Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 02 (2026)

# A New Framework for Detecting Anomalies in Network Traffic Using Supervised Learning Techniques

**Dr. Selvanayaki K., Mr. Panjatcharam V. G.**

Research Supervisior, Department of Computer Science and Applications, VET Institute of Arts and Science (Co-education) College, Erode.

, Research Scholar, Department of Computer Science, VET Institute of Arts and Science (Co-education) College, Erode.

**Abstract** The new framework for detecting anomalies in network traffic is detailed model for novel network security, specifically as cyber threats resume to grow in intricacy. This work investigates Anomaly detection in network traffic is a critical component of modern network security, specifically as cyber threats continue to grow in intricacy. This framework investigates the efficiency of numerous supervised learning methods for classifying anomalies in network data, with a identifiable focus on their facility to handle challenges such as class imbalance and high dimensional feature spaces. The estimated methods include Isolation Forest, Naïve Bayes, Light GBM and Support Vector Machine (SVM) classifiers. Across complete investigation, both supervised and unsupervised techniques are analysed and contrasted using key performing metrics such as accuracy, precision and recall.The proposed anomaly detection framework follow four major stages: Preprocessing, Enhancement, feature Extraction and Selection. New results establish that the framework attains higher exactness and robust overall accuracy. Using the sample dataset, This study highlights the comparative strengths of supervised and unsupervisied models and provides a reliable framework for effective network anomaly detection.

**Keywords**: Anolmaly Detection, Support Vector Machine(SVM), Preprocessing, Enhancement, Feature Extraction, Feature Selection, Swarm Intelligence, Ant Colony Optimization (ACO), Particle Swarm Optimization(PSO). Intrusion Detection System (IDS).

## I. Introduction

Network anomaly detection involves continuously monitoring network traffic to identify unusual or suspicious patterns that deviate from established normal behaviour, which may signal security threats such as cyber-attacks or system malfunctions. By defining a baseline of typical activity, these systems can detect deviations such as unexpected spikes in traffic and enable security teams to respond quickly and mitigate potential risks. Detecting anomalies in network traffic flows is a challenging task due to the inherent complexity and dynamic nature of network data.

Most existing machine learning–based detection methods focus primarily on network-level anomalies while overlooking anomalies in individual user behaviour. In real-world scenarios, abnormal network activities can affect user interests and system integrity. To address this limitation, this paper proposes an anomaly detection model based on time-decay closed frequent patterns. The model extracts closed frequent patterns from each user's network traffic and incorporates a time-decay factor to appropriately weigh recent versus historical behavior. Given the evolving nature of user activities, the framework also includes a model update strategy to maintain detection accuracy over time. Moreover, the use of closed frequent patterns enhances interpretability by providing clear explanations for detected anomalies.

Experimental results indicate that the proposed method effectively identifies abnormal user behaviour, achieving network anomaly detection performance comparable to state-of-the-art techniques and significantly outperforming baseline approaches. Figure 1 illustrates a sample of the network dataset used in this study.

I.1 Swarm Intelligence

Swarm Intelligence is inspired by the collective behaviour of many simple agents that operate through decentralized control and self-organization. While each agent has limited individual capability, their coordinated interactions enable the group to achieve superior problem-solving and decision-making performance compared to individual members. With the rapid increase in network attacks in recent years, intrusion detection systems have become a major focus of research. This paper introduces the design and implementation of a software architecture that integrates Particle Swarm Optimization (PSO) with Non-Negative Matrix Factorization (NMF) to classify computer network behavior based on sequences of system calls.

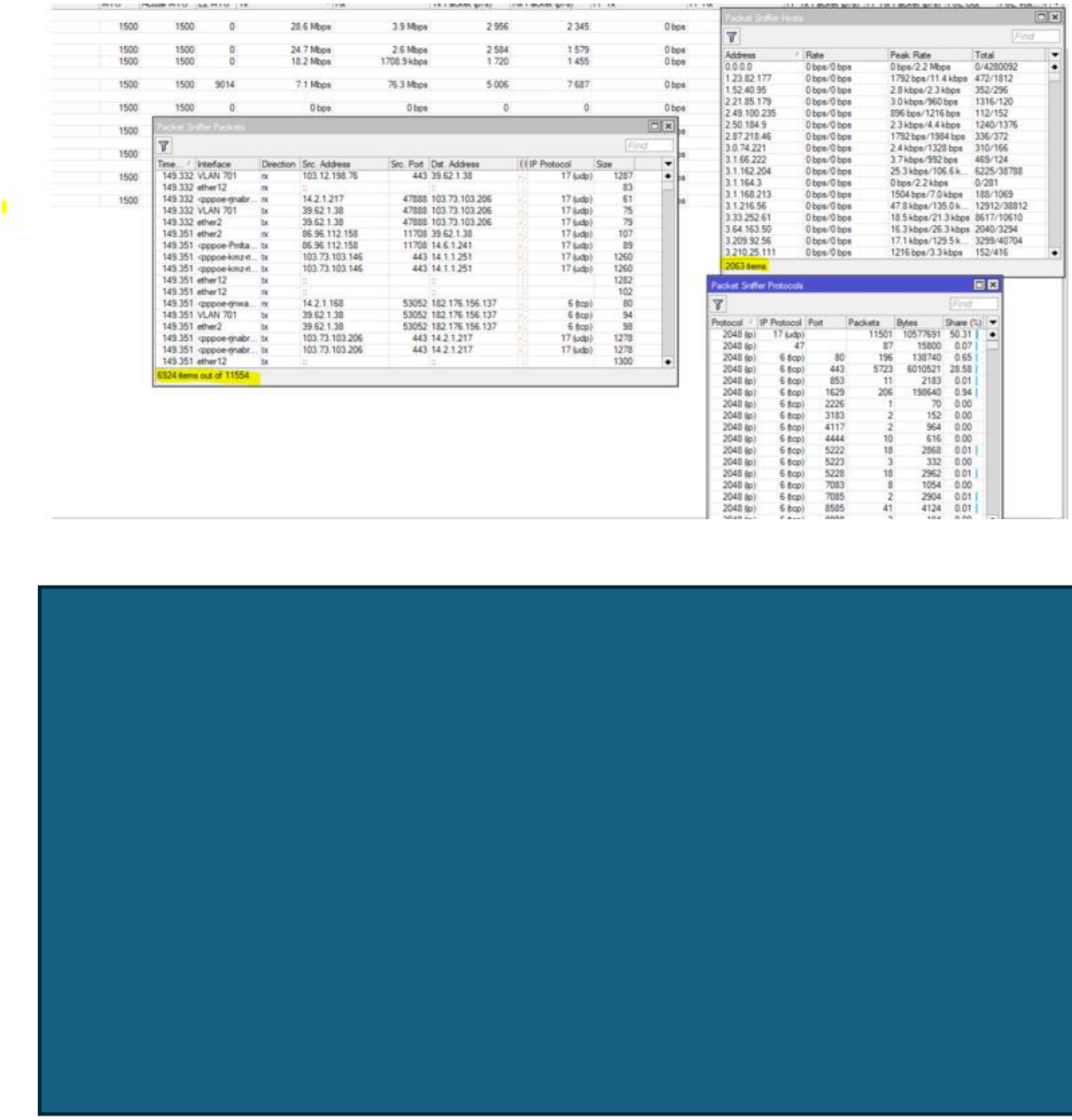


**Fig 1 : Sample KDDC Data for Network Anomaly Detection**

II. **Existing Methods**

The following table 2 summarizes methods and remarks on major research trends emerging after 2010. One study presents a software architecture designed using Ant Colony Optimization (ACO), where ACO is integrated with a Non-Negative Matrix Factorization approach to classify computer network behavior based on sequences of system calls. According to Dukka, the advancement of efficient swarm optimization techniques is largely inspired by the peer-to-peer learning behavior observed in social colonies. Swarm Intelligence (SI) has become deeply embedded in IoT (Internet of Things) environments and IoT-based systems, enabling logical and autonomous operational control.

Moises et al. [1] describe the Particle Swarm Optimization (PSO) algorithm as an evolutionary computation technique characterized by low computational complexity, the ability to avoid local optima, and minimal dependence on input parameters when compared to other evolutionary algorithms such as Genetic Algorithms (GA). Khushboo et al. [5] note that PSO has gained significant attention in the research community due to its potential to meet the increasing demand for reliable and intelligent Intrusion Detection Systems (IDS). Diptam et al. [3] further explain that PSO is inspired by the behavior of birds, fish, or bees searching for food sources—initially unaware of the optimal location but eventually converging on it through cooperative communication.

Artificial Immune Systems (AIS) have also been widely explored over the past decade, particularly for anomaly detection. Much of this research focuses on negative selection, as it naturally aligns with detecting anomalies. Sanju et al. [7] demonstrate that Swarm Intelligence integrated with data mining can produce lightweight yet robust techniques capable of effectively detecting and identifying data flows. In neural network-based approaches, the system learns to interpret the behavior of various daemons and users for intrusion detection. A major advantage of neural networks is their tolerance to uncertainty and noisy data, as well as their ability to infer patterns without prior knowledge of data regularities.

Stephanie et al. [8] emphasize that detecting network intrusions remains a significant challenge in cybersecurity, driven by increasingly sophisticated attack strategies and complex traffic patterns. Although machine learning offers promising solutions, issues such as class imbalance, feature complexity, and model efficiency continue to pose difficulties. Their study aims to evaluate and enhance both supervised and unsupervised machine learning models for network anomaly detection using the KDDCup'99 dataset, with the goal of improving accuracy, scalability, and interpretability in real-world settings.

**TABLE 2 : Existing Methods**

| Sno | Authors | Year | Methods | Remarks |
|---|---|---|---|---|
| 1 | Apoorv [2] et al. | 2018 | Ant Colony Optimization (ACO), Anomaly-based Intrusion Detection System | Classifying a computer network behaviour properly |
| 2 | Moises et al | 2010 | Digital Signature of Network Segment (DSNS), Particle Swarm Optimization-based clustering (PSO-Cls) | PSO-Cls algorithm, swarm intelligence is combined with K-means clustering, in order to achieve high convergence rates. |
| 3 | Sanju[7] et al | 2018 | Baysien Network, Machine Learning Algorithm, Support Vector Machine (SVM) | Classify the output properly |

| | | | | |
|---|---|---|---|---|
| 4 | Diptam Dutta [3] et al | 2013 | Simulink Model and ANN, Artificial Immune Systems (AISs) | The distance measure adopted in this work is the Euclidean distance, which consists of the straight-line distance between two points is clearly |
| 5 | Stephanie [8] et al | 2025 | Isolation Forest, Naive Bayes, XGBoost, LightGBM, SVM, Random Forest, and Logistic Regression. | The results bring to light the importance of model selection in optimizing performance and minimizing false positives in network security applications. |
| 6 | Ali Rahmani [1] et al | 2014 | Graph Based System for Enhancement | Discovers contextual outliers in sequential data. |
| 7 | Khushboo [5] et al | 2010 | Particle Swarm Optimization, Intelligent ToIntrusion Detection | To Provide Satisfied Output and optimized output. |

Apoorv explains that a software architecture is modelled and implemented which uses Ant Colony Optimization (ACO), It is combined with Non-Negative Matrix Factorization method for classifying a computer network behaviour as a sequence of system calls. Dukka said that the development of various efficient swarm optimization methods is largely due to the peer-to-peer learning behaviour of social colonies. SI is deeply engaged in the realm of IoT (Internet of Things) and IoT-based systems to control the operations logically. Moises et al[1] explains PSO algorithm is an evolutionary computation technique whose main characteristics include low computational complexity, ability to escape from local optima, and small number of input parameters dependence, when compared to other evolutionary algorithms, e.g. genetic algorithms (GA). Khushboo et al[5] Particle Swarm Optimization is currently attracting considerable interest from the research community, being able to satisfy the growing demand of reliable & intelligent Intrusion Detection System (IDS).Diptam et al[3] explains PSO is based on the principles that flock of birds, school of fish, or swarm of bees searches for food sources where at the beginning the perfect location is not known. However, they eventually they reach the best location of food source by means of communicating with each other. Artificial Immune Systems (AISs) have been extensively researched in the last decade, mainly for anomaly detection. Much research has been conducted on using negative selection, as that model lends itself conveniently to anomaly detection. Sanju et al [7] demonstrates Swarm intelligence has been integrated with data mining to generate lightweight but robust methods to detect and identify the flow of data effectively. In the neural network approach, it learns to interpret the nature of the divergent daemons and users into the intrusion detection system. The prime beneath of neural networks is their strength to uncertainty information and unreliable data and their ability to determine the results from data without having previous knowledge of the data regularities. Stephanie [8]et al explains the discovery of network incursions rests a essential contest in cybersecurity due to developing attack strategies and complex network traffic patterns. While machine learning offers promising solutions, problems such as data imbalance, feature complexity, and model efficiency persist. This study aims to estimate and improve both supervised and unsupervised machine learning models for anomaly detection in network traffic, using the KDDCup'99 dataset, to improve accuracy, scalability, and interpretability in real-world applications.

## III **Proposed system**

In this work, we employed the One-Class SVM (OC-SVM) technique to detect anomalies in real network traffic. Our contribution includes an automatic hyperparameter selection method based on the EROS similarity index ,

which effectively determines the learning configuration. This approach produced highly satisfactory results. Novelty detection offers several advantages for complex environments such as network traffic analysis; most notably, it does not require faulty or attack data. Since anomalies can occur in numerous and unpredictable forms, it is impractical to characterize all possible cases or collect representative datasets for each. Future challenges include handling streaming data and achieving real-time processing, which can be demanding due to resource constraints and large data volumes. Additionally, exploring novelty detection within distributed frameworks remains an important direction for further research.

Support Vector Machines (SVMs) aim to identify parameters (w, b) that place a separating hyperplane at the maximum possible distance from the nearest training samples of each class, thereby minimizing generalization error. This maximum distance is known as the margin. SVMs were initially developed for linearly separable classification tasks but were later extended to handle non-linearly separable data. Soft-margin SVMs allow certain samples to violate the margin, and non-linear decision boundaries are achieved by mapping input data into a higher-dimensional feature space using a non-linear function $\Phi(x)$. Although samples may not be linearly separable in the original space, they may become separable in this transformed feature space. When the separating hyperplane in the feature space is mapped back to the input space, it forms a non-linear boundary. To prevent overfitting in the presence of noise, slack variables $\xi$ are introduced, and the parameter $C > 0$ (Eq. 2) controls the balance between minimizing classification error on the training data and maximizing the margin. The following Table 3 explains methods and techniques for detecting network anamalies.

**TABLE 3 : Proposed Method to Detect Network Anomaly**

| Sno | Stages in Proposed Method | Methods and Techniques | Purpose |
|---|---|---|---|
| 1 | Data Collection & Sample Data | GPA Tool | collection of real traffic samples and generation of the DSNS |
| 1 | Preprocessing | Min-Max Scaling or Standardization & Normalization | Convert the network data to a common scale or range, such as transforming edge weights to a [0,1] range. |
| 2 | Enhancement | Graph-based smoothing methods and Outlier Detection | Identify and handle outliers that could skew the results. |
| 3 | Feature Extraction | Heuristic-Based Detection | To identify suspicious behavior based on known patterns or characteristics. |
| 4 | Feature Selection | Filter Methods, Correlation Based Feature Selection, | The process of choosing the most relevant features from the extracted set to improve model performance and reduce complexity. |
| 5 | Classification | Machine learning Methods : OC -Support Vector Machine, k-nearest neighbour (kNN), ANN, Particle Swarm | Receiver Operating Characteristic (Roc) curve and Area Under the Curve(AUC) measure the performance of a classification model across different thresholds. |

| | | | |
|---|---|---|---|
| | | Optimization (PSO), Convolutional Neural Networks (CNNs). | |

**III.A. Dataset**

In this paper, we conducted experiments using two datasets: BUCT and KDD99. We first assessed user behavior anomaly detection performance on the BUCT dataset, followed by evaluating network anomaly detection performance on both BUCT and KDD99. The BUCT dataset consists of real network traffic captured from the library of the Beijing University of Chemical Technology. It contains traffic generated by 274 users and two web servers. Three days of network traffic were collected: two days were used to construct the user behavior model, while the remaining day was reserved for testing.

Since the BUCT dataset does not include malicious traffic, we used Scapy to synthesize SYN-flood attack traffic. Scapy is a powerful Python-based interactive packet manipulation tool and library capable of crafting or decoding packets from a wide range of protocols.

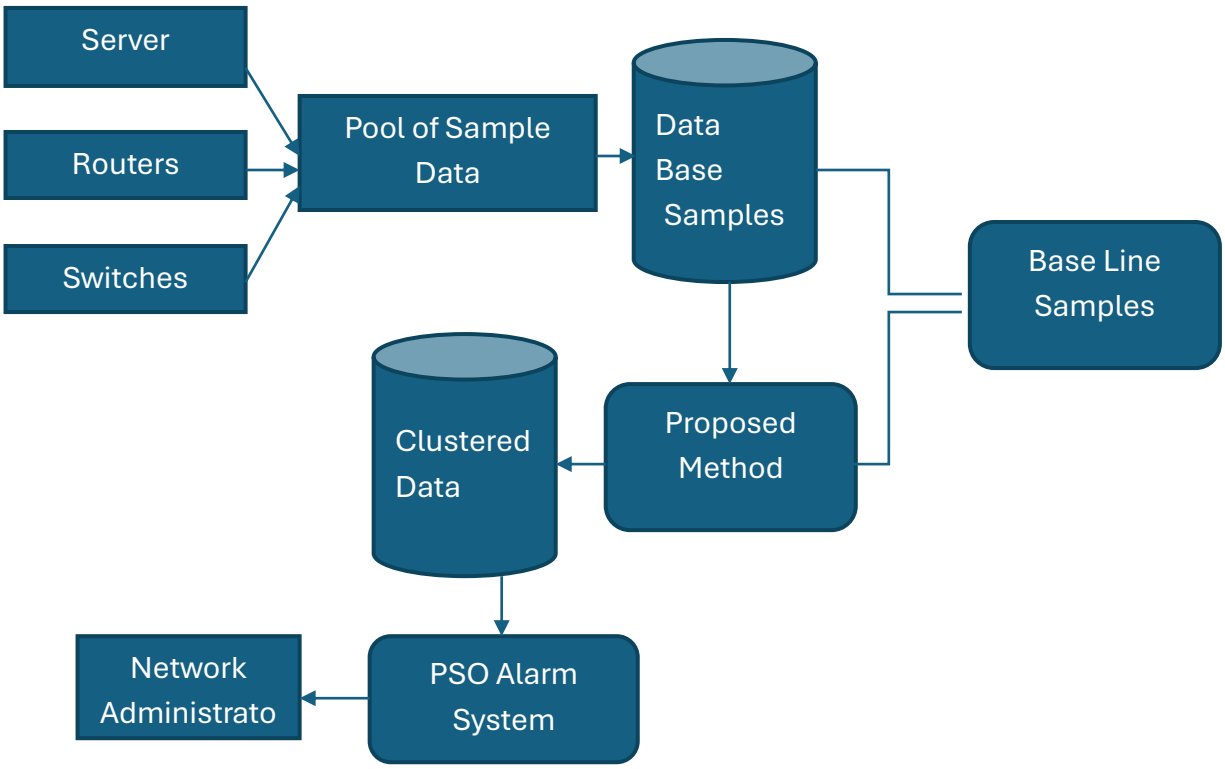



**Fig 2 : Entire structure of Network Anomaly Detection**

A two-layer feed-forward Artificial Neural Network (ANN) was designed with sigmoid activation functions in both the hidden and output layers. The hidden layer consists of 12 neurons, while the output layer contains 6 neurons. The neural network was trained using the Neural Network Pattern Recognition Tool in MATLAB 7.11.0.584 (R2010b). Training was performed through a backpropagation-based conjugate gradient algorithm optimized with scaled Particle Swarm Optimization. This configuration achieved a mean squared error of approximately $10^{-2}$ and a best validation performance of 0.021587 at epoch 33. After training, a basic Simulink model of the ANN-based attack classifier was generated.

Figure 2 illustrates the components of the proposed network anomaly detection system. The GBA tool is responsible for collecting real traffic samples and generating the DSNS. The PSO-Cls module computes the cluster centroids using both real traffic data and the DSNS [6]. The PSO Alarm module then evaluates the distance between the derived centroids and incoming traffic samples [9] to identify potential anomalies.

The anomaly detection process is divided into two primary stages:

1. **Clustering Stage (PSO-Cls System):**

Traffic data obtained from SNMP objects and their corresponding DSNS are collected every 300 seconds and analyzed separately. For each 300-second interval, traffic data and DSNS are clustered simultaneously. A centroid is then computed for each cluster, representing the expected behavior of the traffic samples within that cluster [10]. The pseudocode for the clustering and centroid calculation procedure is provided in the algorithm section. The resulting clustered data and centroids serve as inputs for the next stage.

2. **Anomaly Evaluation Stage (PSO Alarm System):**
In this step, the PSO Alarm module examines the clustering results to determine whether anomalies are present in the analysed interval. It evaluates how close each traffic sample is to its corresponding cluster centroid . The Euclidean distance—defined as the straight-line distance between two points—serves as the distance metric. A sample is flagged as anomalous if its Euclidean distance from its cluster centroid exceeds a defined threshold λ. When this occurs, the PSO Alarm system triggers an alert to notify the network administrator.

**Proposed Algorithm**

Input Fuction

Parameter List :

1. Swarm size : SZ
2. Positive acceleration constants: x1 and x2
3. Inertia weight : W
4. Maximum generation : MG
5. Fitness Threshold : MF
6. Fitness : F

Output: Global best position with KDD Dataset

Algorithm

Step 1: Initialize a population of particles with random positions and velocities on d=1,...,41 NSL-KDD features dimensions pbesti=0, Gbest=0, Iter=0.

Step 2: while Iter < MG or gbest < MF do

Step 3: for i = 1 to number of particles m do

Step 4: F(i)=Evaluate(i)

Step 5: if F(i) > Fi (pbesti) then

Step 6: F(pbesti)= F(i)

Step 7: Update pid = xid

Step 8: end if

Step 9: if F(i) > Gbest then

Step 10: Gbest=Fitness(i)

Step 11

Step 12: end if

Step 13: for each dimension d do

Step 14: Update the velocity vector.

Step 15: Update the particle position.

Step 16: end for

Step 17: end for

Step 18: Iter= Iter+1

Step 19: end while

Step 20: Return the Global best position. 3

PSO- CLs System Input: real traffic, DSNS

Output: clustered traffic and DSNS, cluster centroids.

Iteration: 100 Best (PSO): 0.00447367421062735

[ 2 4 6 8 9 10 11 15 19 20 24 25 27 31 33 34 35 37]

[0 0 0 0 0 0 0 0 109 13 0.0 0.12 0.0 0.06 0.0 1.0 0.0]

**3.bPreprocessing**

In network anomaly detection, the choice of feature scaling technique depends on the characteristics of the network data. When the data exhibits a relatively stable range with few extreme outliers, Min–Max Scaling is often appropriate. However, if outliers are frequent or the data distribution is unknown or unstable, Standardization is generally preferred due to its robustness.

Preprocessing for network anomaly detection using swarm intelligence typically involves three major steps: data cleaning, normalization or scaling, and dimensionality reduction. Data cleaning removes irrelevant, inconsistent, or noisy entries. Normalization or scaling adjusts feature values to a comparable range. Dimensionality reduction—through feature selection or extraction—reduces the number of attributes, improving model performance and mitigating the "curse of dimensionality."

Both Min–Max Scaling (normalization) and Standardization (Z-score normalization) are valid techniques for preparing network traffic data. The choice between them depends on the nature of the data and the requirements of the detection algorithm. Min–Max Scaling rescales features to a fixed interval, typically [0, 1] or [-1, 1], using:

$$y' = \frac{y - \min(y)}{\max(y) - \min(y)}$$

This method ensures that all features lie within the same bounded range, which is advantageous for algorithms sensitive to feature magnitude—such as neural networks using sigmoid or tanh activations, as well as distance-based methods like k-NN or SVM. It also preserves the original distribution of the data.

Standardization, on the other hand, transforms features to have a mean of 0 and a standard deviation of 1. It is generally preferred when network data contain outliers or exhibit unstable statistical properties, as it is less affected by extreme values.

**C.Enhancement**

Graph-based smoothing techniques leverage the structural relationships within a graph to reduce noise by adjusting data points toward the values of their neighboring nodes. These methods can also support outlier detection, as nodes that deviate significantly from the smoothed patterns or are inconsistent with their neighbors

can be flagged as anomalies. Common approaches include using graph-based similarity measures to identify abnormal nodes and transforming data into graph representations for analysis with models such as Graph Autoencoders (GAEs).

**III.D .Feature Extraction & Feature Selection**

Heuristic-based anomaly detection in network traffic identifies threats by analyzing behavioral patterns rather than relying solely on predefined signatures. This proactive approach evaluates characteristics such as traffic flow, behavior, and code structure to detect potentially malicious activities, including zero-day attacks and polymorphic malware. It operates through four main mechanisms.

1. Behavioral Analysis:

The system monitors network behavior to identify suspicious actions, such as unusual connection attempts or deviations from established traffic profiles.

2. Pattern Recognition:

Algorithms analyze network traffic to detect patterns that do not conform to expected or normal behavior.

3. Rule-Based and Statistical Methods:

These techniques apply predefined rules or statistical models to calculate a threat score. If this score exceeds a threshold, the activity is flagged as anomalous.

4. Proactive Threat Identification:

By focusing on how traffic behaves rather than relying on known signatures, these systems are capable of detecting new and previously unseen threats that traditional signature-based methods would overlook.

This approach offers several advantages, including the ability to detect unknown or zero-day attacks, identify evolving threats such as polymorphic malware, and provide a dynamic and adaptive defense that is more flexible than static signature-based systems.

An AI framework for network anomaly detection that integrates swarm intelligence with machine learning combines the collective decision-making strengths of swarm algorithms with the predictive capabilities of machine learning models[1]. In such a hybrid system, swarm intelligence can be used for tasks like feature selection or model optimization, while machine learning algorithms classify network traffic as normal or anomalous based on the extracted or optimized features.

This combined approach provides several key benefits:

1. Improved Accuracy: By leveraging the strengths of both swarm intelligence and machine learning, the system achieves higher detection performance than using either technique alone.

2. Adaptability: The framework can continuously learn and adjust to new or evolving network patterns and attack strategies.

3. Scalability: Distributed swarm intelligence enables the system to efficiently process and analyze large volumes of network traffic data.

4. Enhanced Resilience: Swarm-based optimization contributes to overall system robustness by improving threat detection and response capabilities.

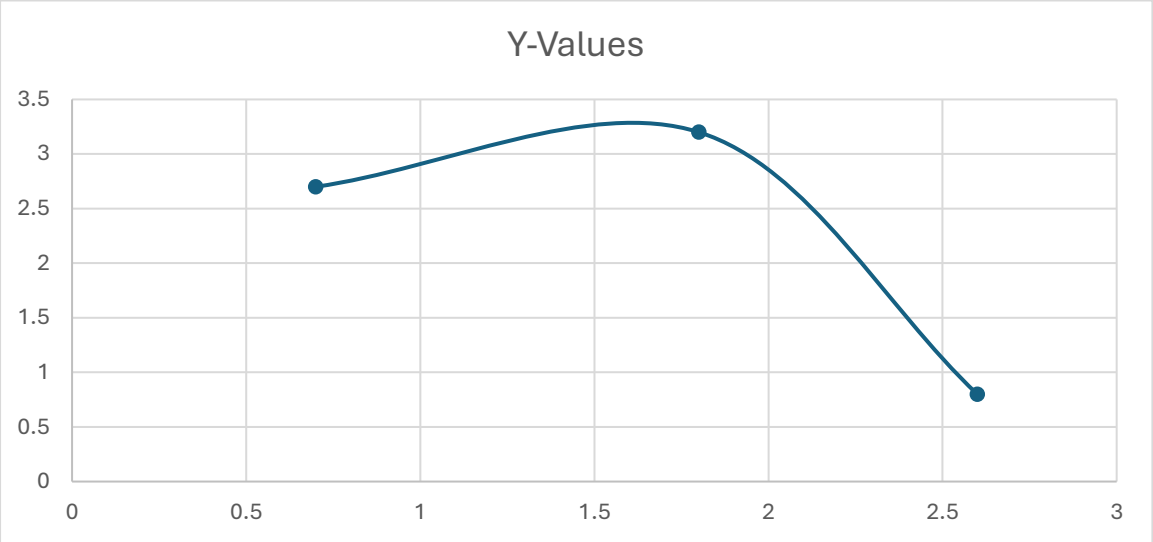
Experiments & Results

Iteration:  Best (ONE-PSO): 0.011073684210526362

Iteration: Best (ONE- PSO): 0.00526315789473684688

Iteration : Best(ONE-PSO) : 0.00526315789493684688

The following chart explains the detection rate is very high using swarm intelligence.




**Conclusion**

The primary objective of this framework is to present a comprehensive overview of the requirements and effectiveness of Intrusion Detection Systems (IDS). This paper details the various types of IDS, their lifecycle, application domains, attack categories, and associated tools. These systems have become essential components of modern security practices, particularly in corporate environments and for network users. Additionally, Intrusion Prevention Systems (IPS) are discussed to highlight preventive security measures.

The IDS lifecycle phases are clearly outlined, and although significant progress has been made, several challenges remain. Techniques such as anomaly detection and misuse detection are explained, with scope for integrating additional methods. Future work may include the application of the Firefly Algorithm to analyse techniques for capturing high-speed network traffic and enhancing real-time system accuracy, aiming for a 0% false detection rate.

In this study, an ANN–PSO–based Intrusion Detection and Prevention System (IDPS) was developed using MATLAB 7.11.0.584 (R2010b), yielding highly encouraging test results. A further objective is to implement this model on FPGA (Spartan-3E) to support hardware-integrated solutions for monitoring, diagnosis, and intrusion detection. Although the Xilinx library provides most of the required blocks for Artificial Neural Networks, certain functions—such as the sigmoid function—are not included. This work proposes an approximation of the sigmoid function, which will be implemented on the FPGA using Xilinx library components.

References:

1. Ali Rahmani · Salim Afra , Omar Zarour , Omar Addam , Negar Koochakzadeh , Keivan Kianmehr , Reda Alhajj · Jon Rokne ," Graph-based approach for outlier detection in sequential data and its application on stock market and weather data",Elsevier, Knowledge Based Systems,Vol 61,pages 89-97,2014.
2. Apoorv Saxena, Carsten Mueller," Intelligent Intrusion Detection in Computer Networks using Swarm Intelligence**",** International Journal of Computer Applications, Volume 179 - Number 16,2018.
3. Diptam Dutta, Kaustav Choudhury," Network Anomaly Detection using PSO-ANN", International Journal of Computer Applications (0975 – 8887) Volume 77– No.2, September 2013.
4. Dukka Karun Kumar Reddy, Janmenjoy Nayak, H. S. Behera, Vimal Shanmuganathan, Wattana Viriyasitavat Gaurav Dhiman," A Systematic Literature Review on Swarm Intelligence Based Intrusion Detection System: Past, Present and Future", Archives of Computational Methods in Engineering , Volume 31, pages 2717–2784, (2024)

5. Khushboo Satpute, Shikha Agrawal,Jitenra Agrawal, Sanjeev Sharma,"A Survey on Anomaly Detection in Network Intrusion Detection System Using Particle Swarm Optimization Based Machine Learning Techniques", In book: Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) (pp.441-452),2013.
6. Lucas D. H. Sampaio, Bruno B. Zarpelão,Joel J. P. C. Rodrigues,Taufik Abrão, Mario Lemes Proença," Networking Anomaly Detection Using DSNs and Particle Swarm Optimization with Re-Clustering", IEEE Global Telecommunications Conference GLOBECOM 2010.
7. Sanju Mishra Tiwari, Rafid Sagban, Ali Yakoob Al-Sultan, Niketa Gandhi," Swarm intelligence in anomaly detection systems: an overview", ,International Journal of Computers and Applications 43(8):1-10, 2018.
8. Stephanie Ness, Vishwanath Eswarakrishnan,Harish Sridharan,Varun Shinde,Naga Venkata Prasad Janapareddy,Vineet Dhanawat,"Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques",IEEE Access, Jan 2025.
9. Panjatcharam V G,Selvanayaki K,"An Intelligent AI Framework for Detecting Anomalies in Network Traffic Using Swarm Intelligence and Machine Learning",Industrial Engineering Journal,Volume :53,Issue 10,2024.
10. Patcha and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," International Journal Computer Networks: The of Computer Telecommunications Networking, 2007.