Approach for Intrusion Detection System using Recurrent Neural Network

Priya U Kadam[1], Research Scholar Jaipur National University India
Dr. Shilpa Sharma[2], Research Scholar Jaipur National University India

Abstract- Intrusion detection is prominent up and coming zone, as an ever increasing number of complex information is being put away and handled in arranged frameworks. With extensive use of internet service, there is constant threat of intrusions and misuse. Thus Intrusion Detection system is most important component of computer and its network security. Intrusion Detection System is software based monitoring mechanism for a computer network that detects presence of malevolent activity in the network. IDS system have gathered consideration by maintaining high safety levels ensuring trusted and safe announcement of the information between dissimilar organizations. Intrusion detection systems classify computer behavior into two main categories: normal and distrustful activities. Many perspectives for intrusion detection have been proposed before but none shows acceptable results so we investigate for better upshot in this field. The proposed study likewise takes a diagram of various 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 wide augment utilization of computer systems in today's general public, especially the sudden surge in hugeness of e-business to the world riches, has made PC system asylum a global priority. Since it is not in fact practicable to fabricate a plan without any vulnerabilities, interruption recognition has occur for an essential range of analyze. For the most part a gatecrasher is characterized as a framework, project or person who tries to and may get to be unbeaten to break into a data framework or execute an activity not formally permitted. We imply interruption as any arrangement of procedures that endeavor to trade off the honesty, privacy, or availability of a network asset. The demonstration of identifying procedures that endeavor to trade off the honesty, attentiveness, or accessibility of a network asset can be implied as interruption discovery. An interruption location framework is a gadget or programming application that screens system and/or framework exercises for resentful exercises or approach infringement and produces data to an administration position. Interruption 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 [2] is a plan for distinguishing interruptions and reporting them definitely to the best possible power.

II. LITERATURE SURVEY

Zero-day Malware Detection based on Supervised Learning Algorithms of API call Signatures, Igor Santos et al.: In this paper, malware is detected by the dynamic analysis of the samples where they have considered the frequency of API calls as their feature. The proposed system unpacks malware and disassembles the binary executable so that it can retrieve the assembly code. The assembly program is further used to pull out the API calls from the code and also some relevant machine code is extracted. These features are used in combination. Finally, they have mapped API calls with MSDN library to analyze the malicious code. They have used Similarity-based detection methods for identifying unknown malware and classifying them into their respective families based on these features. In this section we refer to the significant past literature that uses the various intrusion detection techniques. Most of the researchers concentrate on genetic algorithm for creating the rules. For network intrusion detection, there are many proposed algorithm using well known KDDCUP99 dataset and only few are using real time network data. Some author uses GA for deriving classification rules and for providing optimal solution. Some author uses fuzzy algorithm for defining fuzzy membership function. There are several papers related to IDS which has certain level of impact in computer and network security.

According to Manoj s. Koli [1] proposed a An Advanced method for detection of botnet traffic using Internal

Intrusion Detection, confirmed that on the advanced method for detection to improve the security by identifying and tracking the attacker using machine learning, ranking and Voronoi clustering is proposed the paper ensure reducing the size of data set and high detection accuracy. A data set called ISOT has been used keeping in mind the processing delay in the large scale network UDP and TCP are examined to recognize achieve instruction growth in network traffic is taken care of machine learning modules act like deep neural network various botnet techniques are provided DNA based method is developed by the system help. The paper also uses characteristics of the network flow to detect the botnet intrusion despite packet payload content, which helps in encryption of packet.

According to Thabet Kacem et. Al. [2] proposed An ADS-B Intrusion Detection System, an automatic dependent surveillance-broadcast IDS technique are proposed by using ADS-B techniques. HMAC data set has been used to increase the performance of air traffic control. The methods operate with minimal overhead. The future scope says for ADS-B position to be valid, its distance from the corresponding one at a time task to be within the safe zone. ADS-B as emerged as an alternative to current radio, radar standards in aircraft signaling superior location accuracy are the provided by GPS using the cyber-physical environment the attack detection is confirmed. A mechanism is proposed to exchange the keys used for the HMAC algorithm securely. ATC Centre initiates firm handshakes with ATC's that control another zone in the flight path to transfer the private key over public key infrastructure (PK1) schemes.

Shengyi Pan et. Al. [3] Developing a Hybrid Intrusion Detection System using Data Mining for power system. This paper stated that using common path mining a hybrid IDS using data mining is developed for a power system that uses data logs the approach is an automated approach to build the hybrid IDS. One of the important advantages is detection accuracy which is up to 73%. But this method is not at all suitable for big data problem capturing such as data logs is also tricky. The system leverages features of signature-based and specification based IDS. The data mining technique that aggregates audit logs from multiple system devices to learn the standard path. The automated approach eliminates the need to manually analysis and manually code pattern.

According to Mehdi Ezzarii [4] proposed a system The well-known genetic algorithm is based on gene reproduction and mutation. Recent research has pointed out that additional information embedded alongside individual chromosomes transmits data into future offspring. This additional transmission of information into child generations outside DNA is known as epigenetic. Additional information is considered as the epigenetic factor that helps us to define randomness crossover and mutation used in classical genetic algorithm. This paper also presents a state of art where we try to explore epigenetic algorithms within the context of Intrusion Detection System. We discuss the methodology used in genetic algorithm and how our approach can perform detection of intrusions for an efficient security

According to Flow anomaly based. [5], this paper based on the flow anomaly Intrusion Detection System for Android mobile devices this approach uses ANN (Artificial Neural Network) on Android Operating System to detect anomaly behaviors in android mobiles. Accuracy and detection rate of this methodology reaches 85% and 81% respectively. Imitation is considered regarding CPU, memory and battery power this work endeavors to identify a lightweight, scalable an efficient IDS for an android environment various services are provided for addressing public attacks. The data streams are analysed by using efficient machine learning algorithms. The future scope includes the improvement in accuracy and detection rate.

Trae Hurley et. Al. proposed in [6] HMM Based Intrusion Detection System for software-defined networking, A Hidden Markow model based IDS is developed for software-defined networking (SDN). SDN network can help monitor the overall security of a system by analyzing the web as a hole and making choices to defend the network based on the data from the entire network it includes uses of ANN IDS. This methodology allows greater dynamic control of a networking environment. The paper consists of the advantages like increased in the range of activities and also is the increase of security application. It has shown that machine learning application holds the potential to be used to access the risk in networking environment for the future scope expanding the feature vector used by HMM in determining the maliciousness of a set data are to be added.

According to Mariem Belhor et. Al. [7] proposed a system Intrusion Detection based on genetic fuzzy classification system, Fuzzy systems have been used to solve several classification problems. Genetic-fuzzy systems hybridize the approximate reasoning method of fuzzy systems with the learning capability of evolutionary algorithms. In this paper a novel intrusion detection method is presented, capable of detecting normal and intrusive behaviours, which extracts both accurate and interpretable fuzzy IF-THEN rules from network dataset for classification. This method uses the fuzzy association rule based classification method for high dimensional problems based on three stages to obtain an accurate and compact fuzzy rule based classifier with a low computational cost.

According to Sharad Awatade et. Al. [8] proposed a system Improved EAACK: Develop Secure Intrusion Detection System for MANETS using hybrid cryptography. EAACK

uses the concepts of hybrid cryptography techniques to reduce the network overhead caused by digital signature. By providing Hybrid Cryptography technique to EAACK Scheme, it will become difficult for attacker to break the network as well as retrieved the data.

According to Mayank Agarwal et. Al. [9] Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system, this paper based on the Intrusion detection system for PS-Poll DOS attack in 802.11 networks using real-time discrete event system. This approach uses RTDES on real-time discrete event system for detecting DOS attack. One of the important advantages is high accuracy and detection rate, but one of the major drawbacks is a loss of frames. Detect the PS-DOS attack require encryption change in protocol or installation of proprietary hardware.

Md Zahangir Alom et. Al. [10] proposed a Network Intrusion Detection for cyber security on neuromorphic computing system. Referred that cyber security is severe issues in the cyberspace. The paper includes the demonstration of a neuromorphic cognitive computing approach for network IDS for cyber security using deep learning. This method uses Discrete Vector Factorization. The NSL-KDD dataset is used to increase accuracy and classification up to 90.12% and 81.31% respectively. Deep learning achieves human-level performance in particular for recognition tasks, in-depth learning approach combining the features of extraction classification. The future scope includes the challenge of determining the representation of data in spiking format for the use in the True- North-System.

III. RESEARCH METHODOLOGY

The redundant and irrelevant features in data have caused a long-term problem in network traffic classification. These features not only slow down the process of classification but also prevent a classifier from making accurate decisions, when third party generates the flash events on vulnerable network.

In this research work, system define GA based rule creation system according to their feature selection method that worked on NIDS as well as HIDS. Genetic algorithm is an optimization algorithm, which is used for finding optimal solution.

Ensemble approach with various classification algorithm will provide a best detection with NIDS in all type of sub attacks with master class.

With the proposed research work our aim to generate st ong rules and increase the detection rate for DOS, PROBE, U2R and R2L for NIDS and HIDS.

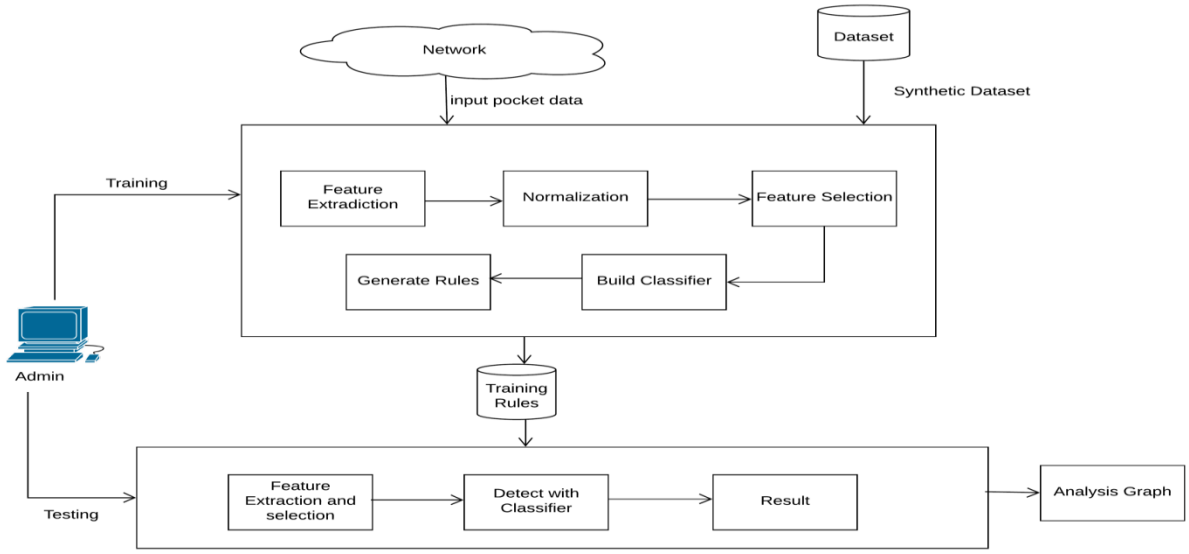

Fig.1: System Architecture

Training Phase

In this Phase, Genetic algorithm is used where, we first initialize the chromosomes and group of chromosomes we say as population is created. Once the population is created crossover is applied to obtain new generation of chromosomes. Mutation is applied for updating bit value of attributes of chromosomes randomly. The fitness function will define the fitness value of each chromosome and a selection criterion is applied for selected optimal rules. When variation is completed then Genetic algorithm will get terminated. The outputs of genetic algorithm are genetic rules. The output of genetic algorithm that is genetic rules is

given as an input to fuzzy logic. In this phase probability of each attribute is calculated which is used for classification of data as attack or normal

Step 1: System first collect network traffic from network audit data using packetX Lib and Wincap driver or some synthetic dataset like KDDCup99, NSLKDD, ISCX and WSNTrace etc.

Step 2: Select features of each connection and apply Genetic Algorithm (GA) for rule creation.

Step 3: Once rule created store it into local database directory called as BK rules.

Testing Phase:

In this Phase, Fuzzy rules are given as an input to the Neural Network algorithm for the classification of sub attack. Here system collect the network traffic data using PacketXLib and Wincap Driver. On each instance neural network algorithm will be applied. Transfer function will be used for calculating each node weight .Using Defined threshold , sub attacks can be classified.

Step 1: System collect the network traffic data using PacketX Lib and Wincap driver or NSLKDD

Step 2: Read each instance and apply ensemble (J48, ANN, NB) algorithm.

Step 3: Calculate the weight using given functions for each connection.

Step 4: Finally classify the each attack with sub attack type using define threshold (e.g. DoS, PROBE, U2R, R2L, Network attacks, Active Attack, Passive Attack, Advance attack etc)

Algorithm Recurrent Neural Network

4. Recurrent Neural Network

Input : Training Rules Tr[], Test Instances Ts[], Threshold T.

Output : Weight w=0.0

Step 1 : Read each test instance from (TsInstnace from Ts)

Step 2 : TsIns $= \sum_{k=0}^{n} \{Ak \ldots An\}$

Step 3 : Read each train instance from (TrInstnace from Tr)

Step 4 : $TrIns = \sum_{j=0}^{n} \{Aj \ldots \ldots Am\}$

Step 5 : w = *WeightCalc(TsIns, TrIns)*

Step 6 : if (w >= T)

Step 7 : Forward feed layer to input layer for feedback FeedLayer[] ← {Tsf,w}

Step 8 : optimized feed layer weight, Cweigt ← FeedLayer[0]

Step 9 : Return Cweight.

IV. DATASET DESCRIPTION

The inherent drawbacks in the KDD cup 99 dataset [4] has been revealed by various statistical analyses has affected the detection accuracy of many IDS modeled by researchers. It contains essential records of the complete KDD data set. There are a collection of downloadable files at the disposal for the researchers.

Table 1: Dataset Description

Id	Name	Description
1	KDDCUP99	41 Attributes with 23 sub classes for all 4 classes.
2	NSLKDD	41 Attributes with 38 sub classes for all 4 classes.
3	Botnet	12 attributes including class as normal and abnormal
4	ISCX	29 attributes including class as normal and abnormal
5	NUSW-NB15	It contains 49 attributes binary,0 for normal and 1 for attack records
6	WSNtrace	12 attributes including class as normal and abnormal

V. RESULTS AND DISCUSSIONS

To evaluate the proposed system performance analysis we have used various data set for system testing which is already define in table 1. Each data set contains different features as well as different kind of attacks. Once the system has train according to specific data set, it generates training rule accordingly. The average accuracy for entire system with all data set is around 90%.

In our experimental setup we have done various experiments, the confusion Matrix has been calculated for each data set according to to label assign by testing algorithm. The testing data set which is basically and label when we deals with the system testing. The classification accuracy should we generate according to two given threshold, the threshold value has set initially 0.70. The optimum threshold for this research it's around 0.60, which displays better accuracy than others. The proposed RNN algorithm is the part of deep learning which is basically supports for LSTM (Long Short Term Memory). The below Figure 2 2 shows average accuracy of proposed system with classical machine learning algorithms.

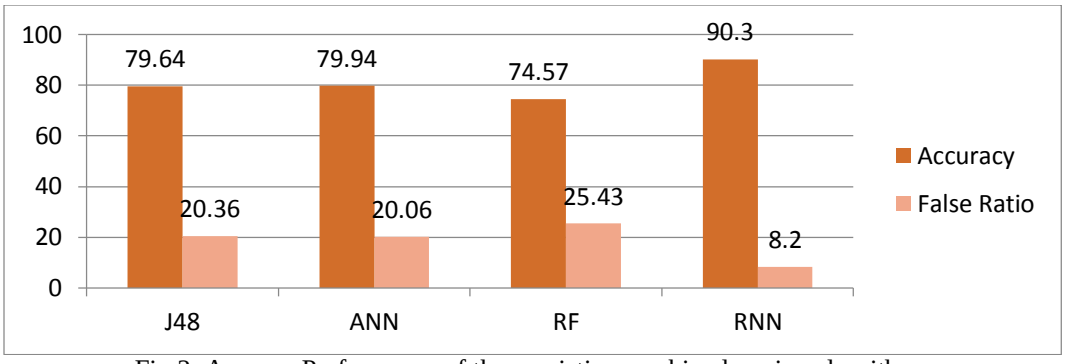

Fig.2: Average Performance of three existing machine learning algorithms

We did not compare our algorithm with other malware detection algorithms because our binary and malicious files did not match the format of the files required to run these algorithms. Additionally, it did not make sense to compare the accuracy between algorithms tested on different datasets. As a result, we compared our algorithm to other state of the art machine learning classifiers using the data obtained by extracting certain features from the original binary and malicious files. We tested the data with the support vector machine classifier [8], ANN classifier [9], and the RF classifier [10]. The resulting accuracies are displayed for the DLL features (Table 2), strings features (Table 3), and byte sequence features (Table 4).Fig 3shows the result Other Classifier Performance on DLLs whose values taken from table 2. Fig 3 is graphical representation of table 2.Fig 4 shows the result of Other Classifier Performance on Strings whose values taken from table 3. Fig 4 is graphical representation of table 3.Fig 5 shows the result of Other Classifier Performance on Byte Sequences whose values taken from table 4. Fig 5 is graphical representation of table 4.

Table 2: Other Classifier Performance on DLLs

ML Classifier	TP	TN	FP	FN	Accuracy	False Ratio
J48	466	124	58	10	79.64%	20.36%
ANN	444	100	82	32	79.94%	20.06%
RF	360	113	33	21	74.57%	25.43%

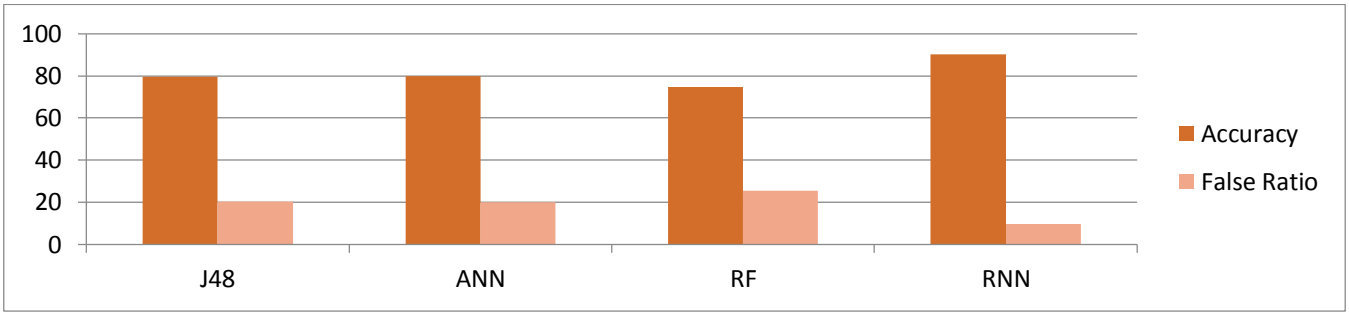

Fig.3: Other Classifier Performance on Dataset's

Table 3: Other Classifier Performance on Strings

ML Classifier	TP	TN	FP	FN	Accuracy	False Ratio
J48	440	130	52	36	86.63%	13.37%
ANN	442	125	57	34	86.17%	13.83%
RF	438	116	66	38	84.19%	15.81%

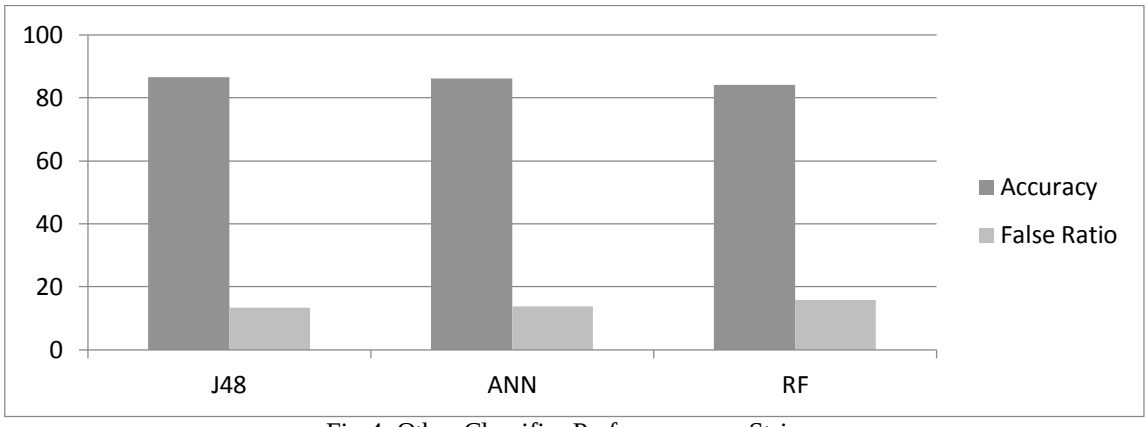

Fig.4: Other Classifier Performance on Strings

Table 4: : Classifier Performance on different WSNTrace data

ML Classifier	TP	TN	FP	FN	Accuracy	False Ratio
J48	471	90	92	5	85.26%	14.74%
ANN	460	76	106	16	81.46%	18.54%
RF	362	85	61	19	80.27%	19.73%

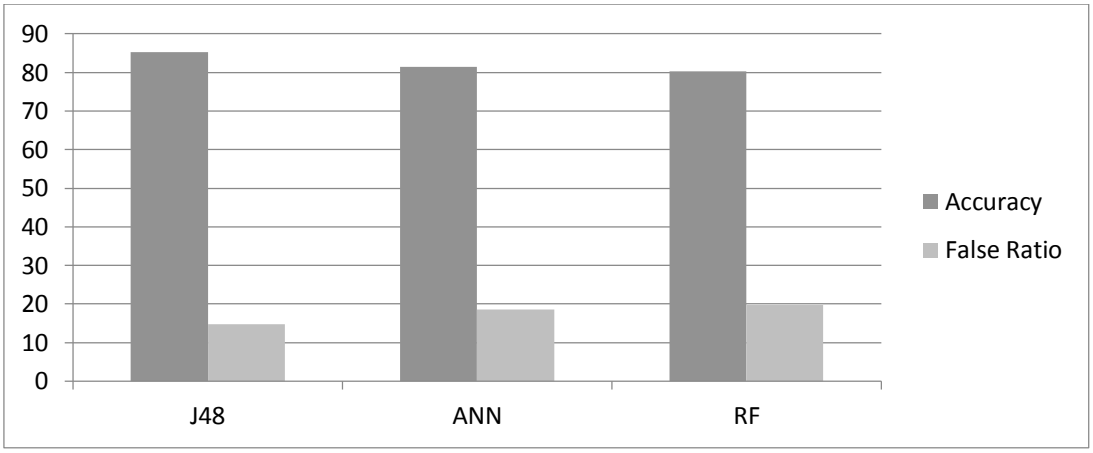

Fig.5: Classifier Performance on different WSNTrace data

Table 5: System confusion matrix with different population size

Pop Size	GA Variation	DOS		Probe		U2R		R2L	
		TP	FN	TP	FN	TP	FN	TP	FN
100	1	99%	1%	98%	2%	70%	30%	98%	2%
500	1	99%	1%	99%	2%	84%	16%	97%	3%
1000	1	99%	1%	99%	1%	87%	13%	98%	2%
2000	1	99%	1%	98%	2%	71%	29%	99%	1%
3000	1	99%	1%	98%	2%	76%	24%	97%	3%
4000	1	99%	1%	98%	2%	77%	23%	99%	1%
5000	1	99%	1%	98%	2%	67%	33%	98%	2%

Table 6: System accuracy for all attacks using different population size

Packet Size	Attacks Found			
	DOS	Probe	U2R	R2L
100	5960	940	35	2167
500	5957	936	35	2182
1000	5960	938	36	2151
2000	5960	936	35	2188
3000	5958	939	38	2171
4000	5957	941	38	2196
5000	5950	941	38	2164

The accuracy produced by the other machine learning classifiers significantly varied across the individual feature sets. For the DLL feature sets, the other machine classifiers performed roughly around the same as our classifier. Our classifier performed better than the support vector machine, ANN, and ensemble classifiers for the byte sequences feature set. Finally, all three of the other machine learning classifiers performed much better than our classifier on the strings feature set. However, our combined classifier still yields the highest accuracy of overall accuracy.

VI. CONCLUSION

Since the study of intrusion detection began to gain momentum in the security community roughly ten years ago, a number of diverse ideas have emerged for confronting this problem. Intrusion detection systems vary in the sources they use to obtain data and in the specific techniques they employ to analyze this data. Most systems today classify data either by misuse detection or anomaly detection. Each approach has its relative merits and is accompanied by a set of limitations. It is likely not realistic to expect that an intrusion detection system be capable of correctly classifying every event that occurs on a given system. Perfect detection, like perfect security, is simply not an attainable goal given the complexity and rapid evolution of modern systems.

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. REFERENCES

[1]. Manoj s. Koli, Manik K. Chavan, "An Advanced method for detection of botnet traffic using Interhnal Intrusion Detection", 2017 International Conference on (ICICCT), March 10-11, 2017, Sangli, India.

[2]. Thabet Kacem, Duminda Wijesekera, Paulo Costa, Alexander Barreto, "An ADS-B Intrusion Detection System", 2016 IEEE on ISPA, 2016, Fairfax, Virginia.

[3]. Shengyi Pan, Thomas Morris, Uttam Adhikari, "Developing a Hybrid Intrusion Detection System using Data Mining for power system", IEEE Transactions on, vol. 6, issues. 6, Nov. 2015.

[4]. Mehdi Ezzarii, Hamid Elghazi, Hassan El Ghazi, Tayeb Sadiki, "Epigenetic Algorithm for performing Intrusion Detection System",2016 International Conference on ACOSIS, Oct17- 19,2016, Rabat, Morocco.

[5]. Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.

[6]. Trae Hurley, Jorge E. Perdomo, Alexander Perez-pons, "HMMBased Intrusion Detection System for software-defined networking", 2016 15th IEEE Conference on Machine Learning and Application, Dec 18-20, 2016, Miami, Florida.

[7]. Mariem Belhor, Farah Jemili, "Intrusion Detection based on genetic fuzzy classification system", 2016 IEEE 13th International Conference on Computer Systems and Application (AICCSA), Nov 29 2016-Dec 2, 2016, Sousse, Tunisia.

[8]. Sharad Awatade, Shweta Joshi. "Improved EAACK: Develop Secure Intrusion Detection System for MANETS using hybrid cryptography", 2016 International Conference on computing communication control and automation (ICCUBEA), Aug 12-13, 2016, Maharashtra, India.
[9]. Mayank Agarwal, Sanketh Purwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system",IEEE,vol.4,issue4,2017.
[10]. Md Zahangir Alom, Tarek m. Taha, "Network Intrusion Detection for cyber security on neuromorphic computing system", 2017 International Joint Conference on Neural Networks (IJCNN), May 14-15,2017, USA..
[11]. Dipika Narsingyani, Ompriya Kale, Optimizing False Positive In Anomaly based Intrusion Detection using Genetic Algorithm, IEEE 2015
[12]. Mostaque Md. Morshedur Hassan, "Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 7, September 2013.
[13]. Samaneh Rastegari, Chiou-Peng Lam, Philip Hingston, Statistical rule learning approach to neural network, IEEE 2015.
[14]. Preeti Singh, Amrish Tiwari, An Efficient Approach for Intrusion Detection in Reduced Features of KDD99 using ID3 and classification with KNNGA, IEEE 2015.
[15]. Geethapriya Thamilarasu, Genetic Algorithm based Intrusion Detection System for Wireless Body Area Networks, IEEE 2015

.