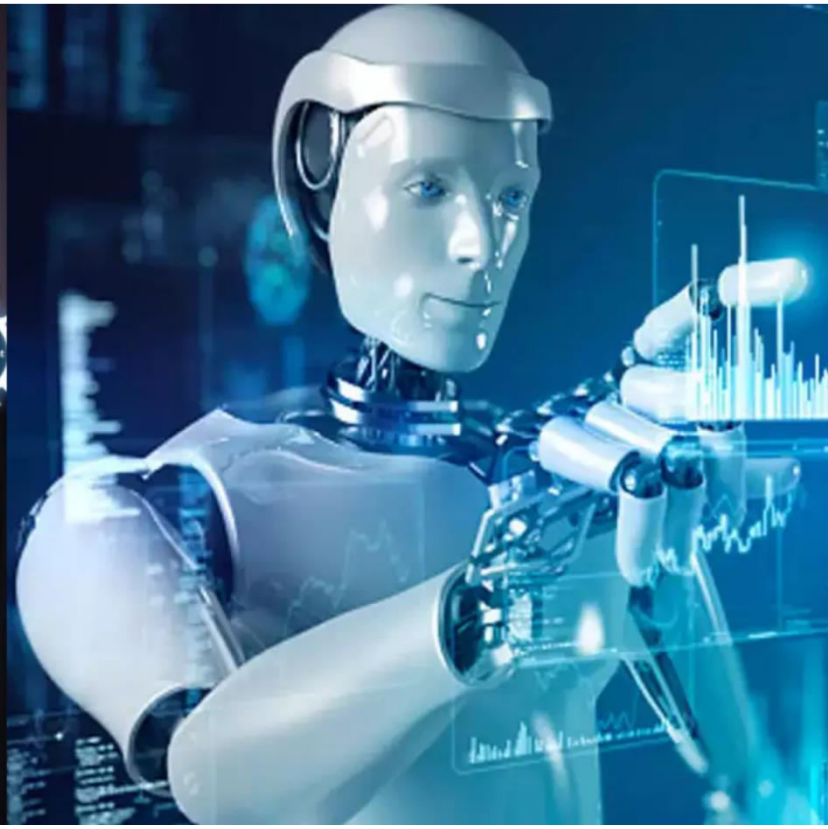
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

A Method for Detecting Intrusions that uses Decision Trees and Multi-Layer Perceptron Neural Networks

J.Archana, Dr.A.S. Aneetha

Research Scholar, Department of Computer science, Vels Institute of Science Technologies and Advanced Studies (VISTAS), Pallavaram, Chennai, India

Associate Professor, Department of Computer science, Vels Institute of Science Technologies and Advanced Studies (VISTAS), Pallavaram, Chennai, India

ABSTRACT: One of the biggest issues facing today's computer networks is the rise in online attacks. Therefore, it is essential for any computer network to establish security measures to stop such assaults. Intrusion Detection Systems (IDS) can more accurately identify assaults and system irregularities with the aid of machine learning and data mining techniques. However, the majority of the techniques that have been researched in this area, such as rule-based expert systems, are unable to effectively detect attacks that exhibit patterns that differ from those that are anticipated. Artificial Neural Networks (ANNs) can be used to classify the data and detect attacks, even in cases when the dataset is restricted, nonlinear, or incomplete. This study presents a technique that combines the Multi-Layer Perceptron (MLP) ANN with the Decision Tree (DT) algorithm. Artificial Neural Networks (ANNs) can be used to classify the data and detect attacks, even in cases when the dataset is restricted, nonlinear, or incomplete. This study proposes a technique that combines the Multi-Layer Perceptron (MLP) ANN with the Decision Tree (DT) algorithm to detect attacks with excellent accuracy and dependability.

KEYWORDS: Intrusion Detection Systems; Decision Tree; Neural Networks; Machine Learning

I. INTRODUCTION

Attackers and intruders can enter computer networks and cause major issues for users and network administrators by using anomalous traffic. For instance, an attack on a business firm could result in the loss of a large quantity of money and sensitive data. IDSs can identify attack traffic and report it to the firewall, even when they are unable to block it on their own. The reported traffic is then blocked by the firewall. Therefore, a strong firewall combined with a well-trained intrusion detection system can effectively safeguard the network. Since the IDS presented in this study is offline, it has undergone training prior to being installed on a network. An online IDS, on the other hand, is taught using online traffic data and operates in real time.

Multi-Layer Perceptron (MLP) Neural Networks [2] and Decision Trees (DT) [1] are two of the most well-known and widely used machine learning techniques for classification. DTs can identify anomalous traffic by creating a variety of intentional restrictions. By examining the dataset's features, this algorithm is able to create these rules. It is possible to diagnose the type of traffic once a DT has been trained. The tree can be subjected to network traffic. Certain rules that are generated in the tree's node would decide the record's trail. The sort of traffic will be discernible when the record reaches a leaf. An other effective method for categorisation is the use of artificial neural networks.Widely employed in machine learning, artificial neural networks (ANNs) are computational models modelled after the biological nervous system of the human brain. A typical artificial neural network (ANN) is portrayed as a network of interconnected "neurones" that can calculate values from inputs [3]. In this research, the dataset is classified using an MLP ANN. By establishing connections between attack entries, this network can assist the system in accurately identifying them.In many situations, ANNs can help the system function dependably even when there is ambiguity. The KDD CUP 99 [4] dataset from the Defence Advanced Research Projects Agency (DARPA) was used for the simulation. Along with one label that creates its 42 records, this dataset includes 41 network traffic features that have

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

been analysed [4].

KDD Cup 99 is one of the most well-known datasets in IDS-related research. For categorisation purposes, DT and MLP have been applied numerous times. The authors of [5] categorise attack and normal records using DT and NN-oriented methods. The Decision Support System (DSS) is used in their hybrid intrusion detection system (IDS) to detect misuse in KDD CUP 99. Support Vector Machines (SVM), DT, and Simulated Annealing (SA) are utilised by the authors in [6]. They employ SVM and SA for feature selection and SVM and DT for classification in their cumbersome approach. Their approach has a high rate of anomaly detection. The authors of [7] suggest a multi-level hybrid classifier that blends unsupervised Bayesian Clustering and supervised DT.Their approach has a high rate of anomaly detection. The authors of [7] suggest a multi-level hybrid classifier that blends unsupervised Bayesian Clustering and supervised DT. The performance of their approach on KDD CUP 99 demonstrates a low false alarm rate. To improve performance, these three recent publications combine DT with several different techniques. This shows that when DT is used in conjunction with another carefully selected algorithm, it performs fairly well. In [8], the typical training data is broken down into smaller subsets before a misuse detection model based on DT is constructed.A number of one-class SVM models are applied to the decomposed subsets. Their KDD CUP 99 results indicate a low false alarm rate, which is crucial for an intrusion detection system. The authors of [9] provide an ensemble classification technique that uses the radial basis function and MLP. The outcomes of their approach demonstrate better performance when compared to using these two algorithms exclusively. The K-mean Clustering technique is used by the authors in [10] to divide the dataset into a subspace. They then examine each space using a collection of MLPs.On the DARPA dataset, their model has revealed suitable performance. The authors of [11] employ SVMMLP, which is built on SVM and MLP. Compared to employing SVM alone, their technique has fewer space and time difficulties. The accuracy of their model on DARPA NSL-KDD is observable. To create a potent IDS, the authors of [12] combine the SVM, K-means, and MLP algorithms. When compared to other methods currently in use, their approach shows higher performance.

To improve their performance, the majority of studies combine DT and MLP with other clustering or classification techniques. These two approaches are particularly helpful due to their relative ease of usage. Other techniques, like RBF and SVM, demand greater timeliness. In this sense, it is easier to achieve outstanding outcomes with a hybrid IDS that incorporates the efficacy of both DT and MLP. The important thing is to use a straightforward technique that can be applied to actual networks. In contrast to this paper's approach, the majority of the aforementioned works utilise DT and MLP capabilities with a number of large and complicated tools that could be difficult to apply in the actual world.

Considering all assault kinds as a single primary attack class is one of the main goals of researchers [13–16]. This preprocessing procedure could have encouraging results in reducing false positive and false negative rates, which are essential for an effective IDS in offline detection, where the IDS has adequate time to develop its best potential model. The suggested approach makes use of every feature in KDD CUP 99. This approach makes use of DT and MLP's fruitful prediction capabilities. A well-trained MLP can function exceptionally well on its own. When a DT and this potent neural network are combined, an improved IDS will be produced. The suggested approach takes advantage of the advantages of these two algorithms in a well-constructed dataset.

II. THE SUGGESTED METHOD

The KDD CUP 99 dataset is enormous. Over 4 million records make up the original training dataset. It is therefore laborious to use the original dataset for the sinulation. For simulation purposes, researchers attempt to generate random train and test datasets in this regard. A dataset comprising 43,000 train records and 21,000 test records is chosen for this strategy. Each dataset contains 10,000 records, of which 10,000 are regular records and the remaining 10,000 are assault records. It should be noted that the KDD CUP 99 dataset has an excessive amount of redundancy. The majority of the records in the dataset are exactly the same since many traffic records were recorded by listening to traffic for a brief period of time. As a result, the redundant records are deleted before the random sampling phase begins. Normal and assault records were chosen at random from the irredundant dataset. Attack records were chosen from the original KDD CUP 99 dataset, which contains a variety of attack kinds. This kind of selection makes the model more accurate in real-world applications by simulating real-world settings. the distribution of records between our dataset and the original dataset.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

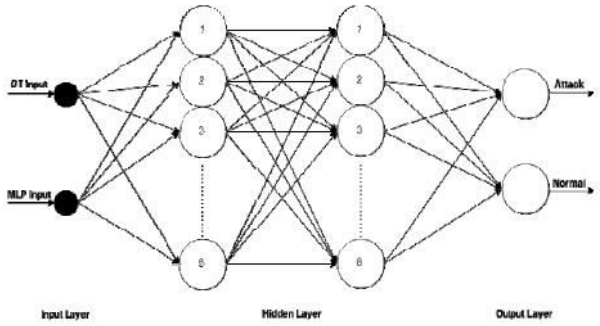
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

One of the most effective and often used classification techniques in IDS is MLP. A researcher can create a well-trained model to categorize desired classes by adjusting the MLP design. The suggested MLP network has been created with two layers, each containing eight neurons, taking into account the link between the number of layers and neurons, the speed, and the simulation's error rate. These neurons were chosen in order to adequately cover the volume of incoming data. Levenberg-Maquardt has been used as the training function for these simulations. Train records make for 43/21 of test records.



The suggested approach attempts to capitalize on the strengths of both algorithms in order to produce superior results, taking into account the performance of the DT and MLP algorithms in the trials conducted for this paper as well as the earlier research of other researchers. The suggested approach is created in two stages. The first step involves training both DT and MLP on a random dataset containing 41 features (referred to as the original dataset from now on) and obtaining the outputs (predicted values). These algorithms use their input records to forecast the type of traffic. They are able to create logical connections between input and output data and create a model that can precisely forecast fresh inputs. As a result, for every input traffic record, the algorithms forecast "Attack" or "Normal". It has two outputs to represent Attack and Normal recordings, eight neurons in each of its two hidden layers, and 41 inputs to cover the 41 features of the original dataset.

TABLE 1: NUMBER OF RECORDS IN EACH DATASET

	DARPA's Training dataset	DARPA's Test dataset	Our Random Training dataset	Our Random Test dataset
Attack	3,925,650	250,436	33,000	11,000
Normal	972,781	60,591	10,000	10,000
Total	4,898,431	311,027	43,000	21,000

The results of DT and MLP are combined with the dataset's original labels to form a new dataset in the second phase. The anticipated values of running DT and MLP on each and every record from the original dataset are thus two unique properties of the new dataset, which we will refer to as the new dataset from now on. Stated differently, each record's DT and MLP findings from phase one are recorded and added to the new dataset. "Normal" or "Attack" are the results of DT and MLP; we substitute the corresponding numbers of "1" or "2" for these values. The label that shows whether this item is "Normal" or "Attack" is then added to the new dataset. The MLP network is then trained once more using the fresh dataset, which consists of three columns (two features and one label). The outcomes are assessed after feeding the newly acquired dataset to the skilled MLP. It has two outputs to represent Attack and Normal recordings, eight neurons in each of its two hidden layers, and two inputs for the results of DT and MLP from the first phase.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

Diagram 3 represents the proposed method.

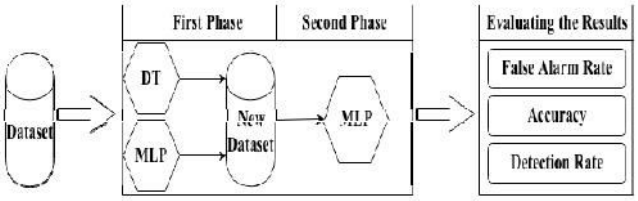


Diagram 3: Block diagram of the proposed method

The new dataset is the result of running DT and MLP on the previous dataset. The well-trained MLP is applied to the fresh dataset in the second phase. Ultimately, three distinct criteria can be used to evaluate the suggested IDS.

False Alarm Rate (FAR) is the metric used in this study to assess the outcomes. Detection Rate (DR) and Accuracy (A). These three factors are important instruments for assessing an IDS's effectiveness. The results of the suggested approach and its improved performance are demonstrated in the next section.

TABLE 2 : RESULTS OF SIMULATIONS ON ORIGINAL DATASET

	A (%)	FAR (%)	DR (%)
DT	97.93	2.19	98.01
MLP	99.71	0.12	99.89

III. RESULTS OF SIMULATION

Every IDS is evaluated using standard criteria to assess whether or not its performance is sufficient. The system may assess its performance in practical applications with the use of these examined evaluation criteria. A suitable IDS would favor lower FAR and greater A and DR. The following is an explanation of these terms:

$$A = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$DR = TP / (TP + FP) \quad (2)$$

$$FAR = FP / (FP + TN) \quad (3)$$

The number of Attack records that were accurately classified as Attack is represented by TP (True Positive), while the number of Normal records that were correctly labeled as Normal is represented by TN (TrueNegative).

FN (False Negative); the amount of Attack records that were mistakenly classed as Normal;
FP (False Positive); the number of Normal records that were mistakenly classified as Attack
The original dataset was used to test the DT and MI.P classification algorithms as a learning system in the suggested approach. And they get their outcomes. The outcomes of these two algorithms are shown in Table 2, which also contrasts their A, FAR, and DR.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

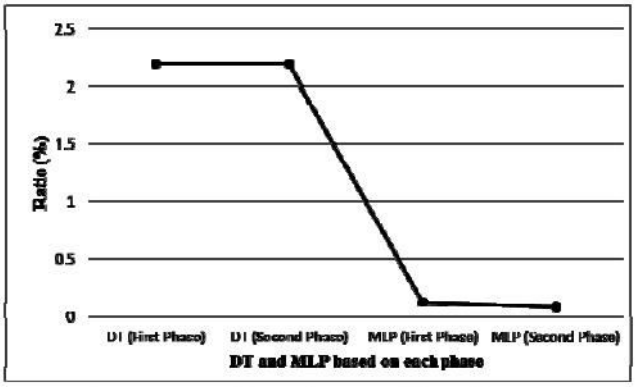
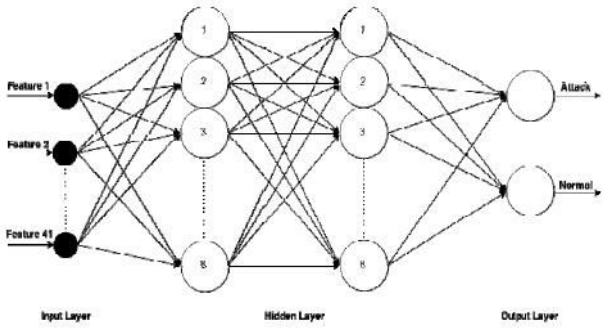


Figure 2. FAR of DT and MLP on different dataset

Both classification techniques function appropriately, as table 2 illustrates. Results from MI.P are better than those from DT. FAR is one of the most crucial factors in assessing an IDS since it shows how reliable it is in practical applications. MLP's efficacy and dependability are acknowledged by its extremely low FAR. FP has a direct impact on FAR. IDSs need to have low FP. Classifying suspected traffic that is actually typical as an attack may appear innocuous, but on a large scale, this characteristic can reduce an intrusion detection system's credibility.

The suggested approach improves the MLP's performance by recommending a new dataset. This indicates that the suggested approach makes use of both algorithms' capabilities to get better outcomes. Although DT outcomes are clearly not as good as MLP, they do assist MI.P perform better. The output dataset from the second phase includes important details on DT's outputs. With the use of this useful DT data, MLP is able to attain better outcomes in the second phase that were not possible in the first.



The effectiveness of DT and MLP on two distinct stages of the suggested approach. Each classification algorithm is preceded by its phase number. In both phases, the MLP performs better. Both datasets have the same DR of DT (98.01), however the output dataset's accuracy (97.94) is somewhat better by 0.01 percent than the original dataset. Although this may seem like a small improvement, it can help IDS correctly identify many records in situations when a large number of records are received in a short period of time. The MLP has the best DR and A (DR: 99.93 and A: 99.75), demonstrating that applying the MLP to the new dataset can result in an enhanced IDS.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

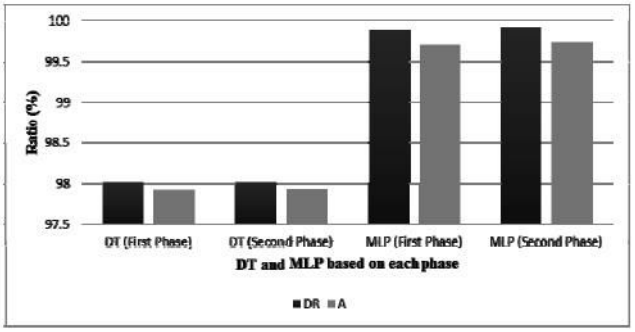


Figure 1. DR and A of two classification algorithms on different phases

FAR of MLP and DT on the new and original dataset. In the second phase, MLP performs better on the new dataset, but FAR in DT remains constant across datasets. In this scenario, MLP on the new dataset has the lowest FAR (0.08). As a result, the FAR, like DR and A, has the best overall outcome when using the suggested strategy. Using MLP with the new dataset could promise respectable performance in the actual world, given that FAR is one of the main criterion of IDS for system dependability.

These outcomes demonstrate the effectiveness of the suggested approach. The suggested approach improves the performance of both algorithms. The new dataset includes useful data regarding the performance of both algorithms. We can demonstrate the capabilities of both classifiers with better outcomes if we consider this dataset as a train dataset and evaluate its performance using classification techniques.

IV. FINAL RESULTS

This study suggests a technique for creating an accurate intrusion detection system with a high detection rate and a low false alarm rate that is based on artificial neural networks and decision trees. There are two distinct stages to this approach. By feeding the classification outcomes of the DT and MLP networks onto the random dataset, a new dataset is created in the first step. The data in the new dataset is once more classified using the MLP network in the second step, after which the outcomes are assessed. This hybrid approach has the potential to produce encouraging results. This technology can provide dependable outcomes in practical applications because to its extremely low FAR. Therefore, it is anticipated that future research will appropriately take into account the potential of such hybrid approaches.

V. ASSERTIVE

The KDD CUP 99 dataset was made publicly available by the UCI machine learning repository, for which the authors of this study are grateful.

REFERENCES

1. S. S. Savatha Sindhu, S. Geetha, and A. Kannan, "Decision tree based light weight strusion detection using a wrapper approach," Expert Systems with Applications, vol. 39, pp. 129. 141, 2012.
2. Chih-Fong Tsai and Chia-Ying Lin, "A triangle area based nearest neighbors approach to intrusion detection" in Pattern Recognition 43 (2010) 222-229
3. Simon O. Haykin "Neural Networks: A Comprehensive Foundation 3rd edition" New York Macmillan, 2009.
4. Y Yi, J. Wa, and W. Xu, "Incremental SVM based on reserved set for network intrusion detection," Expert Systems with Applications, vol. 38, pp. 7698.7707, 6/2011.
5. Ozgur Depren, Murat Topallar, Emin Anarım, M. Kemal Ciliz, "An intelligent intrusion detection system (IDS) for anomaly and mususe detection m computer networks" m Expert Systems with Applications Volume 29, Issue 4, November 2005, Pages 713-722

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404584|

6. Shah-Wei Lina, Kano-Ching Yingb, Chou-Yuan Leee, Zue-Jung Leed, "An imteligent algorithm with feature selection and decision rules applied to anomaly intrusion detection" im Applied Soft Computing Volume 12, Issue 10, October 2012, Pages 3285-3290
7. Cheng Xiang, Png Chan Yong, Lam Swee Meng. "Design of multiple level hybrid classifier for intrusion detection system. using Bayesian clustering and decision trees" in Pattern Recognition Letters Volume 29. Issue 7, 1 May 2008, Pages 918-924
8. Gusung Kina, Seumguin Leeb, Sehun Kima "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection" in Expert Systems with Applications Volume 41, Issue 4, Part 2, March 2014, Pagres 1690-1700
9. M Govindarajan and RM Chandrasekaranl, "Intrusion detection using neural based hybrid classification methods" in Computer Networks Volume 55, Issue 8, 1 June 2011, Pages 1662-1671
10. Hongyng Zheng, Lan Ni, Di Xiao "Intrusion Detection Based on MLP Neural Networks and K-Means Algorithm" in Springer-Advances in Neural Networks ISNN 2005 Lecture Notes m Computer Science Volume 3498, 2005, pp 434-438
11. Yong Hou, Xue Feng Zheng "SVM Based MLP Neural Network Algorithm and Application in Intrusion Detection" in Springer Artificial Intelligence and Computational Intelligence Lecture Notes in Computer Science Volume 7004, 2011, pp 340-345
12. M Chandrashekhar, K Raghuveer "Amalgamation of K-means Chistering Algonthm with Standard MLP and SVM Based Neural Networks to Implement Network Intrusion Detection System" m Sprmger-Advanced Computing Networking and Informatics Ve Volume 2 Smart hnovation Systems and Technologies Volume 28, 2014, pp 273-283
13. M. J. Fadaeieslam, B. Minaei-Bidgoli. M. Fathy, and M. Soryaau, "Comparison of Two Feature Selection Methods in Intrusion Detection Systems," in Computer and Information. Technology, 2007. CIT 2007 7th IEEE International Conference on, 2007, pp. 83-86.
14. S. Zaman and F. Karray, "Lightweight IDS Based on Festures Selection and IDS Classification Scherne," in Computational Science and Engineering, 2009. CSE 09. International Conference on, 2009, pp. 365-370,
15. L. Sang Min, K. Dong Seong, Y YoungHyun, and P. Jong Sou, "Quantitative Intrusion Intensity Assessment Using Important Feature Selection and Proximity Metrics, in Dependable Computing, 2009. PRDC 09. 15th IEEE Pacific Rim International Symposium on, 2009, pp. 127-134
16. PA Raj Kumar and S. Selvakumar, "Distibuted denial of service attack detection using an ensemble of neural classifier," Computer Communications, vol. 34, pp. 1328-1341, 7/15/2011.
17. F. Anuri M. Rezaei Yousefi, C. Lucas, A. Shakery, and N Yazılan, "Mutual mformabon-based feature selection for intrusion detection systems," Joumal of Network and Computer Applications, vol. 34, pp. 1184-1199, 7/2011
18. P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, "Practical real-time intrusion detection using machine learung approaches, Computer Communications, vol. 34, pp. 2227 2235, 12/1/2011
19. C. R. Pereira, R. Y. M. Nakamura, K. A. P. Costa, and J. P. Papa, "An Optumum-Path Forest framework for intrusion detection in computer networks," Engmeering Applications of Artificial Intelligence, val 25, pp. 1226-1234, 9/2012
20. A N Toosi and M Kahan, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers," Computer Communications, vol. 30, pp 2201-2212, 7/31/2007
21. ST. Sarasamma, Q. A. Zhu, and J. Huff, "Hierarchical Kohonenen net for anomaly detection in network security," Systems, Man, and Cybernetics, Part B: Cybermetics, IEEE Transactions on, vol 35, pp. 302-312, 2005
22. Clayton R Pereira, Rodrigoy M. Nakamura, KeltonA P Costa P Papa, "An Optimun-Path Forest framework for intrusion detection in computer networks in Engineering Applications of Artificial Intelligence Voluane 25, Issue 6, September 2012, Pages 1276-1234
23. Carlos A. Catania, Carlos Garcia Garino, Automatic network intrusion detection: Current techniques and open issues "Computers and Electrical Engineering (2012) "http// archive suci edu"

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details