. Varal, Dr. S. K. Wagh
Department of Computer Engineering, Modern Education Society College of Engineering, Pune, India

ABSTRACT

Intrusion discovery is prominent up and coming zone, as an ever growing number of complex information is being put away and handled in arranged frameworks. With wide use of internet service, there is constant risk of intrusions and misuse. Thus Intrusion Detection system is most important constituent of computer and its network security. Intrusion Detection System is software centered monitoring mechanism for a computer network that searches presence of wicked activity in the network. IDS system ought to congregated contemplation by sustaining high safety levels safeguarding reliable and secure transmission of the information between different organizations. Intrusion discovery systems categorize computer activities into two main categories: normal and distrustful activities. Many perspectives for intrusion detection have been proposed afore but none displays satisfactory results so we examine for better result in this field. The proposed study likewise takes a diagram of several kinds of arrangement strategies for Intrusion Detection System (IDS). We additionally research in these extraordinary methodologies, their exactness and also false positive proportions.

Keywords : Intrusion Detection system; soft computing; classification techniques.

I. INTRODUCTION

The security of computer networks has been in the concentration 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 [1]. 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) [13,14,15]. 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. The oddity identification model depicts the typical conduct of a client to recognize this 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.

This paper is sectioned as follows. Some related work on IDS in Sect. II. Section III presents a detailed report of the proposed structure stepwise. Sect. IV. Includes the details of database and experimental setup and outcomes of routine in Sect. V. Finally, in Sect.VI. and Sect. VII. presents conclusions and outlines ideas for future work.

II. RELATED WORK

According to Sharmila Kishor Wagh and et. al. [1] proposed Survey on Intrusion Detection System using Machine Learning Techniques. Detecting the anomalies based on three techniques i.e. statistical based, knowledge based and last is machine learning based. This paper reviews different machine approaches for Intrusion detection system. Main advantage of machine learning is flexibility and adaptability capture of interdependencies. Based on the training data set, this model is automatically build using machine learning techniques for intrusion detection. This paper also presents the system design of an Intrusion detection system to reduce false alarm rate and improve accuracy to detect intrusion. Authors have presented an overview of machine learning technologies which are being utilized for the detection of attacks in IDS and system design of effective IDS. This paper also used data mining techniques to find patterns and intelligent relationships in data and infer rules that allow the prediction of future result. The security of information in computer based systems is a major concern to researchers. The work of IDS and methodologies which has been a major focus of information security related research. Machine learning is a vast and advanced field still relatively immature and definitely not optimized for IDS.

According to Sharmila Kishor Wagh [2] proposed an effective semi-supervised method to reduce false alarm rate and to improve detection rate for IDS. Because of complicated chain of computers, there are number of possibilities increased for intrusions and attacks. So, there is a need to find out the solution to secure the computers. Now a days, many industries faces different kinds of attacks. Therefore Intrusion Detection System(IDS) plays very important role for computer security. Machine learning is very efficient technique to solve the problem of IDS. It is very expensive to get the labelled data as well as it takes more time.

Sharmila Kishor Wagh [4] proposed Effective semi-supervised approach towards intrusion detection system using machine learning techniques, this system presents a new semi-supervised approach for the intrusion detection, which decreases false alarms effectively with the capacity of a high detection rate. Semi - supervised learning approach used a less amount of labelled data and a large quantity of unlabelled data. To improve classification accuracy of semi-supervised IDS, presented the model used, named as a novel self-learning construction. Train data is given as a input to the system which is labelled data. The obtained output from the training data is a input to the supervised classifier with unlabelled dataset for testing. For data correct predicted label entropy is calculated. Most confident data is added in training input for threshold calculation. Entropy is used for selecting most confident data from correctly classified test dataset. Various statistical methods used for new data selection. Optimizing the security administrator's efforts and creating the alert mechanism is improved the network security.

In Basant Subba et. al. [5] this paper presented a simple Artificial Neural Network (ANN) based IDS model. The feed forward and the back propagation algorithms used by presented Intrusion Detection System (IDS) with different optimization techniques. This technique used for minimizing the overall computational overhead, also maintain a high performance level. Experimental results shows that the performance in terms of accuracy and detection rate of the proposed ANN based IDS model on NSL-KDD dataset is at par and in some cases better than other IDS

models. This proposed model is very suitable for analysing the detection of intrusions and real time deployment.

Levent Koc and Alan D. Carswell proposed work given in [6] utilizes data mining procedures in intrusion discovery frameworks for the grouping of the system occasions as either typical or assault. Naive Bayes (NB) strategy is a straightforward, effective and well known information mining technique that is based on contingent autonomy of characteristics presumption. Hidden Naive Bayes (HNB) is an expanded type of NB that keeps the NB's effortlessness and proficiency while unwinding its freedom presumption. frameworks exploratory research guarantees that the HNB paired classifier model can be connected to interruption discovery issue. Test comes about utilizing great KDD 1999 Cup interruption discovery dataset show that HNB double classifier has better execution regarding location precision contrasted with the conventional NB classifier. Framework disclosed the expanding need to apply information mining strategies to arrange organize assault occasions. A straightforward and broadly utilized information mining strategy is checked which is called Naive Bayes (NB) classifier, it shows dependency on the freedom of qualities suspicion. A paired classifier display in view of the Hidden Naive Bayes (HNB) technique as an augmentation to NB to decrease its naivety supposition is presented. s connected this new classifier strategy to the testing system interruption recognition issue and tried execution of framework's model utilizing the very much perceived KDD'99 interruption identification dataset. framework's test ponder results demonstrate that the HNB twofold characterization display expanded with EMD discretization and CONS highlight determination channel strategies has better general outcomes in wording of detection accuracy, error rate and area under ROC curve than the traditional NB model.

Eman Abd EI Raoof Abas [7] used artificial immune system network based intrusion detection. In framework's structure authors propose utilizing KDD Cup database set for intrusion identification and apply R-piece calculation of counterfeit invulnerable framework system, it is utilized for anomaly discovery .An upgraded highlight determination of harsh set hypothesis utilized for improving tedious. Around the productive execution of interruption location authors accomplish the exactness and less tedious in discovery activities. Authors make similarity between interruption identification framework and fake invulnerable framework, these assistance us to accomplish framework's objective. AIS give speculation, multilevel guard and collaboration between cells. RST Solve the issue of the unpredictability of KDD cup informational collection by diminishing 41 elements to six components . Enhanced RST likewise utilized for increment the execution of IDS by adding distinctive weights to the estimations of the six elements .The rate of true positive(TP) and true negative(TN) become relatively high.

Naila Belhadj Aissa and Mohamed Guerroumi [8] focus on a grouping based identification procedure utilizing a hereditary calculation named Genetic Clustering for Anomaly-based Detection (GC-AD) is proposed. GC-AD utilizes a divergence measure to frame k bunches. It, at that point, applies a hereditary procedure where every chromosome speaks to the centroids of the k groups. Authors acquaint a certainty interim with refine the bunches so as to get segments that are more homogeneous and register, expand bunch fluctuation as. The exactness of framework's system is tried on various subset from KDD99 dataset. The outcomes are talked about and contrasted with kmeans grouping calculation. This paper presents a peculiarity based recognition conspire that uses an unsupervised bunching method joined with a hereditary procedure. The main purpose for CG-AD (Clustering Genetic for Anomaly-based Detection) is to get an ideal homogenous apportioning of typical and oddity cases. Because of the calculation of a certainty interim in the fitness function, a bunch of rejected instances is generated.

Mohammed A. Ambusaidi et. al. [9] proposed a shared data based calculation that scientifically chooses the ideal component for characterization. This shared data based element choice calculation can deal with directly and nonlinearly subordinate information highlights. Its adequacy is assessed in the instances of system interruption discovery. An Intrusion Detection System (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is manufactured utilizing the components chose by framework's proposed highlight determination calculation. The execution of LSSVM-IDS is assessed utilizing three

interruption identification assessment datasets, to be specific KDD Cup 99, NSL-KDD and Kyoto 2006+ dataset. The assessment comes about demonstrate that framework's element choice calculation contributes more basic components for LSSVM-IDS to accomplish better precision and lower computational cost contrasted and the cutting edge techniques. In this paper, an administered channel based component determination calculation has been proposed, to be specific Flexible Mutual Information Feature Selection (FMIFS). FMIFS is a change over MIFS and MMIFS. FMIFS proposes a change to Batiste's calculation to lessen the excess among highlights. FMIFS disposes of the excess parameter _ required in MIFS and MMIFS. Practically there is no particular system or rule to select the appropriate values for this parameter.

III. PROPOSED SYSTEM

The proposed system worked with ensemble approach, 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. The attribute selection done with the help of weka 3.7 tool. The below figure show the proposed system architecture and execution.

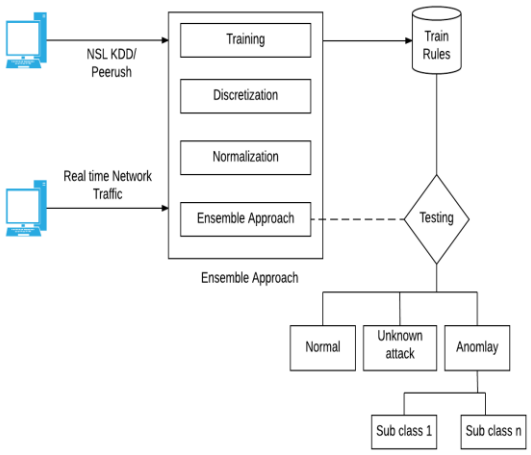

Figure 1 : Proposed System Architecture

A. System Modules:

Basically there are two phase in the proposed system, we have taken NSL KDD dataset for system training as well testing purpose.

Module 1: Training Phase

Step 1: Upload training data for feature extraction.
Step 2: Apply discretization approach on dataset.
Step 3: Apply Normalization approach on dataset.
Step 4: Generate normalize and discretize dataset

Module 2: Testing Phase

Step 1: Upload Testing data for detection.
Step 2: Apply discretization approach on dataset.
Step 3: Apply Normalization approach on dataset.
Step 4: Generate normalize and discretize dataset.
Step5: Apply ensemble approach {NB, J48, Winnow, JRip, PART, IBK, BayesNet}
Step 6: classify all attacks.
Step 7: Show detection results.

IV. DATASET DETAILS

The KDD cup 99 dataset [9] contains several statistical analyses which affects to detect the accuracy of many IDS model. NSL-KDD data set [3] is a modified version of its precursor. It consist, important records of the KDD data set. Downloaded files are listed in the table 1.

Table 1 : List of NSL-KDD Dataset Files and Their Description

S. No.	File Name	Description
1	KDDTrain+.ARFF	The complete NSL-KDD train set with binary labels in ARFF format
2	KDDTrain+_20Percent.ARFF	A 20% subset of the KDDTrain+.arff file
3	KDDTest+.ARFF	The complete NSL-KDD test set with binary labels in ARFF format
4	KDDTest-21.ARFF	A subset of the KDDTest+.arff file which does not include records with difficulty level of 21 out of 21

There are 41 attributes presented to clarify number of features of the flow. In this labels are given to each

records as an attack type or as normal. There are 4 attack classes which are further grouped as DoS, Probe, R2L and U2R.

V. RESULTS AND DISCUSSION

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 can't be focus on misclassified 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.

The below figure 2 show existing systems of the overall attack found ratio in ensemble approach with the help of table 3. In result, the X-axis presents number of packets (network flow size) provided for finding results and Y-axis presents how many packets correctly classified in particular attack types.

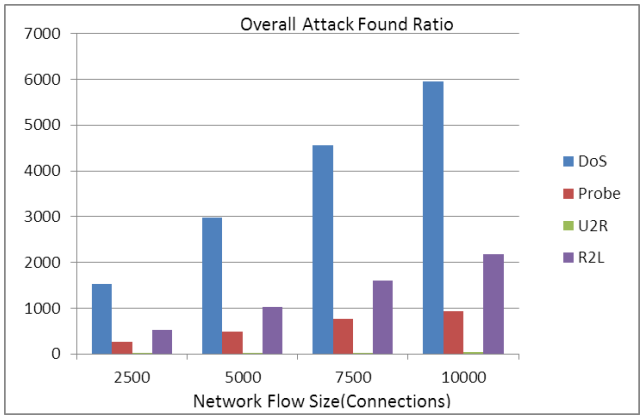

Figure 2 : Overall attack found ratio in ensemble approach

In the second experiment, in figure 3 we have test existing system with different classifiers on weka environment. In that true positive and false negative find according to particular classifiers. It shows detection rates of anomalies. This result analyze by table 2. In result, the X-axis presents number of classifiers with particular attack and Y-axis presents percentage of data classified as true positive or false negative.

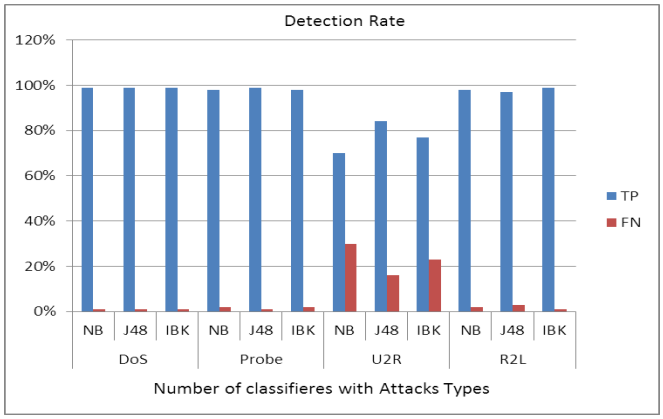

Figure 3 : Average detection rate for three classifiers in ensemble approach

Network flow Size (Connections)	Attacks Found			
	DOS	Probe	U2R	R2L
2500	1520	265	8	523
5000	2987	498	17	1029
7500	4562	765	22	1602
10000	5960	936	35	2188

Table 2 : Confusion matrices base system analysis

Here the table 2 and table 3 show the overall detection rate with false ratio of system.

Packet Size	Classifier	DOS		Probe		U2R		R2L	
		TP	FN	TP	FN	TP	FN	TP	FN
5000	NB	99%	1%	98%	2%	70%	30%	98%	2%
5000	J48	99%	1%	99%	2%	84%	16%	97%	3%
5000	Winnow	99%	1%	99%	1%	87%	13%	98%	2%
5000	JRip	99%	1%	98%	2%	71%	29%	99%	1%
5000	PART	99%	1%	98%	2%	76%	24%	97%	3%
5000	IBK	99%	1%	98%	2%	77%	23%	99%	1%
5000	BayesNet	99%	1%	98%	2%	67%	33%	98%	2%

Table 3 : Overall attack found ratio with ensemble approach

VI. 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 overcome the limitations of the previous methods. Lots of complexities are present in this modern era, so the perfect solution and the perfect detection is not an attainable goal.

After the completion of this survey we can conclude there are different techniques that can used for detection, some soft computing as well as some classification approaches are effective for detect the different attacks. Some system has work on signature base anomaly detection with creation of different rules. KDD cup dataset has used for training and testing purposed. Finally every system shows the maximum accuracy for attack detection, but none of these are has focused on unknown attack detection or misuse detection.

VII. FUTURE WORK

We propose to embed the multi-classification approach (ensemble) with different algorithm on NIDS as well as HIDS as future enhancement. The second challenging task for future enhancement is to detect and prevent the attack into both online and offline environment with forensic approach.

VIII. REFRENCES

[1] 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.

[2] Wagh SK, Kolhe SR, "Effective intrusion detection system using semi-supervised learning", In Data Mining and Intelligent Computing (ICDMIC), International Conference on 2014 Sep 5 (pp. 1-5). IEEE, 2014.

[3] Wagh S, Khati A, Irani A, Inamdar N, Soni R., "Effective Framework of J48 Algorithm using Semi-Supervised Approach for Intrusion Detection", International Journal of Computer Applications, 1;94(12), Jan 2014.

[4] 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.

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

[6] Levent Koc and Alan D. Carswell , "Network Intrusion Detection Using a HNB Binary Classifier", IEEE 2015.

[7] Eman Abd EI Raoof Abas, "Artificial immune system based intrusion detection: anomaly detection and feature selection", IEEE, 2015.

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

[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.

[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] Hachmi Fatma and Limam Mohamed , "A two-stage technique to improve intrusion detection systems based on data mining algorithms", IEEE, 2013.

[14] Saeed Khazaee and Maryam Sharifi Rad , "Using fuzzy c-means algorithm for improving intrusion detection performance", IEEE, 2013.

[15] Shadi Aljawarneha,, Monther Aldwairi, Muneer Bani Yassein , "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model", Elsevier ,2017.

[16] Eduardo K. Viegas, Altair O. Santin , Luiz S. Oliveira , "Toward a reliable anomaly-based

intrusion detection in real-world environments “, Elsevier , 2017.

[17] Amira SayedA. Aziz, Sanaa EL-OlaHanafi, Aboul EllaHassanien, “Comparison of classification techniques applied for network intrusion detection and classification”, Elsevier , 2016.

[18] Weiwei Chen, Fangang Kong, “A Novel Unsupervised Anomaly Detection Approach for Intrusion Detection System”, IEEE, 2017.