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. 9, Issue 3, March 2020

Intrusion Detection using Soft Computing Approaches: an Overview

Gauri Vilas Rasane, Dr. Sunil Rathod

P.G. Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engg, Pune, India

Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engg, Pune, India

ABSTRACT: The wide spread use of internet services, there is sudden surge in importance of e-commerce to the world economy made network security a priority at international point of view. Since it is not feasible to build a complete secure system, IDS has become a crucial area of research. 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. Each day new type of attacks is being faced by industries. 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

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. 9, Issue 3, March 2020

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.

II. RELATED WORK

In [1] Ali A. Ghorbani , Wei Lu and Mahbod Tavallaee proposd a Network Intrusion Detection and Prevention In first section system discusses different attack taxonomies and then presents the details of a large number of known vulnerabilities and strategies to lunch attacks. Then detection approach describes the attack analysis method used in IDS. In second phase different detection approaches that are currently available to detect intrusions and anomalous activities are given. Misuse, rule-based, model-based, anomaly and specification-based detection approaches are explained in detail. In third section provide the Intrusion detection systems collect their data from various sources. Such sources include log files, network packets, system calls, or a running code itself. In this chapter, we provide detail information as to how this very important step in the life of different intrusion detection systems (i.e. host-based, network-based and application-based) can be accomplished.
The fourth section denote the Theoretical Foundation of Detection, understanding the strengths and weaknesses of the machine learning and data mining approaches helps to choose the best approach to design and develop a detection system. Several approaches to the intrusion detection research area are introduced and analyzed in this area. The fifth section presented the Architecture and Implementation, intrusion detection systems can be classified based on their architecture and implementation. This classification usually refers to the locus of the data collection and analysis. This chapter introduces the centralized, distributed and agent based intrusion detection systems.

In sixth chapter Alert Management and Correlation Intrusion Detection Systems trigger too many alerts that usually contain false alerts. Decreasing false positives and improving the knowledge about attacks provides a more global view of what is happening in a network. Alert management and correlation addresses the issue of managing large number of alerts by providing a condensed, yet more useful view of the network from the intrusion standpoint. The correlation function can relate different alerts to build a big picture of the attack. The correlated alerts can also be used for cooperative intrusion detection and tracing an attack to its source. This chapter introduces different approaches to cluster, merge and correlate alerts. The evaluation criteria provides a number of approaches that can be used to evaluate the potential intrusion detection systems for accuracy, performance, completeness, timely response, cost and intrusion tolerance & attack resistance. The objective is to enable automated reasoning and decision-making to aid in human-mediated or automatic response. The cost benefit analysis of response is the key for effective response.

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. 9, Issue 3, March 2020

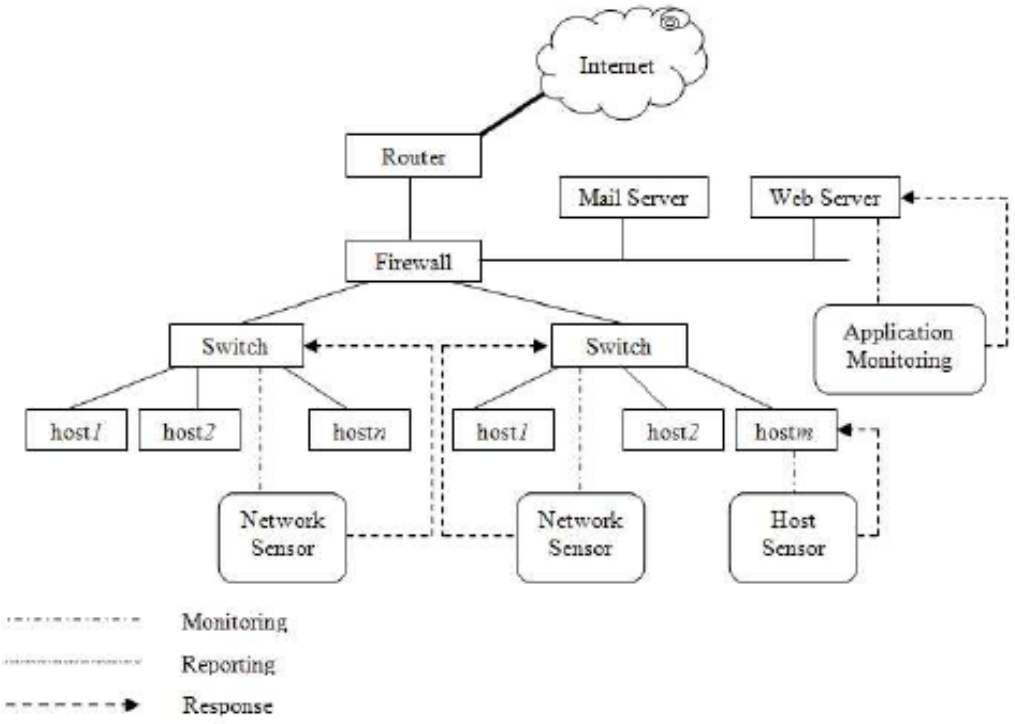


Figure 1: Distributed intrusion detection architecture

This system focus on distributed intrusion detection system, the above diagram shows the overall system.
In [2] Rafeeq Ur Rehman proposed a Intrusion Detection Systems with Snort

This system provides the basic information about how to build and install Snort itself. Using the basic installation and default rules, we will be able to get a working IDS. It can create the Snort rules, different parts of Snort rules and how to write your own rules according to your environment and needs. It is very important, as writing good rules is the key to building a detection system. In fourth section of this book about input and output plug-ins. Plug-ins are parts of the software that are compiled with Snort and are used to modify input or output of the Snort detection engine. Input plug-ins prepare captured data packets before the actual detection process is applied on these packets. Output plug-ins format output to be used for a particular purpose. For example, an output plug-in can convert the detection data to a Simple Network Management Protocol (SNMP) trap. Another output plug-in is used to log. It also provides information about using MySQL database with Snort. MySQL plug-in enables Snort to log data into the database to be used in the analysis later on. It will find information about how to create a database in MySQL, configure a database plug-in, and log data to the database. describes ACID, how to use it to get data from the database we configured in fourth section, and how to display it using Apache web server. ACID is a very important tool that provides rich data analysis capabilities. we can find frequency of attacks, classify different attacks, view the source of these attacks and so on. ACID uses PHP (Pretty Home Page) scripting language, graphic display library (GD library) and PHPLOT, which is a tool to draw graphs. A combination of all of these results in web pages that display, analyze and graph data stored in the MySQL database. Snort output data into databases. The overall system is devoted to information about some other useful tools that can be used with Snort.

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. 9, Issue 3, March 2020

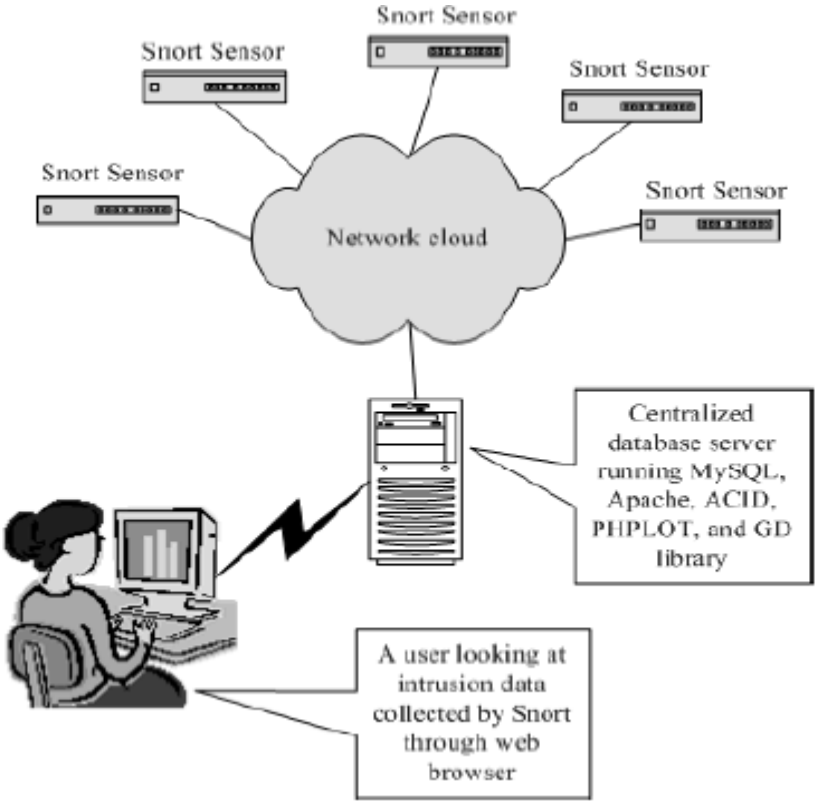


Figure 2 Multiple Snort sensors in the enterprise logging to a centralized database server.

There are alerts are any sort of user notification of an intruder activity. When an IDS detects an intruder, it has to inform security administrator about this using alerts. Alerts may be in the form of pop-up windows, logging to a console, sending e-mail and so on. Alerts are also stored in log files or databases where they can be viewed later on by security experts. we will find detailed information about alerts later in this book. Snort can generate alerts in many forms and are controlled by output plug-ins. Snort can also send the same alert to multiple destinations. For example, it is possible to log alerts into a database and generate SNMP traps simultaneously. Some plug-ins can also modify firewall configuration so that offending hosts are blocked at the firewall or router level.

In [3] Stephen Northcutt, Judy Novak Network Intrusion "System first phase, ranging from an introduction to the fundamental concepts of the Internet protocol to a discussion of Remote Procedure Calls (RPCs). We realize that it has become stylish to begin a book saying a few words about TCP/IP, but the system Judy and I have developed has not only taught more people IP but a lot more about IP as well—more than any other system ever developed. We call it "real TCP" because the material is based on how packets actually perform on the network, not theory. Even if we are familiar with IP, give the first part of the book a look. We are confident we will be pleasantly surprised. Perhaps the most important chapter in Part I is , "Stimulus and Response." Whenever we look at a network trace, the first thing we need to determine is if it is a stimulus or a response. This helps we to properly analyze the traffic. Please take the time to make sure we master this material; it will prevent analysis errors as we move forward.

The book continues in Part II, "Traffic Analysis" with a discussion of traffic analysis. By this, we mean analyzing the network traffic by consideration of the header fields of the IP and higher protocol fields. Although ASCII and hex signatures are a critical part of intrusion detection, they are only tools in the analyst's tool belt. Also in Part II, we begin

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

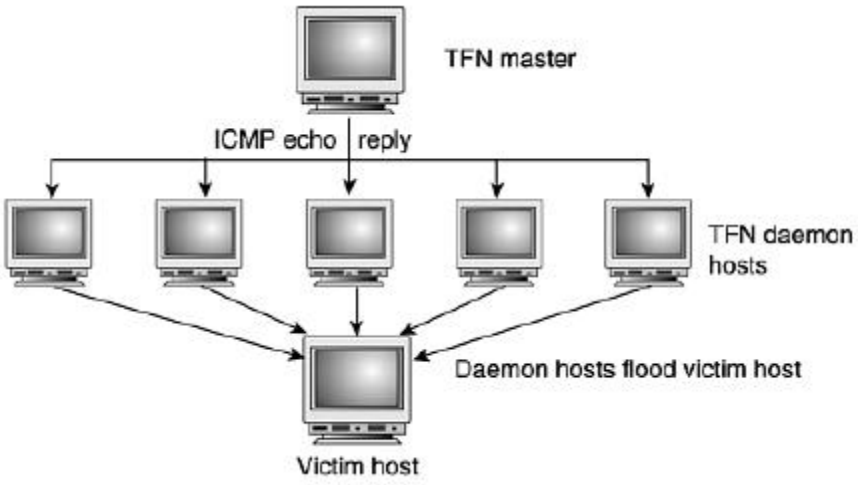
International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

to show we the importance of each field, how they are rich treasures to understanding. Every field has meaning, and fields provide information both about the sender of the packet and its intended purpose. As this part of the book comes to a close, we tell we stories from the perspective of an analyst seeing network patterns for the first time. The goal is to help we can prepare for the day when we will face an unknown pattern. Although there are times a network pattern is so obvious it almost screams its
message, more often we have to search for events of interest. Sometimes, you can do this with a well-known signature, but equally often, we must search for it. Whenever attackers write software for denial of service, or exploits, the software tends to leave a signature that is the result of crafting the packet. This is similar to the way that a bullet bears the marks of the barrel of the gun that fired it, and experts can positively identify the gun by the bullet. In Part III of the book,
"Filters/Rules for Network Monitoring" we build the skills to examine any field in the packet and the knowledge to determine what is normal and what is anomalous. In this section, we practice these skills both with TCPdump and also Snort. In Part IV, we consider the larger framework of intrusion detection. We discuss where we should place sensors, what a console needs to support for data analysis, and automated and manual response issues to intrusion detection. In addition, this section helps arm the analyst with information about how the intrusion detection capability fits in with the business model of the organization.



Tribe Flood Network TFN

Figure 3: Statistical Techniques for Network Security

First, statistical models, particularly those intended for classification and profiling purposes, should account for particular features of the organization of the data. An important aspect of data used for profiling users relates to the "clustering" of network traffics within a group (or sub-network system), and statistical models used for profiling should take into account the group effect, which is known as intra-group correlation. In addition to the clustered nature of the data, the classification model ought to be able to account for the differential amounts of information across groups, measured in terms of the number of observations per group. If two groups have the same anomaly-free rate, there would be greater confidence in the estimate for the group having large volume traffic than for the one with small volume traffic. Credible statistical models must be designed to accommodate groups with widely varying sample sizes. Groups with less traffic may have observed rates at the extreme ends of the range, but such rates may not accurately reflect their "true" behavior. Although some methodologies specifically exclude groups with fewer observations than a minimum threshold, it is almost impossible to remove all of the variability in sample sizes across groups.

Second, hierarchical regression models can address the design issues that typically occur in profiling efforts, in that clustered data and differential information can be addressed by hierarchical generalized linear models. These models

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. 9, Issue 3, March 2020

are used commonly in education and healthcare, where students or patients are clustered within classrooms or hospitals and classroom or hospitals sizes vary. Hierarchical models explicitly quantify intra-group variation and produce better group-specific estimates for small groups. This particular approach avoids "regression to the mean," a statistical concept describing the tendency for groups who have been identified as outliers in the past to become less extreme in the future. This approach also affords a more realistic assessment of the role of chance in the observed variation between groups.

Third, with regard to these model performance measures, classification models should be evaluated by measures of discrimination, calibration, and goodness-of-fit. The decision about what constitutes a "good" or "good enough" model will be based more on subjective considerations than on predefined criteria, but the model performance will depend on the degree to which traffic characteristics contribute to the outcome and the availability of variables that reflect variables associated with the outcome. Also, these models should be developed and validated in different samples to assess robustness, and such evaluations should be conducted repeatedly. If validation has not been performed, then that should also be reported. In addition, models developed from traffic data from one network system should be validated against data from a different network system as possible. This validation should include a comparison of how much agreement in classification exists between two systems' models.

Fourth, all models are not good and there is no one "gold-standard" model that can be used to compare the model performances. A stream of network traffic data with many positive predictor variables might not represent a true attack, and a look-alike normal stream could present a novel attack due to uncertain factors and users changing their behavior. In general, increasing the sensitivity could reduce the false positive alarm rate, and increasing the specificity could reduce the false negative alarm rate. The objective of a good statistical model is to demonstrate high values in sensitivity, specificity, and correctly classified rate. To achieve this goal, the process of selecting predictive variables must consider the issues of sampling variability and stability of variables' statistical significance. Fifth, we may be able to develop a robust attack-specific model to detect a particular type of attack or abnormal event but we cannot develop a model that can cover all types of attacks. Such a model may not exist. Finally, model parameters should be updated frequently to take into account new attacks and user behavior changes in over time.

Both intrusion and prevention systems needs to be constructed with a hybrid approach. Use of a hybrid modeling approach means that a final classification decision on real time traffic, such as anomaly-free or anomalous, should be made based on a vote of multi-classification algorithms; the assignment of a membership of user behavior patterns for any new user or group should also be voted by multi-profiling algorithms. A hybrid model that integrates and combines more than one classification of models and algorithms has shown potential for reducing false positive and false negative alarms rates that could maximize the strengths and minimize the weaknesses of each individual classification and profiling approaches. Uncertainty should be treated carefully and rationally. Statistics is concerned with how data change people beliefs. A confidence level is the desired "certainty" we wish to have in making a conclusion about the population. Both the probability and the confidence interval provide measurements for the uncertainty. Probability tells us how likely the observed traffic belong to a particular pattern and the confidence interval tells us the error margin for the estimate of interest. Statistical simulation techniques, such as bootstrap and Monte Carlo can be used to acquire a probability of the outcome in which we are interested. Standard error can be utilized to the calculate confidence interval that depends on the confidence level selected.

The decision of whether we should use a statistical modeling approach or an alternative approach depends on many known and unknown factors, and should be gauged by the accuracy of predicting results. Although in many situations, an empirical historical data-based statistical modeling approach is used as a principal tool for intrusion detection and prevention, it can also be used as an alternative tool to benefit both researchers and network administrators for making better evidence-based decisions.

In [5] M. Jordan, S.L. Lauritzen, J.P. Lawless, V. Nair proposed Computer Intrusion Detection and Network Monitoring A Statistical Viewpoint System proposed an Ongoing information mining base IDS, they display an outline

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. 9, Issue 3, March 2020

of our exploration progressively information mining-based interruption identification frameworks (IDSs). Framework simply centered around issues identified with sending a data mining-based IDS in a continuous domain. Framework portray our ways to deal with location three sorts of issues: precision, productivity, and ease of use. To enhance precision, information mining projects are utilized to break down review information and concentrate includes that can recognize typical exercises from interruptions, they utilize counterfeit oddities alongside ordinary and/or interruption information to deliver more compelling abuse and peculiarity recognition models. To enhance productivity, the computational expenses of elements are investigated and a various model expense based methodology is utilized to create identification models with minimal effort and high exactness. One of the first things an attacker needs to do is determine information about the hosts on your network. This includes a list of the **IP** addresses that are valid and the operating systems and services running on the hosts. Network mapping refers to the act of obtaining a list of the hosts on the network, whereas probing refers to methods for determining specific information about individual machines. A typical attack involves first mapping the network to determine the active machines, followed by probing select machines to determine the operating system and services running on the machine, selecting a service for which a known vulnerability exists, and launching an attack against the selected machine and service. We will discuss each of these topics in tum. The overall system can focus the network basic intrusion creation and detections.

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

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. 9, Issue 3, March 2020

calculation. This paper proposes a peculiarity based recognition conspire that uses an unsupervised bunching approach joined with a hereditary procedure. The basic reason 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 cluster of rejected instances is created.

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. This is attractive practically speaking since there is no particular system or rule to choose the best value for this parameter.

III. DATASET DETAILS

The inherent flaws in the KDD cup 99 dataset [9] has been publicized by several statistical analyses has affected the detection accuracy of many IDS modeled by researchers. NSL-KDD data set [3] is a refined version of its precursor. It contains vital records of the complete KDD data set. There are a group of downloadable files at the disposal for the researchers. They 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+.TXT	The complete NSL-KDD train set including attack-type labels and difficulty level in CSV format
3	KDDTrain+_20Perce nt.ARFF	A 20% subset of the KDDTrain+.arff file
4	KDDTrain+_20Perce nt.TXT	A 20% subset of the KDDTrain+.txt file
5	KDDTest+.ARFF	The complete NSL-KDD test set with binary labels in ARFF format
6	KDDTest+.TXT	The complete NSL-KDD test set including attack-type labels and difficulty level in CSV format
7	KDDTest-21.ARFF	A subset of the KDDTest+.arff file which does not include records with difficulty level of 21 out of 21
8	KDDTest-21.TXT	A subset of the KDDTest+.txt file which does not include records with difficulty level of 21 out of 21

In each record there are 41 attributes clarifying different features of the flow and a label allocated to each either as an attack type or as normal. The essentials of the characteristics namely the attribute name, their description and sample data are listed in the tables. Table 2 encompasses type information of all the 41 attributes present in the NSL-KDD data set. The 42nd attribute holds data about the several 5 classes of network connection vectors and they are categorized as one normal class and four attack class. The 4 attack classes are further grouped as DoS, Probe, R2L and U2R. The explanation of the attack classes.

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. 9, Issue 3, March 2020

Table 2 : Mapping Attack Class with Attack Type

Attack Class	Attack Type
DoS	Back, Land, Neptune, Pod, Smurf,Teardrop,Apache2, Udpstorm, Processtable, Worm (10)
Probe	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint (6)
R2L	Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named (16)
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps (7)

IV. 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 4. 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.

Table 1 : Overall attack found ratio with 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

V. CONCLUSION

Since the study of intrusion detection began to gain drive in the security community coarsely ten years before, a number of diverse ideas have emerged for challenging this problem. Intrusion detection systems vary in the sources they use to get data and in the exact techniques they rendezvous to examine this data. Most systems nowadays categorize data either by abuse detection or anomaly detection. Each method has its relative advantages and is accompanied by a set of limitations. It is likely not realistic to expect that an intrusion detection system be capable of correctly categorizing each event that happens 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.

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. 9, Issue 3, March 2020

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.

REFERENCES

[1] Ali A. Ghorbani , Wei Lu and Mahbod Tavallaee proposed a Network Intrusion Detection and Prevention.
[2] Rafeeq Ur Rehman proposed a Intrusion Detection Systems with Snort.
[3] Stephen Northcutt, Judy Novak Network Intrusion.
[4] Yun Wang proposed Statistical Techniques for Network Security.
M. Jordan, S.L. Lauritzen, J.P. Lawless, V. Nair proposed Computer Intrusion Detection and Network Monitoring A Statistical Viewpoint
[5] Levent Koc and Alan D. Carswell , Network Intrusion Detection Using a HNB Binary Classifier in IEEE 2015
[6] Eman Abd EI Raoof Abas Artificial immune system based intrusion detection: anomaly detection and feature selection IEEE 2015.
[7] Naila Belhadj Aissa, Mohamed Guerroumi , A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems IEEE 2015
[8] 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
[9] Fatemeh Barani , A Hybrid Approach for Dynamic Intrusion Detection in Ad Hoc Networks Using Genetic Algorithm and Artificial Immune System in IEEE 2014.
[10] Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis , Intrusion Detection System Using Genetic Algorithm Science and Information Conference 2014
[11] 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
[12] Hachmi Fatma and Limam Mohamed , A two-stage technique to improve intrusion detection systems based on data mining algorithms IEEE 2013.
[13] Saeed Khazaee and Maryam Sharifi Rad ,Using fuzzy c-means algorithm for improving intrusion detection performance IEEE 2013
[14] Shadi Aljawarneha,, Monther Aldwairi, Muneer Bani Yassein , Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model Elsevier 2017
[15] Eduardo K. Viegas, Altair O. Santin , Luiz S. Oliveira ,Toward a reliable anomaly-based intrusion detection in real-world environments Elsevier 2017
[16] Amira SayedA. Aziz, Sanaa EL-OlaHanafi, Aboul EllaHassanien, Comparison of classification techniques applied for network intrusion detection and classification Elsevier 2016
[17] Weiwei Chen, Fangang Kong, A Novel Unsupervised Anomaly Detection Approach for Intrusion Detection System IEEE 2017