
Magudeeswaran PREMKUMAR*, Tharai Vinay Param SUNDARARAJAN, Gopalakrishnan MOHANBABU

Abstract: Security in wireless frameworks is a significant and difficult task because of the open environment. The Denial of Service (DoS) is as yet significant endeavour to make an online assistance inaccessible. The objective of this attack is to keep the authentic nodes from getting to the administrations. Intrusion detection systems assume an essential job in identifying DoS attacks that improve the performance of the system. However massive information from the system presents huge difficulties to the discovery of DoS attack, as the identification framework needs adaptable techniques for gathering, storing and processing a lot of information. In order to defeat these difficulties, this paper proposes Hybrid Intrusion Detection System (HIDS) framework dependent on different MLP strategies. In this article HIDS utilizes Naive Bayes (NB), irregular random forest (RF), decision tree (DT), multilayer perceptron (MLP), K-nearest neighbours (K-NN) and support vector machine (SVM) for better outcomes. The NSL-KDD dataset and UNSW-NB15 dataset are taken to examine the detection accuracy. The experiment results show that the proposed defence system is accomplished with high accuracy, high detection rate and low false alarm rate in both the datasets.

Keywords: denial-of-services; DoS defense; hybrid mechanism; intrusion detection; machine learning; wireless sensor networks

1 INTRODUCTION

The wireless sensor networks (WSNs) are comprised of several tiny sensors networked with low- yield wireless environments. Wireless sensor networks are used in many devices for healthcare tracking, habitat monitoring, military applications, battlefield monitoring, smart grid and so on. These networks are endowed with sensing, data pre processing and communication modules. In this thousands of sensors were composed with a network Id which is rapidly deployed to collect the critical information from hostile and unattended circumstances. A node is comprised of four components: one sensing unit, one transceiver, one processor and one battery. The first component is comprised of optical converters and sensors. The phenomenon detected by sensing system is transformed by A to D converters into full digital signals. The processing circuit is followed by a small storage unit and the sensor node communicates with the other nodes. The transceiver system acts as a connection between the node and the network. Li, Ni MH batteries or power-saving devices such as solar powered systems use an electrical unit. The main constraint of the network is the node having limited energy resources, minimal processing memory, safety risks, and the least communication and processing capacity. In a device, DOS attacks are propelled by remotely controlled, powerful, and narrowly disseminated botnet PCs of zombies. Many of the traffics or administration demands are at the same time or persistently sent to the objective framework. The objective framework gets unusable, reacts gradually, or crashes totally because of the attack [7].

The distinguishing proof of the first aggressors is hard for the protection strategies on the grounds that the aggressors have caricature IP addresses and secured inside zombies with the intention of heavily influenced by them [5].

The DoS attack detection based on the KNN classifier is shown in Fig. 1. The KNN algorithm is based on the iterative relocation of a dataset separated into k clusters. The average square distance between cluster centres and data points is reduced.

Before the last attacks are dispatched, the DoS attacks are manual and must perform a few steps, including identification of traded off devices to create zombies on the internet, port filtering, and sending malware. Parameters unique to the target can also be organised by the perpetrators, while the others can be tracked using mechanised instruments [12]. The FC resembles the attacks as far as numerous clients access a framework simultaneously. In FC there is an unusual and abrupt ascent in real rush hour gridlock as a result of extraordinary occasions. It is hard for guarded frameworks to recognize FC unusual traffic from DoS attacks. Some entrenched audit and reviews on DoS attacks and safeguard strategies are accessible in the writing, including [12]. The survey researches the guard techniques that are sent for distinguishing, alleviating, as well as forestalling DoS attacks. It orders DoS safeguard strategies as indicated by the class of defencelessness, the level of mechanization, effect, and elements. In addition, this survey incorporates typical testing sets and assessment strategies. The aim of this survey is to expand the scope and shape the course of DoS study. It leads to some open research issues and gives a few thoughts for future research.

2 RELATED WORKS

In the last few years, several detection and mitigation mechanisms have been recorded for DoS attacks aimed at service providers [16]. For DoS attack detection, many methods have been suggested so far. Some of the latest work in DoS attack detection is summarized in this section.

Song and Liu [10] are proposing an authentic detection method that employs dynamic K-Nearest Neighbours (K-
NN) algorithm. In their proposed approach, a Storm delivery system is used which considers three aspects: feedback in real-time, the sophistication of processing bulk information, and analytical method.

Hameed and Ali [3] suggest a counter base detection method for DoS attacks that performed better while targeting DoS flood attacks using Hadoop technology. However, their research is unable to achieve effective delays in processing.

The Singular Value Decomposition (SVD) based detection mechanism is proposed [11] to construct the multi classifiers. The SVD is helpful to determine system reliability when extended to realtime scenario. The reports of SVD indicate improvement in performance when compared with the previous algorithms.

Bovenzi et al. [20] are proposing a DoS attack mitigation system in a Hadoop cluster using a Genetic Algorithm (GA). This DoS detection incorporates different patterns to overcome the weakness in a static pattern matching approach. Their research shows that the number of nodes is proportional to GA conclusion time. Their outcome would be more valuable if they had tested their algorithm's accuracy.

"Rehman et al propose an IDS model for recurrent neural networks using a deep learning approach. To construct the model, they use the LSTM architecture and their evaluation results using the performance metrics demonstrate that the deep learning approach is successful for IDS", [17].

The DNN is the well-known Network Intrusion Detection (NIDS) deep learning models that are tested by Green et al [16]. Both LSTM and Autoencoder are the basis of the NIDS and their reviews rely on the algorithm's consistency and precision. It claims that the coder can attain 98.9 percent accuracy, while only 79.2 percent accuracy is attained by the LSTM model.

"Deep learning is computationally more expensive and is expected to outperform conventional learning algorithms", [2]. It was proposed in [5] that computation is one way of minimising latency in DNNs. Conversely, it is also a daunting job to assess maximum parallelism in DNNs.

The proposed article incorporates advantages of the fuzzy system to construct classification models in a distributed environment, while providing reliable and efficient isolation of DoS attacks. Based on the selected machine learning algorithm the detection rate ability is limited and proposed method dynamically select the method using fuzzy.

3 HYBRID IDS FOR DOS DETECTION

The proposed framework is shown in Fig. 2. The framework consists of distributed system, different classification algorithms and fuzzy logic. To distinguish traffic packets, classification algorithms are used to locate DoS attacks from regular traffic. To facilitate the execution of these classification algorithms, we implemented a distributed framework using spark. The fuzzy logic approach is used to provide the intended classification algorithm to manage the process with a dynamic collection. Defense framework technique begins at the client side and data are compared with database after some advancement. In this event that if the data id not attacked, it continue to the agent side. Finally, if data is interpreted as natural, it goes to the server.

![Figure 2 Hybrid IDS framework for DoS detection](image)

According to AI algorithms, the detection of DoS attacks involves Artificial Neural Network (ANN) [1], Fuzzy Sets (FS) [2] and Genetic Algorithm (GA) [3]. The ANN uses the unsupervised learning to detect the attacks more effectively. In case of huge and complex networks, the above methods will not be scaled easily. The traditional machine learning (ML) approaches are employed to detect the attacks such as Decision Tree (DT) [3], Support Vector Machine (SVM) [9], Logistic Regression [2], Naive Bayes (NB) [3] etc.

Computational complexity, a qualitative metric, can indicate the difficulties involved with the development and overhead of a novel system. Its utility comes from its usage in comparing different techniques. According to [12], a type of network should be utilized that has features such as research functionality, extensibility, fidelity, repeatability, and programmability.

3.1 Naive Bayes Algorithm

It is extensively used in numerous classification applications [3]. Naive Bayes provides better accuracy compared to other algorithms. But it requires some features as it relies on the principle of independence of attributes. The posterior probability $P(h \mid D)$ of the classification is given by the Eq. (1).

$$P(h \mid D) = \frac{P(D \mid h) P(h)}{P(D)}$$ (1)
| where \( P(h \mid D) \) represents predictor probability given class, \( P(h) \) is prior class probability, and \( P(D) \) is the predictor's prior probability. In our data set, binomial model is used to segregate the attacker and normal node. For this the Bernoulli distribution is selected from other distributions like Gaussian, Multinomial, etc.

### 3.2 Random Forest Algorithm (RF)

"The RF algorithm is specifically proposed to have high prediction accuracy by integrating a large number of decision trees", [7]. The training processes are re-sampling, feature selection, tree bagging and DT growing. The RF algorithm is an ensemble of \( K \) trees \( T_1(X), T_2(X), \ldots, T_K(X) \), where \( X = X_1, X_2, \ldots, X_1 \) is 1 dimension input vectors. The resulting ensemble produces \( K \) outputs \( \psi_1 = T_1(X), \psi_2 = T_2(X), \ldots, \psi_K = T_K(X) \). The training \( \psi \) can be obtained from regression trees by considering the mean prediction results will be denoted as

\[
\Phi = \frac{1}{T} \sum_{\alpha=1}^{T} f_\alpha(x')
\]  

where \( T \) represents tree bagging time and \( f_\alpha \) denotes regression tree.

### 3.3 Decision Tree (DT)

Based on supervised learning, the decision tree builds the model in order to estimate the output from the previously learned attribute. In certain implementations, the decision tree is commonly used [8]. "The characteristics of DT in the training dataset affect the predictive accuracy, as the number of characteristics increases and the predictive accuracy decreases", [4]. Two variants are used to construct to determine the correct split in each stage. The Entropy in the data set \( X \) is given by

\[
H(X) = \sum_{x=1}^{\zeta} - f_x \log(f_x)
\]  

The Gini impurity is fluctuation entropy the calculation from previous to after port \( X \) is cleaved on an aspect.

\[
G = \sum_{x=1}^{\zeta} f_x(1 - f_x)
\]  

where \( f_x \) denotes \( x^{th} \) label frequency at a given node and \( \zeta \) is unique labels.

### 3.4 Multilayer Perceptrons (MLP)

The MLP is a dynamic attack detection using the concept of feature selection and feedback. In this, original features initially determine the number of input neurons that will be modified during the iteration process of feature selection. The logistic sigmoid function is defined as

\[
P(\alpha) = \frac{1}{1 + \exp(-\alpha)}
\]  

The model's output layer has a single neuron. The outcome of the MLP is according to conditional probability of attack and normal nodes. Given a training data set consisting of input vectors along with the related target vectors. The cross-entropy function is

\[
H = \frac{1}{N} \sum_{n=1}^{N} \left[ r_n \ln y_n + (1 - r_n) \ln(1 - y_n) \right]
\]  

### 3.5 SVM Classification

In different applications a non-linear classifier gives better accuracy. The essential standard of SVM is finding the ideal straight hyperplane in the segment space that maximally withdraws the two objective classes [9]. In this the SVM classifier is built by Gaussian kernel which is denoted as

\[
\Phi(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma} \right)
\]

\( \sigma \) signifies width of the capacity. The SVM classifier is developed by picking Gaussian kernel and its parameters. The hyper plane margin that is parameters are identified by the SVM training which is denoted as hyper parameters.

### 3.6 K-Nearest Neighbors

This algorithm is based on the iterative relocation of a dataset separated into \( k \) clusters. The average square distance between cluster centres and data points is reduced. The function is specified as

\[
D = \sum_{i=1}^{k} \sum_{x \in C_i} \sum_{m=1}^{l} \|W_{mm}(C_{mi} - x_m)\|^2
\]

Figure 4 Feature selection in KNN classifier
where \( X \) denotes the dataset, center of each cluster is represented as \( C_i \) and the \( W \) represents weight factor. The algorithm uses a minimal number of labelled data to restrict the collection of initial centre points in order to boost classification accuracy.

### 4 EXPERIMENTAL SETUP AND DISCUSSIONS

Presently two datasets, to be specific NSL-KDD and UNSW-NB15 are acquainted with identify the attacks in the system. Here the test settings are offered to approve the exhibition of the proposed hybrid IDS framework. Additionally the hybrid IDS is correlation made with some significant parameters.

#### 4.1 NSL-KDD Dataset

The NSL-KDD is one of the worldwide accepted databases for detecting the attacks in arrangements. "KDDTrain and KDDTrain 20% are the two sets for preparing and KDDTest+ and KDDTest-21 are testing sets which incorporate 41 highlights. Right now the system situations like typical, output to discover known vulnerabilities (Probing), DoS, U2R and R2L which is shown in Tab. 1", [3].

#### 4.2 UNSW-NB15 Dataset

The UNSW-NB15 using the IXIA Perfect Storm tool is proposed by Moustafa and Slay in Cyber Range Lab of the ACCS. It is entirely different from NSL-KDD that reflects the complex and huge threat to environment. This new data set has 49 features and it develops more categories of attacks. The some standard attack categories in the current scenario as shown in Tab. 2.

#### 4.3 Evaluation Metrics

The above hybrid technique evaluation is based on the classification of DOS attack from normal network. Based on correct or incorrect classification the represented as True positive (\( \text{True}_\text{pos} \)), True Negative (\( \text{True}_\text{neg} \)), False Positive (\( \text{False}_\text{pos} \)) and False negative (\( \text{False}_\text{neg} \)).

The accuracy defines how the algorithm identifies the DOS attacks from normal nodes in the dataset perfectly which is represented as

\[
\text{Accuracy} = \frac{\text{True}_\text{pos} + \text{True}_\text{neg}}{\text{True}_\text{pos} + \text{True}_\text{neg} + \text{False}_\text{pos} + \text{False}_\text{neg}}
\]

The detection rate is ratio of the algorithm identifies the DOS attacks from normal nodes in the dataset correctly which is represented as

\[
\text{Detection rate} = \frac{\text{True}_\text{pos}}{\text{True}_\text{pos} + \text{False}_\text{neg}}
\]

The False Alarm Rate (FAR) is defined as the ratio that the nodes incorrectly recognized as the attacks. It can be denoted as

\[
\text{FAR} = \frac{\text{False}_\text{pos}}{\text{True}_\text{neg} + \text{False}_\text{pos}}
\]

### 5 RESULTS AND DISCUSSION

The hybrid IDS is contrasted with other different ML methods, such as NB [3], LR [12], XGBoost [13], Gradient Boosting Decision Tree (GBDT) [3], K Nearest Neighbours [6], Multilayer Perceptrons [6] and Support Vector Machine (SVM) [9] with various kernel functions. The performance of the distinctive datasets and confusion matrix is given in Tabs. 3 and 4 respectively.

The presentation of the hybrid IDS and LR calculations shown in Fig. 3 that are superior to the next contrasting calculations of both KDDTestC and KDDTest-21.

The proposed method is presented as Algorithm 1 during the model training phase. The first cluster centres are calculated using a modest number of labelled data. Until the process converges, the equation is employed to determine the similarity between other unlabeled data and the original cluster centres.

The proposed system gets the accuracy of 99.963% for KDDTest+ dataset and 99.892% for KDDTest-21 individually. The experimental results of both datasets are given in Tab. 3.

The proposed detection system is evaluated in terms of detection rate, false alarm rate and accuracy of the detection. The detection parameters for varying classifiers are shown in Figs. 5 to 7. Average value of detection parameters the data sets are varied and tabulated. According to the result in Tab. 3, the detection rate of...
HIDS is 99.6%, which can be improved by increasing the number of training samples. The false alarm rate is 0.02867. The false alarm rate is high as the network is analyzed only with few attackers. As the number of attackers increases the false alarm rate reduces. Accuracy is 98%.

The proposed detection system is evaluated in terms of detection rate, false alarm rate and accuracy of the detection. The detection parameters for varying classifiers are shown in Figs. 5 to 7. Average value of detection parameters the data sets are varied and tabulated. According to the result in Tab. 3, the detection rate of HIDS is 99.6%, which can be improved by increasing the number of training samples. The false alarm rate is 0.02867. The false alarm rate is high as the network is analyzed only with few attackers; as the number of attackers increases the false alarm rate reduces. Accuracy is 98%.

| Table 3 Performance comparison with different datasets |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Classifiers     | Accuracy (%)    | False Alarm (%) | Detection Rate (%) |
| Hybrid IDS NB   | 99.963          | 0.049           | 99.6%            |
| GBDT            | 99.957          | 0.053           | 99.6%            |
| SVM             | 99.940          | 0.058           | 99.6%            |
| RF              | 99.933          | 0.062           | 99.6%            |
| MLP             | 99.920          | 0.063           | 99.6%            |

Discernibly, the HIDS performance is superior to other measurements in detection provided that all documents are known as regular in LR and results of the assessment are correct. The similarities between these two system groups could lead to dissatisfaction of LR with the perception of Normal/DoS (Tab. 3).

| Table 4 Confusion Matrix with different datasets |
|-----------------|-----------------|-----------------|-----------------|
| Type of data set | KDD TEST+       | KDD TEST-21     | UNSW-NB15       |
| Prediction       | Accuracy (%)    | False Alarm (%) | Detection Rate (%) |
| Hybrid IDS NB    | 99.963          | 0.049           | 99.6%            |
| GBDT             | 99.957          | 0.053           | 99.6%            |
| SVM              | 99.940          | 0.058           | 99.6%            |
| RF               | 99.933          | 0.062           | 99.6%            |
| MLP              | 99.920          | 0.063           | 99.6%            |

From the above difference, the HIDS accomplishes great outcomes in the recognition of DOS attacks, followed by a low FAR. Our proposed strategy is additionally contrasted with various recognition techniques with confirm the acquired outcomes. This section utilizes the NSL-KDD dataset and UNSW-NB15 on the grounds that it is the most broadly utilized correlation dataset. Right now, proposed strategy is contrasted and Hybrid-ADE [15], semi supervised learning method [14], Marliboost [2] and their results are summarized in Tab. 5.

| Table 5 Proposed Hybrid IDS Comparison with various DoS detection methods |
|-----------------|-----------------|-----------------|
| Approach        | Accuracy (%)    | FAR (%)         |
| Proposed Hybrid IDS | 99.963          | 0.049           |
| Hybrid-ADE      | 98.128          | 0.328           |
| Semi supervised learning | 97.399          | 0.449           |
| Marliboost      | 96.739          | 0.510           |

Our hybrid IDS technique achieves 99.963% accuracy with 0.049% FAR. Hybrid-ADE is used to minimise network traffic information loss and thereby achieve accuracy improvement which uses density calculation auto encoder's hidden layer. Semi-supervised learning method uses method of that eliminates irrelevant normal network traffic data to increase the accuracy of detection. Among this beginning state of art draws near, its FAR is the most noteworthy. Despite the fact that by utilizing an outfit technique, Marliboost accomplished the least FAR of 0.010%, it is restricted as far as exactness.

6 CONCLUSION

This article tries to offer an insight into DoS attacks detection in wireless environments based on Hybrid IDS. It consists of modules for capturing network traffic and detection. The collection module makes the use of micro-batch processing system to collect and extract traffic features. In detection module the malicious nodes are isolated based on Hybrid IDS Based classification algorithm. We performed experiments more on the location accuracy of the identifiable proof module to
measure the influence of the proposed hybrid IDS structure. The examination uses two customary datasets for benchmarks: NSL-KDD, UNSW-NB15. The test outcomes give extraordinary results with an accuracy of 99.963% and 98.751%, and the FAR of 0.049% and 1.08%, exclusively, as rather than various DoS attack acknowledgment methods. Because of the restrictions of the exploratory establishment, this article examined the HIDS structure in real time data. In future work, we plan to move on the appropriate state of the proposed Hybrid IDS, use DoS devices to initiate attacks, and then verify the performance of the system.

7 REFERENCES


Contact information:

Dr. Magudeeswaran PREMKUMAR, Assistant Professor  (Corresponding author)  
Department of ECE,  
SSM Institute of Engineering and Technology,  
Dindigul, Tamilnadu, India  
E-mail: prem3skumar@gmail.com

Dr. Tharai Vinay Param SUNDARARAJAN, Professor  
Department of ECE,  
Sri Shakti Institute of Engineering and Technology,  
Coimbatore, Tamilnadu, India  
E-mail: suryhp@yahoo.co.in

Dr. Gopalakrishnan MOHANBABU, Professor  
Department of ECE,  
SSM Institute of Engineering and Technology,  
Dindigul, Tamilnadu, India  
E-mail: shamyubabu@gmail.com