Abstract—This paper proposes a novel approach for intrusion detection and diagnosis. The proposed approach uses Sequential Backward Floating Search for feature selection and fuzzy ARTMAP for detection and diagnosis of attacks. The optimal vigilance parameter for the fuzzy ARTMAP is chosen using a genetic algorithm. The reduced set of features decreases the computation time by 0.789s. A classification rate of 100% and 99.89% is obtained for the detection stage and diagnosis stage, respectively.

I. INTRODUCTION

An increasing dependence of companies and government agencies on computer networks has increased the importance of protecting these systems from attacks. A single intrusion of a computer network can result in the loss or unauthorized utilization or modification of large amounts of data. Hence, it is important to have a detecting and monitoring system to protect important data. Intrusion Detection System (IDS) uses a lot of features. However, some of the features may be redundant or contribute little if anything to the detection process. The purpose of this study is to identify important input features in building an IDS that is computationally efficient and effective. Experimental results indicate that significant input feature selection is important to design an IDS that is lightweight, efficient and effective for real world detection systems[1]. Punch et al.[2], Pei et al. [3], Huang et al.[4] addressed the problem of feature selection by using a genetic algorithm (GA) to find an optimal or nearly optimal weighting of features for k-Nearest Neighbor classifiers. Huang et al. [4] apply a GA to find an optimal subset of features for a Bayes classifier and a linear regression classifier. Gartner et al. [5] used support vector machines to find optimal feature weights for a Bayes classifier. Stein et al.[6] combined a genetic algorithm with the decision tree classifiers to find optimal subset of features for decision tree classifiers. Mukkamala et al.[7] used data mining paradigms for feature selection and classification based on Bayesian networks and Classification and Regression Tress (CART). To the best of our knowledge, using sequential backward floating search has not been tried for intrusion detection. In this paper, we use a sequential backward floating search to find an optimal subset of features for fuzzy ARTMAP.

II. OVERVIEW OF INTRUSION DETECTION SYSTEMS

An intrusion is an unauthorized access or use of computer system resources. Intrusion detection systems are software that detect, identify and respond to unauthorized or abnormal activities on a target system. Intrusion detection techniques can be categorized into misuse detection and anomaly detection. Misuse detection uses patterns of well-known attacks or vulnerable spots in the system to identify intrusions. However, only known attacks that leave characteristic traces can be detected this way. Anomaly detection, on the other hand, attempts to determine whether deviations from the established normal usage patterns can be flagged as intrusions [7]. Although misuse detection can achieve a low false positive rate, minor variations of a known attack occasionally cannot be detected [7]. Anomaly detection can detect novel attacks, yet it suffers a higher false positive rate. Earlier studies on intrusion detection have utilized rule-based approaches to intrusion detection, but had a difficulty in detection new attacks or attacks that had not previously define patterns[8][9][10].

In the last decade, the emphasis has shifted to learning by example and data mining paradigm. Neural Networks have been extensively used to detect both misuse and anomalies pattern [11][12][13][14][15][16][17]. Recently, kernel-based methods such as support vector machine(SVM) and their variants are being used to detect intrusion [14][18][19]. One way in which fuzzy ARTMAP differs from many previously fuzzy pattern recognition algorithm is that it learns each input as it is received online, rather than performing an off-line optimization of a criterion function.

III. THEORETICAL BACKGROUND

A. Feature Selection

In complex classification domains, some features may be redundant since the information they add is contained in other features. Extra features can increase computation time, and can impact the accuracy of the IDS. Feature selection improves classification by searching for the subset of features, which best classify the training data. In this paper, we propose floating search to select the optimal or near optimal subset of features. Wrapper method which utilizes the prediction ability of some learning machine is used in this work to evaluate the feature subset. The wrapper method is defined as [20]

\[ W = R_{acc}(S_w) \]  

(1)

Where \( S_w \) is a feature subset candidate, and \( R_{acc} \) is calculated by 10-fold cross validation method. For training sample \( S \), \( R_{acc} \) is defined as:
\[ R_{\text{acc}} = \frac{1}{T} \sum_{j=1}^{T} (y_j = \tilde{y}_j) \]  

(2)

Where \( l \) is the number of training examples, \( y_j \) is the predicted value by fuzzy ARTMAP. The feature subset corresponding to the largest \( W \) is the most significant one. Sequential Backward Floating Search (SBFS) is based on sequential backward search (SBS) and sequential forward search (SFS), where SBS eliminates the least important features one by one, while SFS adds the most important features one by one. The least important feature \( f_x \) means that for a subset \( S_w \), there exists \( f_x \) such that \( R_{\text{acc}}(S_w - f_x) > R_{\text{acc}}(S_w - f_x f_x S_w) \), while the most important feature \( f_x \) means that for a subset \( S_w \) which satisfies \( S_w S_u = S \) and \( S_w S_u = 0 \), there exists \( f_x \) such that \( R_{\text{acc}}(S_w + f_x) > R_{\text{acc}}(S_w + f_x f_x S_w) \), where \( S_w = f_x \) means \( f_x \) is removed or added to \( S_w \). Fuzzy ARTMAP is used as the criterion function to find the most important features from \( S \) is a function used to find the most important feature for \( S_u \) from \( S \). The sequential backward floating search method proceeds as follows [21]:

**Step 1: Initialization**, use SBS method to remove the least important feature and set the value for the current number and the target number of features. Since we want to eliminate the features from maximum to \( l \), we set the target number to \( l \).

**Step 2: Exclusion**, use basic SBS method to remove the feature.

**Step 3: Conditional inclusion**, find the most significant feature, among the excluded features, if it is not the feature just eliminated, the basic SFS is used to add the feature to the current subset.

**Step 4: Continuation of conditional exclusion**, continue to find the most significant feature among the excluded features with respect to the subset obtained in Step 2, if \( R_{\text{acc}} \) of the subset is not higher than that of any subset with the same number of feature, then go to Step 2; otherwise go to Step 3.

**Step 5: Display**, if the number of features in the current subset is greater than the target number, go to Step 1; otherwise, display the optimal feature subset.

### B. Fuzzy ARTMAP

The fuzzy ARTMAP architecture is a self-organized learning system [22]. This kind of network has supervised training and pertains to the Adaptive Resonance Theory (ART) family: its structure is based on the adaptive resonance theory and is similar to the fuzzy ART network, which employs calculus based on fuzzy logic. This network is composed of two fuzzy ART modules: ART\(_a\) and ART\(_b\), interconnected by an inter-ART using an associative memory module as illustrated on Fig. 1 [22].

A fuzzy ARTMAP has an internal controller that ensures autonomous system operation in real time. The inter-ART module has a self-regulator mechanism named match tracking, whose objective is to maximize the generalization and minimize the network error. The \( F_a \) layer is connected to the inter-ART module by the weights \( w_{jk}^{ab} \). The steps of the fuzzy ARTMAP algorithm are summarized below.

1) **Input data**: The input patterns of the ART\(_a\) is represented by the vector \( a = [a_1 \ldots a_{Ma}] \) and the input patterns of the ART\(_b\) is represented by the vector \( b = [b_1 \ldots b_{Mb}] \).

2) **Parameters**: There are three fundamental parameters for the performance and learning of the fuzzy ART network [23].

- The chosen parameter, \( (\alpha > 0) \): acts on the category selection.
- Training rate, \( (\beta \epsilon [0, 1]) \): controls the velocity or the learning rate of the network adaptation. \( \beta = 1 \) permits the system to adapt faster while \( 0 < \beta < 1 \) allows the system to adapt slowly.
- Vigilance parameter, \( (\rho \epsilon [0, 1]) \): controls the network resonance. The vigilance parameter is responsible for the number of the categories formed. If the vigilance parameter is very large, it produces good classification, providing the generation of several categories and this means that the network has less errors but has less capacity of generalization. If it is very small, it will generate few categories and the network will have more capacity for generalization, but more possibility of making mistakes.

3) **Algorithm structure**: The ARTMAP network performs the processing of two ART networks, ART\(_a\) and ART\(_b\).

After the resonance is confirmed in each network, \( J \) is the active category for the ART\(_a\) network and \( K \) is the active category for the ART\(_b\) network. The next step is to verify using the match tracking, if the active category on ART\(_a\) corresponds to the desired output vector presented to ART\(_b\). The vigilance criterion is given by:

\[ \frac{|y^b \land w^{ab}_{jk}|}{|y^b|} = \rho_{ab} \]

(3)

It works by causing a little increment on the vigilance parameter only enough to exclude that category and select another category which will be active and reintroduced on the process until the active category correspond to the desired output.

4) **Learning**: After the input has completed the resonance state resonance by the vigilance criterion, the weight adaptation is implemented. The adaptation of the ART\(_a\) and ART\(_b\) module weights is given by:

\[ w^{new}_{jk} = \beta |I \land w^{old}_{jk}| + (1 - \beta)w^{old}_{jk} \]

(4)

The adaptation for the inter-ART module is performed as follows [23]:

\[ w^{ab}_{jk} = 1 \quad w^{ab}_{jk} = 0 \quad \text{for} \quad k \neq K \]

### C. Genetic Algorithm

GA solves problems by applying the principles of evolutionary biology, such as crossover, mutation, reproduction and natural selection [24]. The GA search process consists
of the following steps: 1) Generation of a population (pool) of candidate 2) Evaluate the candidate in the population, candidate with the lowest fitness are discarded and make way for a new set of chromosomes. 3) Replacement sets of chromosomes are created by the genetic operations of crossover and mutation on the fit individuals. 4) Steps 1 to 3 are repeated for a given number of generations until a specified fitness level is attained or a maximum number of generations is exceeded [25]. The genetic algorithms represent input data from the problem by an encoding and use the genetic operations to iteratively evaluate solutions from the population of potential solutions to determine the global optimum [24]. The GA evaluates candidate solutions through a fitness function by maximizing this fitness function. The fitness function contains information from the problem space and is the mechanism by which properties of the problem space is transferred to the GA, which is dependent of the problem. The genetic operations are important since they add an element of randomness to the search process, allowing a wider range of the solution space to be explored.

IV. PROPOSED INTRUSION DETECTION SYSTEM

The proposed intrusion detection system has four main phases which are data preprocessing and feature selection, selection of an optimal vigilance parameter using a GA, classification of attacks and testing of the IDS. Fig. 2 shows the block diagram for the intrusion detection system. Data preprocessing is done to make preprocessing to make it easier for the network to learn. Optimal set of feature were selected using sequential backward floating search. Fuzzy ARTMAP requires the selection of the vigilance and the selection of this parameter is important in classification system. A method for automatic selection of an optimal vigilance parameter using a GA is presented. The process of selecting an optimal vigilance parameter is shown in Fig. 3.

The following fitness function was utilized:

$$E = \frac{CI}{CT}$$

(5)
Where CI represent example that were correctly classified and CT is the total number of examples.

In this paper, a floating point representation is used as it allows the search to be faster [24]. An arithmetic crossover was proven to provide better stability in generating solutions [24]. A non-uniform mutation is used as it aids in the search of an optimal solution by allowing for faster convergence and greater accuracy [24]. Sequential Backward Floating Search is used to select the optimal or near optimal set of features. After a good set of feature are selected these features are used to train a fuzzy ARTMAP. The classification stage is divided into stages. The first stage detects if there is an intrusion or not; while the second stage determines the nature of the attack. Fig. 4 shows the block diagram of the classification phase.

![Block diagram of the classification phase](image)

V. EXPERIMENTATION

A. Intrusion Detection Database

The investigation in this paper is entirely based on the data obtained in the experimentation done by [26]. The data contains 24 attack types that could be used be classify into four main classes, which are probing, denial of service (DoS), user to root(U2R) and root to user (R2U). The data contains 5 symbolic features and 38 numerical features. The symbolic features were converted into binary using ASCII. The data is normalized using linear normalization. For the detection stage, all the attacks are put as one class, this is to determine if there is an intrusion or not. In the diagnosis stage, the attacks are classified into four classes.

1) Probing: Probing is a class of attacks where an attacker scans a network to gather information or find vulnerabilities [27]. An attacker with a map of machines and services that are available on a network can use the information to look for exploits. There are different types of probes some of them are abuse the computer’s legitimate features and social engineering.

2) Denial of service: DoS is a class of attacks where an attacker makes some computing or memory resource too busy or too full to handle legitimate request, thus denying legitimate users access to a machine [27]. There are different ways to launch DoS attacks through abusing the computers’ legitimate features by targeting the implementations bugs and exploiting the system’s misconfiguration. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users.

3) User to root: User to root exploits are a class of attacks where an attacker starts out with an access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system [27]. Most of the common exploit in this class of attacks are regular buffer overflow, which are caused by regular programming mistakes and environmental assumptions.

4) Remote to User: A remote to user (R2U) attack is a class of attacks where an attacker sends packets to a machine over a network, then exploits machine’s vulnerability to illegally gain local access as a user. There are different types of R2U attacks; the most common attack is done using social engineering [27].

VI. RESULTS AND DISCUSSION

We first demonstrate how different vigilance parameter affect the classification error. Fig. 5 shows the error rate versus the number of training cycles for various vigilance parameters and the optimal vigilance parameter. As it can be seen from the graph the optimal vigilance converges to zero error after five training cycles. Sequential Backward Floating Search is used to select the optimal set of features. The feature selection process takes 140s seconds on average with a pentium IV with processing speed of 3.2 GHz. Fig. 6 shows criterion function as the number of features are increased. It can be seen that the criterion function increase to one, which is the maximum value when the number of features is twelve. Hence, 12 feature were selected to be the optimal number.
Fig. 5. Error rate for various vigilance parameter and the optimal vigilance selected using a genetic algorithm

Fig. 6. The criterion function for different number of features

Table I shows the classification rate for 12 selected features and 41 features. The duration of the training process of the fuzzy ARTMAP with 12 feature is approximately 0.998 seconds with a Pentium IV having a processor speed of 3.2GHz. Using the same machine, the training took 1.789s for 41 features. It can be seen that the reduced feature give similar accuracy to the 41 features in a shorter time. This is important for real-time intrusion detection. The overall classification rate of the system is 99.89% Table II shows the detection rate and the false alarm rate for the diagnosis stage for the four different classes. The DoS and probe attacks gave a detection rate of 100% and false alarm rate of 0% while the U2R and R2U gave detection rate of 99.87% and 99.78%, respectively.

Table I

<table>
<thead>
<tr>
<th>Class</th>
<th>41 Features</th>
<th>Optimal Set of Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>DoS</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>R2U</td>
<td>99.87</td>
<td>99.78</td>
</tr>
<tr>
<td>U2R</td>
<td>99.70</td>
<td>99.70</td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>Class</th>
<th>DR (%)</th>
<th>FR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>DoS</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>R2U</td>
<td>99.87</td>
<td>1.4</td>
</tr>
<tr>
<td>U2R</td>
<td>99.70</td>
<td>1.5</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS

A procedure for detection and diagnosis of network attacks using Fuzzy ARTMAP has been presented. The optimal subset of features were selected using Sequential Backward Floating Search. The optimal vigilance parameter for the Fuzzy ARTMAP was chosen using GA. The Fuzzy ARTMAP is used to detect and classify the attacks. It is found that reducing the features from 41 to 12 reduces the computation time by 0.789s. The optimal set of features and Fuzzy ARTMAP give a classification rate of 100% for the detection phase and 99.89% for the diagnosis. Future work will look at the implementation of on-line training for real-time intrusion detection system.

REFERENCES


