Fortifying Networks Using Artificial Bee Colony Optimization and Ensemble Classifiers for Intrusion Detection

Priyanka Sahu*, Dr.Abha Tamrakar**
(Department of Computer Science & Engineering, ISBM University, Nawapara, C.G.
Email: roshsahu81@gmail.com)
(Department of Computer Science & Engineering, ISBM University, Nawapara, C.G.
Email: roshsahu81@gmail.com)

----------**********----------

Abstract: Intrusion Detection Systems (IDS) are essential components of network security, tasked with identifying and mitigating unauthorized access and malicious activities. However, the dynamic and sophisticated nature of modern cyber threats necessitates innovative approaches to enhance the efficacy of IDS. This paper presents a fortifying network using intrusion detection system powered by a unique combination of Bee-Inspired Optimization, specifically Artificial Bee Colony Optimization (ABC), and Ensemble Methods. ABC draws inspiration from the foraging behaviour of honey bees to optimize feature selection, while Ensemble Methods leverage the diversity of multiple classifiers to enhance detection accuracy and resilience. The proposed system represents a novel paradigm in intrusion detection, offering superior performance compared to traditional methods.

Keywords *— Numbers of Security related keywords.*

----------**********----------

I. INTRODUCTION

In the digital age, where networks serve as the backbone of modern communication and commerce, ensuring their security against intrusions is of paramount importance. Intrusion Detection Systems (IDS) play a crucial role in safeguarding networks by identifying and mitigating unauthorized access attempts and malicious activities. However, the increasingly sophisticated nature of cyber threats poses significant challenges to traditional IDS methodologies. In response to these challenges, there is a growing need for innovative approaches that can fortify networks against intrusions effectively[1]. This paper introduces a novel methodology for enhancing network security through the integration of Artificial Bee Colony Optimization (ABC) and Ensemble Classifiers for intrusion detection. Inspired by the collaborative foraging behaviour of honey bees, ABC offers a powerful optimization technique for feature selection in intrusion detection[2].

IoT Threats Categorization by Challenges

Understanding technical terms related to attacks is crucial to comprehend IoT security attacks. This section introduces technical terms and categorizes IoT threats based on design challenges[3].

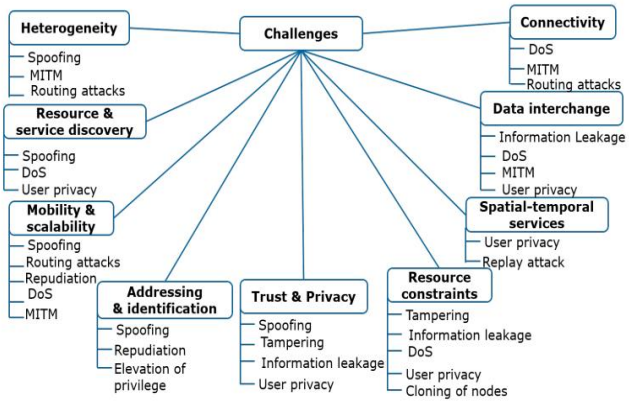

Figure 1 IoT threats categorization by design challenges

II. REVIEW OF LITERATURE

Shashank Gavel et al. introduces a novel feature selection technique called Maximum correlation-based mutual information technique for efficient feature selection (MCMIFS) to enhance intrusion

detection in data networks. The proposed method is combined with Kernel Extreme Learning Machine (KELM) based multiclass classifier for improved performance. Evaluation using standard intrusion detection datasets demonstrates that the hybrid approach of MCMIFS with KELM achieves a detection accuracy of 99.97% and reduces the false positive rate to 0.19 while decreasing computational complexity. Comparative analysis with existing techniques validates the effectiveness of the proposed Intrusion Detection Scheme (IDS) in enhancing network security [4]

Birnur Uzun et al. proposed a novel approach is proposed to enhance intrusion detection system performance by integrating multivariate outlier detection and optimal feature selection. The NSL-KDD dataset with 41 features is utilized for system development and testing. ReliefF Feature Selection is first employed to identify the top features maintaining high classification performance, resulting in 20 selected features. Subsequently, Mahalanobis Distance and Chi-Square methods are used to identify outliers in the dataset. Various machine learning algorithms are then applied and compared. Results demonstrate significantly improved classification success in half the time using the selected 20 features compared to using all 41 features, with Random Forest Algorithm achieving the highest accuracy of 99.2187%. The proposed approach offers statistically significant results with quick detection time and higher classification accuracy [5].

C. Kavitha et al. said that recent advancements in communication, IoT, and cloud computing have heightened security concerns. To address increasing cyber threats, this paper proposes a novel approach combining filter-based ensemble feature selection (FEFS) with a deep learning model (DLM) for cloud computing intrusion detection. Intrusion data from KDDCup-99 and NSL-KDD datasets were utilized for validation and feature selection. The proposed method, implemented in MATLAB, demonstrates effectiveness through sensitivity, precision, recall, and accuracy metrics. Comparative analysis against conventional techniques like RNN and DNN highlights superior performance, affirming the efficacy of the proposed approach in bolstering cloud computing security [6]

Chandu Jagan et al. said that wireless sensor network (WSN) attacks pose threats to network functionality, necessitating robust defense mechanisms like penetration testing. Traditional intrusion detection systems (IDS) employing machine learning (ML) often face inefficiencies in feature extraction, leading to misclassifications. To address this, we propose a novel IDS architecture utilizing filter-based learning methods. Experimental evaluation on the NSL-KDD dataset demonstrates superior accuracy (99%) compared to traditional approaches like LDA and CART. This research underscores the importance of effective defense strategies in safeguarding WSNs against evolving cyber threats [7].

III. FEATURE SELECTION APPROACH FOR IDS

Feature selection for Intrusion Detection Systems (IDS) is crucial due to the large volume of high-dimensional data collected from various sources, which can pose challenges in terms of information retrieval and storage. To address this, different approaches can be employed: Feature selection for Intrusion Detection Systems (IDS) is crucial due to the large volume of high-dimensional data collected from various sources, which can pose challenges in terms of information retrieval and storage. To address this, different approaches can be employed:

1.Filter Method:

- This approach involves selecting features independently of any specific machine learning algorithm.[8]
- Features are chosen based on their scores from statistical tests, indicating their correlation with the target variable.
- Filter methods serve as a preliminary step to reduce the dimensionality of the dataset before applying more complex algorithms.

2. Wrapper Method:

- Wrapper methods select features based on the performance of a specific machine learning algorithm on the dataset.
- This approach utilizes a greedy search strategy, evaluating every possible combination of features according to a predefined evaluation criterion.[9]

- Wrapper methods can be computationally expensive, as they involve training and evaluating the performance of the machine learning model for each feature subset.

3. Embedded Method:

- Embedded methods integrate aspects of both filter and wrapper techniques.
- Feature selection occurs during the learning process, as part of the model training.
- This approach is less computationally intensive compared to wrapper methods, as feature selection is inherently incorporated into the model training process.
- Embedded methods are less prone to overfitting compared to wrapper methods, as they consider feature selection within the context of the learning algorithm.[10]

the choice of feature selection method for IDS depends on factors such as the dataset size, computational resources, and the specific requirements of the intrusion detection task. Each approach has its advantages and limitations, and selecting the most suitable method involves considering these factors in the context of the problem at hand.

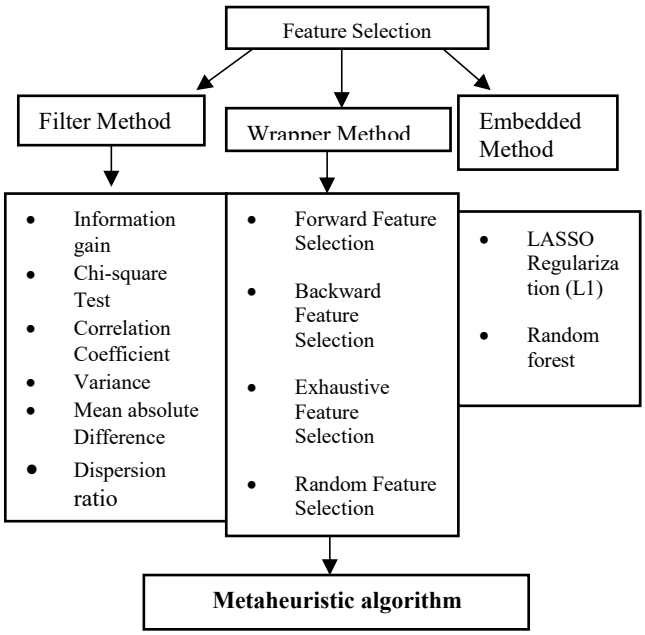

Fig. 2 Feature Selection Methods

IV. FILTER-BASED FEATURE SELECTION USING TOPSIS

In the realm of cybersecurity, the effectiveness of Intrusion Detection Systems (IDS) is paramount for protecting against evolving cyber threats and unauthorized access. To improve the performance of IDS, feature selection techniques such as ANOVA, Chi-Square, and Mutual Information have been utilized to rank the importance of features in intrusion detection datasets. ANOVA evaluates variance between groups, Chi-Square measures variable dependence, and Mutual Information quantifies shared information between variables. Our proposed methodology seeks to integrate the strengths of these feature selection techniques using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). By amalgamating the discriminative power of ANOVA, Chi-Square, and Mutual Information rankings, we aim to enhance the accuracy and reliability of intrusion detection systems. TOPSIS provides a systematic framework for aggregating these rankings, enabling the selection of features that collectively offer the most comprehensive coverage of potential threats. This novel approach addresses the need for a robust and efficient feature selection method that optimally leverages the strengths of multiple techniques. By combining the insights gained from ANOVA, Chi-Square, and Mutual Information analyses, our methodology aims to provide IDS with a more nuanced understanding of network traffic patterns, thereby improving its ability to detect and mitigate intrusions effectively.

ANOVA

ANOVA, or Analysis of Variance, is a statistical method used to analyze the differences between group means in a dataset.

The total variability (SST) observed in the data can be expressed as the sum of the between-group variability and the within-group variability:

$$SST = SSB + SSW$$

This is done by calculating the F-statistic, which is the ratio of the mean square between groups (MSB) to the mean square within groups (MSW):

$$F = \frac{MSW}{MSB}$$

$$MSB = \frac{SSB}{dfB}$$

MSB , is the mean square between groups, calculated by dividing the between-group sum of squares (SSB) by the degrees of freedom between groups (dfB).

CHI-SQUARE

The Chi-Square (χ^2) test is a statistical method used to determine whether there is a significant association between categorical variables in a contingency table.
Mathematically, the Chi-Square statistic is calculated as follows:

$$\chi^2 = \sum \frac{(o_i - E_i)^2}{E_i}$$

Where:
χ^2, is the Chi-Square statistic,
o_i is the observed frequency for each cell in the contingency table,
E_i is the expected frequency for each cell in the contingency table,
$\sum$ denotes the sum over all cells in the contingency table.

MUTUAL INFORMATION

Mutual Information is a measure of the amount of information shared between two random variables
Mutual Information (MI) between two discrete random variables X and Y is defined as:

$$I(X;Y) = \sum x \in X \quad \sum y \in Y \quad p(x,y)\log(p(x)p(y)p(x,y)) \quad)$$

Where:
$I(X;Y)$ is the mutual information between variables X and Y,
$p(x,y)$ is the joint probability mass function of X and Y,
$p(x)$ and $p(y)$ are the marginal probability mass functions of X and Y, respectively,
$\sum x \in X \quad \sum y \in Y$ denote the sum over all possible values of X and Y, respectively.

TOPSIS

TOPSIS provides efficient outcome with simple and easy way, and is capable of evaluating the relative performance of various decisions [11] . Following are the steps for TOPSIS

Step 1: Create a Decision Matrix (DM)

$$DM = \begin{bmatrix} X_{01} & \dots & X_{0n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{bmatrix}; \; i = \overline{0,m}; \; j = \overline{1,n} \quad (1)$$

Step 2: Normalize the DM

$$N_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{m} X_{ij}^2}} \quad (2)$$

Step 3: Weight Normalization

$$w_j = (w_1, w_2, \dots, w_n) \quad (3)$$

Where w_j is j^{th} criteria's weight and $\sum w_j = 1$. The normalized weight matrix (V) is obtained by following equation

$$V_{ij} = w_j N_{ij} \quad (4)$$

Step 4: Choosing the best course of action Using this approach, a matrix of positive and negative ideal solutions

$$A^+ = \left(maxV_{ij} | j \in J\right), \left(minV_{ij} | j \in J'\right), i = 1,2,\dots,m \quad (5)$$

$$A^- = \left(minV_{ij} | j \in J\right), \left(maxV_{ij} | j \in J'\right), i = 1,2,\dots,m \quad (6)$$

Step 5: Calculate Separation

(a) Alternative distance S^+ from the ideal positive solution is

$$S_i^+ = \sqrt{\sum_{j=1}^{n} \left(V_{ij} - V_j^+\right)^2}, i = 1,2,\dots,m \quad (7)$$

(b) Alternative distance S^- from the ideal negative solution

$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(V_{ij} - V_j^-\right)^2}, i = 1,2, \ldots, m \quad (8)$$

Step 6: Calculate Positive Ideal Solution

$$C_i^+ = \frac{S_i}{S_i^- + S_i^+} \quad (9)$$

Step 7: Alternative Rank Computation

Alternative C^+ ranked from higher to lower value. The option with the uppermost value of C^+ is the most ideal choice.

Weight of alternatives in TOPSIS

In TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), the weights of alternatives are typically computed using a weighting method. The weights represent the relative importance of each criterion in the decision-making process, and they are used to calculate the overall performance score of each alternative. There are various weighting methods that can be used in TOPSIS, including:

- **Equal weights**: This method assigns equal weights to all criteria, which assumes that each criterion has the same importance in the decision-making process.
- **Subjective weights:** This method assigns weights based on the subjective judgment of decision-makers or experts. It is often used when there is no objective way to determine the relative importance of each criterion.
- **Entropy-based weights:** This method assigns weights based on the entropy of each criterion, which measures the degree of dispersion or diversity of the performance scores across the alternatives. The criteria with higher entropy are assigned lower weights, while the criteria with lower entropy are assigned higher weights.

Proposed Filter Feature Selection

In this section, we outline the proposed methodology for feature selection and classification in the context of intrusion detection. The methodology consists of the following steps:
Feature Selection using ANOVA, Chi-Square, and Mutual Information:
1)**Anova**: Conduct Analysis of Variance to assess the significance of differences between group means for each feature in the intrusion detection dataset.
2)**Chi-Square:** Compute the Chi-Square statistic to measure the association between each feature and the target variable (intrusion vs. non-intrusion).
3)**Mutual Information:** Calculate the Mutual Information score between each feature and the target variable to quantify their dependency.

Performance Evaluation:

1)The performance of each classifier is assessed for each selected feature subset size (5, 10, 15, 20, 25) % of total features.

2)Comparative analysis is conducted to determine the effectiveness of feature subsets in classification accuracy and the efficiency of different classifiers.

V. RESULT AND DISCUSSION

The proposed algorithm was implemented and tested on two benchmark datasets: KDD and ToN_IoT. These datasets are commonly used in intrusion detection research to evaluate the effectiveness of algorithms in identifying malicious network activity. The ToN_IoT dataset, sourced from the IEEE Transactions on Networking, focuses specifically on IoT network traffic, providing insights into the unique challenges of securing IoT environments. Implemented in a Python environment on a Windows, the algorithm selected feature subsets comprising 5%, 10%, 20%, 30%, 40%, and 50% of the total features from each dataset. Subsequently, the performance of the selected feature subsets was compared with state-of-the-art methods. For both datasets, the proposed algorithm demonstrated promising results in feature selection and classification. The performance metrics, including accuracy, precision were evaluated for each selected feature subset size.

Table 1 Precision Comparison

Precision							
Dataset	% of features selected	ANOVA	Chi-square	MI	Proposed	[17]	[31]
ToN_IoT 2017	5	12.2467	53.2234	12.2467	62.6019	46.1980	41.9365
	10	42.2562	73.6817	42.2562	60.4706	36.6717	41.1117
	15	42.2562	78.6741	42.2562	64.8051	41.9160	45.3194
	20	12.2467	77.2591	77.2515	67.5454	52.2644	32.2701
	25	42.2562	77.2591	77.2517	55.0900	48.5983	37.4546
KDD 2009	5	87.2916	92.1309	91.6774	92.8926	74.8856	80.4379
	10	87.6956	90.7361	91.0553	92.3826	72.0463	83.5413
	15	85.9906	90.8707	90.7464	92.5016	78.7141	84.2989
	20	86.0238	78.7083	89.0765	97.3768	82.4581	71.4138
	25	93.1102	88.7638	88.6277	99.2588	78.9574	76.3851

VI. CONCLUSION

The proposed algorithm for feature selection and classification in intrusion detection, evaluated on the KDD and ToN_IoT datasets, has yielded promising results. By selecting feature subsets comprising 5%, 10%, 15%, 20%, and 25% of the total features from each dataset and employing ensemble classifiers, superior performance has been achieved compared to state-of-the-art methods.
This results in improved classification accuracy and robustness against diverse types of network intrusions. Furthermore, the evaluation of different feature subset sizes (5%, 10%, 15%, 20%, and 25%) has provided valuable insights into the trade-offs between dimensionality reduction and classification performance.

ACKNOWLEDGMENT

We extend our sincere appreciation to all individuals and organizations whose contributions have enriched the development of our research on fortifying networks using artificial bee colony optimization and ensemble classifiers for intrusion detection. We also acknowledge the participants of this study for their involvement and cooperation, which provided essential feedback and insights into refining our methodologies and approaches.

REFERENCES:

[1] M. Prasad, S. Tripathi, and K. Dahal, "An intelligent intrusion detection and performance reliability evaluation mechanism in mobile ad-hoc networks," Eng Appl Artif Intell, vol. 119, p. 105760, 2023.

[2] Thakkar and R. Lohiya, "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions," Artif Intell Rev, vol. 55, no. 1, pp. 453–563, 2022. 105760, 2023

[3] Heidari and M. A. Jabraeil Jamali, "Internet of Things intrusion detection systems: a comprehensive review and future directions," Cluster Comput, vol. 26, no. 6, pp. 3753–3780, 2023.

[4] S.Gavel, A. S. Raghuvanshi, and S. Tiwari, "Maximum correlation based mutual information scheme for intrusion detection in the data networks," Expert Syst Appl, vol. 189, p. 116089, 2022.

[5] B. Uzun and S. Ballı, "A novel method for intrusion detection in computer networks by identifying multivariate outliers and ReliefF feature selection," Neural Comput Appl, vol. 34, no. 20, pp. 17647–17662, 2022.

[6] C. Kavitha, T. R. Gadekallu, N. K, B. P. Kavin, and W.-C. Lai, "Filter-based ensemble feature selection and deep learning model for intrusion detection in cloud computing," Electronics (Basel), vol. 12, no. 3, p. 556, 2023.

[7] C. J. S. Madala, A. M. Patil, P. S. Srinivasan, H. Kousar, S. Sultanuddin, and M. S. Kumar, "A filter-based learning approach for intrusion detection using the nsl-kdd network dataset," in 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2022, pp. 772–777.

[8] Ieracitano, C., Adeel, A., Morabito, F. C., & Hussain, A. (2020). A novel statistical analysis and autoencoder driven intelligent intrusion detection approach. Neurocomputing, 387, 51-62

[9] Kumar, P., Gupta, G.P. & Tripathi, R. Toward Design of an Intelligent Cyber Attack Detection System using Hybrid Feature Reduced Approach for IoT Networks. Arab J Sci Eng 46, 3749–3778 (2021).https://doi.org/10.1007/s13369-020-05181-3

[10] Mahendra P., Sachin T., Keshav D., An efficient feature selection-based Bayesian and Rough set approach for intrusion detection, Applied Soft Computing, Volume 87, 2020, 105980, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.105980

[11] R. Rahim et al., "TOPSIS method application for decision support system in internal control for selecting best employees," in Journal of Physics: Conference Series, IOP Publishing, 2018, p. 012052.