9. BIBLIOGRAPHY

1. Network Intrusion Detection Method Based on GAN Model (https://ieeexplore.ieee.org/document/9240732)

2. A Method for Network Intrusion Detection Based on GAN-CNN-BiLSTM (https://dx.doi.org/10.14569/IJACSA.2023.0140554)

3. GAN-based imbalanced data Intrusion Detection System (https://ieeexplore.ieee.org/document/8587389)

4. Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning (https://ieeexplore.ieee.org/document/8587389)

5. Machine Learning-Based Adaptive Synthetic Sampling Technique for Intrusion Detection (https://www.mdpi.com/2076-3417/13/11/6504)

6. A Deep Learning Model for Network Intrusion Detection with Imbalanced Data (https://www.mdpi.com/2079-9292/11/6/898)

7. A novel statistical analysis and autoencoder driven intelligent intrusion detection approach (https://www.sciencedirect.com/science/article/pii/S0925231219315759)

8. GIDS: GAN BASED INTRUSION DETECTION SYSTEM FOR IN-VEHICLE NETWORK (https://arxiv.org/pdf/1907.07377.pdf)

9. Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network (https://ieeexplore.ieee.org/document/8998253)

10. A Deep Learning Approach for Network Intrusion Detection System (https://ieeexplore.ieee.org/document/8264962)

Improving the model's real-time detection capabilities is another critical area of development. By focusing on increasing the model's processing speed and efficiency, it can enable faster identification and mitigation of network intrusions, which is crucial in high-traffic environments where delays could lead to potential damages. Enhanced real-time capabilities would not only minimize the risk of prolonged intrusions but also provide cybersecurity teams with the responsiveness needed to address threats as they happen, making the system especially valuable in industries where downtime or data breaches can have serious consequences.

8. FUTURE SCOPE

Our rare class attack detection system holds significant potential for further growth and impact in cybersecurity. A major avenue for expansion lies in integrating this system into broader cybersecurity frameworks. By embedding the model into real-time monitoring infrastructures, it could boost the detection and response capabilities of existing security tools, allowing for a more proactive defense against rare and sophisticated attacks. Such integration would offer an additional layer of protection by enabling immediate threat identification and faster response times, ultimately contributing to a more resilient and secure network environment.

A key aspect of the system's future is its adaptability to evolving threats. The cybersecurity landscape is in constant flux, with new attack vectors and sophisticated techniques emerging regularly. To keep pace, our model must be continuously updated and retrained with the latest datasets. This adaptability is essential for ensuring the model remains relevant and effective against the latest cyber threats, helping organizations stay ahead of attackers who constantly refine their tactics. By incorporating recent data and leveraging advanced algorithms, the system could provide a robust line of defense against even the most cutting-edge threats, making it an invaluable asset in dynamic security contexts.

Expanding the system to work with a variety of datasets and environments is also a promising direction. While initially developed using the CICIDS dataset, the system could be broadened to incorporate datasets from different industries, geographical regions, and network types. This expansion would enhance the model's versatility, making it applicable across multiple sectors, including finance, healthcare, and government. By diversifying its training sources, the system can better handle the unique threat landscapes and network configurations of each sector, increasing its utility in a wide range of real-world scenarios.

Additionally, integrating the detection system with advanced analytical platforms and tools could significantly enhance its capabilities. Visualization tools and data analytics platforms would allow cybersecurity professionals to analyze network traffic and attack patterns in real time, gaining deeper insights into trends and vulnerabilities. With these insights, security teams can make more informed decisions, respond more effectively to threats, and adjust strategies based on real-time data, ultimately strengthening an organization's defensive posture.

the generalization of both models on minority classes. This improvement supports the effectiveness of GANs in creating high-quality synthetic data that benefits various machine learning algorithms.

Final Implications

In summary, although accuracy metrics initially appeared better on unbalanced data, the balanced model's confusion matrix showed that the proposed approach of generating synthetic data via GAN led to substantial performance improvements in detecting minority attack classes. By addressing the class imbalance, the system achieved a more reliable performance in real-world intrusion detection scenarios, where detecting rare attacks is critical. This study demonstrates that GAN-based data augmentation is a viable solution for handling class imbalance issues, leading to a more robust and fair intrusion detection model.

7. CONCLUSION

This project focused on enhancing rare class detection in an intrusion detection system by addressing the issue of class imbalance using a Generative Adversarial Network (GAN) with the CICIDS-2017 dataset. The dataset initially suffered from an extreme imbalance, where normal traffic was highly represented, but attack classes were underrepresented, making it challenging for machine learning models to identify attacks accurately. This imbalance caused traditional machine learning algorithms to perform well on accuracy metrics but poorly on recall, precision, and F1-score for minority classes. The results were particularly skewed towards the majority (normal traffic) class, thereby failing to detect attack classes effectively.

To mitigate this, we applied a GAN to generate synthetic samples for the minority attack classes, effectively balancing the dataset. After creating a balanced dataset, we implemented two models—K-Nearest Neighbors (KNN) and Random Forest—to evaluate the impact of the synthetic data. The performance was evaluated based on confusion matrices, which provided deeper insights beyond traditional metrics like accuracy, precision, recall, and F1-score.

While accuracy, recall, precision, and F1-score were initially higher for the unbalanced dataset due to the model's focus on the majority class, the confusion matrices revealed the true improvement achieved with the balanced dataset:

1. **Improved Detection of Attack Classes**: The confusion matrix of the balanced model demonstrated a significant increase in correctly classified samples for each attack class. This improvement suggests that synthetic samples generated by the GAN allowed the models to learn distinct patterns associated with minority classes, enhancing their ability to detect attacks.
2. **Reduced Bias Toward Majority Class**: In the unbalanced dataset, models exhibited a strong bias towards the normal class, often misclassifying attack classes as normal traffic. The balanced dataset reduced this bias, leading to a more equitable classification across all classes, as seen in the confusion matrix of the proposed model.
3. **Balanced Model Performance with KNN and Random Forest**: Both KNN and Random Forest algorithms showed improvements in their ability to classify rare classes with the balanced dataset, indicating that the GAN-generated data enhanced

- **Overall Accuracy**: The proposed model's confusion matrix indicates an increase in correct classifications across all classes, demonstrating a better overall model performance.

These results demonstrate that integrating GANs into the data preprocessing pipeline of intrusion detection systems can significantly enhance the detection of rare and potentially critical cyber threats. This approach not only addresses the challenge of class imbalance but also contributes to the development of more reliable and secure network defense mechanisms.

classification report:

```
Existing Model Evaluation (Unbalanced Data):
Classification Report:

                precision    recall  f1-score   support

        Benign       1.00      1.00      1.00      2550
      DoS Hulk       1.00      1.00      1.00        61
  DoS GoldenEye       1.00      1.00      1.00      1327
   DoS slowloris       0.92      1.00      0.96        24
DoS slowhttptest       1.00      0.97      0.99        38
 Heartbleed Port       0.00      0.00      0.00         0

     micro avg       1.00      1.00      1.00      4000
     macro avg       0.82      0.83      0.82      4000
  weighted avg       1.00      1.00      1.00      4000
```

Figure 17: Unbalanced data Classification report RF 10 estimators

```
Proposed Model Evaluation (Balanced Data):
Classification Report:

                precision    recall  f1-score   support

        Benign       1.00      1.00      1.00      2524
      DoS Hulk       0.81      0.90      0.85      2619
  DoS GoldenEye       0.92      0.83      0.87      2524
   DoS slowloris       0.87      0.89      0.88      2557
DoS slowhttptest       0.49      0.53      0.51      2580
 Heartbleed Port       0.47      0.41      0.44      2506

        accuracy                           0.76     15310
     macro avg       0.76      0.76      0.76     15310
  weighted avg       0.76      0.76      0.76     15310
```

Figure 18: balanced data Classification report RF 10 estimators

Analysis Summary

- **Class Imbalance Issue**: In the existing model, the imbalanced dataset led to poor classification of attack classes, as the KNN model tended to favor the majority class (normal traffic).
- **Improvement with GAN**: The proposed model, with synthetic data generated by the GAN, shows a balanced classification across all classes, especially improving the detection of minority attack classes. This suggests that the GAN effectively generated synthetic data that made the attack classes more representative, thus allowing the KNN model to better learn the patterns in minority classes.

Classification Report:

Existing Model Evaluation (Unbalanced Data):
Classification Report:

	precision	recall	f1-score	support
Benign	1.00	1.00	1.00	2550
DoS Hulk	1.00	1.00	1.00	61
DoS GoldenEye	1.00	1.00	1.00	1327
DoS slowloris	0.96	1.00	0.98	24
DoS slowhttptest	1.00	0.97	0.99	38
Heartbleed Port	0.00	0.00	0.00	0
micro avg	1.00	1.00	1.00	4000
macro avg	0.83	0.83	0.83	4000
weighted avg	1.00	1.00	1.00	4000

Figure 13: unbalanced data classification report RF 20 estimators

Proposed Model Evaluation (Balanced Data):
Classification Report:

	precision	recall	f1-score	support
Benign	1.00	1.00	1.00	2524
DoS Hulk	0.88	0.94	0.91	2619
DoS GoldenEye	0.96	0.87	0.91	2524
DoS slowloris	0.90	0.94	0.92	2557
DoS slowhttptest	0.49	0.51	0.50	2580
Heartbleed Port	0.47	0.44	0.46	2506
accuracy			0.78	15310
macro avg	0.78	0.78	0.78	15310
weighted avg	0.78	0.78	0.78	15310

Figure 14: balanced data classification report RF 20 estimators

Random Forest (for 10 estimators):

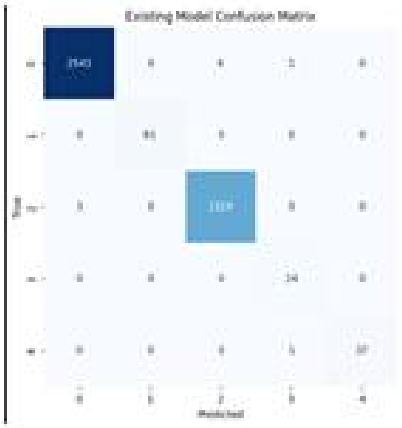


Figure 15: Unbalanced data CM using RF 10 estimators

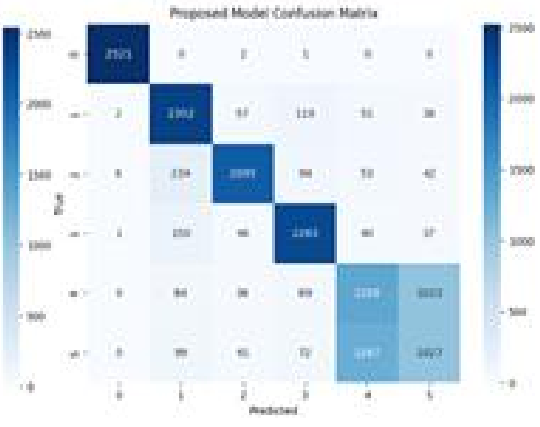


Figure 16:balanced data CM using RF 10 estimators

classification Report:

```
Existing Model Evaluation (Unbalanced Data):
Classification Report:

                precision    recall  f1-score   support

        Benign       0.99      0.99      0.99      2550
      DoS Hulk       0.95      0.97      0.96        61
  DoS GoldenEye       0.98      0.99      0.99      1327
    DoS slowloris       0.95      0.75      0.84        24
DoS slowhttptest       0.94      0.87      0.90        38
 Heartbleed Port       0.00      0.00      0.00         0

      micro avg       0.99      0.99      0.99      4000
      macro avg       0.80      0.76      0.78      4000
   weighted avg       0.99      0.99      0.99      4000
```

Figure 9: Classification report unbalanced data KNN

```
Proposed Model Evaluation (Balanced Data):
Classification Report:

                precision    recall  f1-score   support

        Benign       0.99      0.99      0.99      2524
      DoS Hulk       0.95      0.97      0.96      2619
  DoS GoldenEye       0.98      0.92      0.95      2524
    DoS slowloris       0.96      0.97      0.96      2557
DoS slowhttptest       0.51      0.59      0.55      2580
 Heartbleed Port       0.50      0.43      0.46      2506

       accuracy                           0.81     15310
      macro avg       0.82      0.81      0.81     15310
   weighted avg       0.82      0.81      0.81     15310
```

Figure 10: Classification report balanced data KNN

2. Random Forest (for 20 estimators):

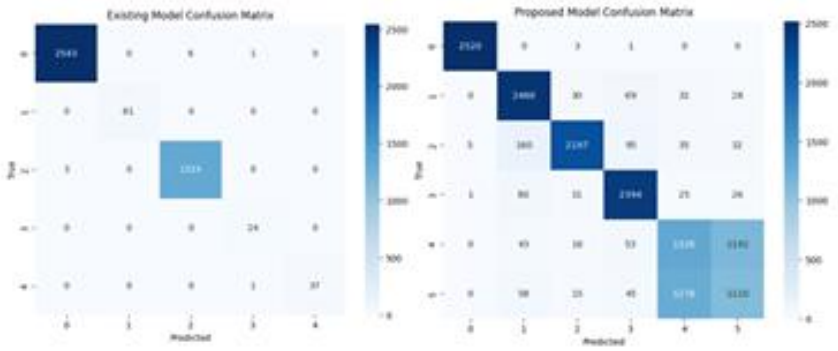


Figure 11: unbalanced data confusion matrix RF Figure 12: balanced data confusion matrix RF

- Class 1 has only 59 correctly classified samples, with a few samples misclassified.
- Class 2 has 1313 correct predictions but also suffers from some misclassifications.
- Classes 3 and 4 show poor detection with just 18 and 33 samples correctly classified, respectively.

This pattern highlights that the model has a strong bias towards the majority class (class 0) but struggles to identify minority classes (attack classes), which is expected due to the unbalanced data distribution in the CICIDS-2017 dataset.

Proposed Model (Balanced Data with GAN)

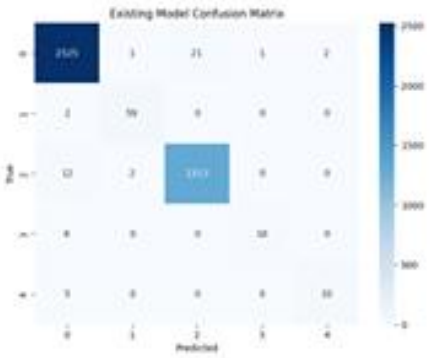
The proposed model's confusion matrix, after balancing the dataset with a GAN, shows significant improvements in identifying attack classes. Here's a breakdown by each class:

1. **Class 0 (Normal Traffic)**: The model correctly classified 2503 samples, similar to the existing model. However, there are fewer misclassifications into other classes compared to the existing model.
2. **Class 1 to Class 5 (Attack Classes)**: The classification for attack classes is much more balanced and accurate in the proposed model:
    - Class 1 has 2523 correctly classified samples, indicating substantial improvement from 59 in the existing model.
    - Class 2 has 2333 correct predictions, an improvement from 1313 in the existing model.
    - Classes 3 to 5 show dramatic improvements in correct classifications, with 2445, 1553, and 1429 correctly classified samples, respectively. This suggests that the proposed model is able to detect minority attack classes far better than the existing model.

6. RESULTS

The project "Enhancing Rare Class Detection using GAN Model in Intrusion Detection System" yielded significant improvements in the detection of rare class attacks in the CICIDS-2017 dataset. The key results are as follows:

1. For KNN Model:



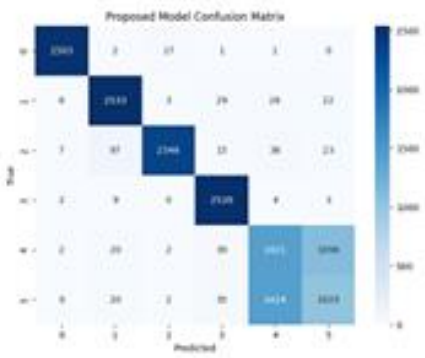


Figure 7: Existing_model_confusion_matrix_KNN Figure 8:Proposed_model_confusion_matrix_KNN

Based on the confusion matrices provided for the existing and proposed models, here is a detailed analysis of the results:

Existing Model (Unbalanced Data)

The confusion matrix of the existing model shows that the dataset is highly imbalanced, with the majority of the samples in the normal (class 0) and attack classes (classes 1 to 4) not well-distributed. Here's a breakdown by each class:

1. **Class 0 (Normal Traffic)**: The model has successfully classified 2525 samples correctly as class 0 but has some misclassifications into other classes, though these are minimal (21 misclassified as class 2, and a few in other classes).
2. **Class 1 to Class 4 (Attack Classes)**: The performance is considerably lower. For example:

generation did not introduce inaccuracies in these cases. This consistency highlights the GAN's ability to supplement rare class samples without degrading model performance.

2. **Improvement with GAN (TC_4, TC_6):**
    - In TC_4, the KNN model initially misidentifies the attack as "DoS Slowhttptest" before GAN enhancement but corrects it to "DoS Slowloris" after GAN enhancement, aligning with the actual output. Similarly, in TC_6, the Random Forest model initially misclassifies "DoS GoldenEye" as "Benign," but after GAN enhancement, it accurately detects "DoS GoldenEye."
    - These results suggest that the GAN-generated samples may have helped the model better distinguish between rare class attack types. This improvement indicates that the synthetic data added by the GAN helps reduce misclassification rates for rare class attacks.
3. **Correct Benign Classification (TC_3, TC_7):**
    - In TC_3 and TC_7, both KNN and Random Forest models correctly identify benign instances both before and after GAN enhancement. There are no discrepancies in detecting benign traffic, which is classified as "Normal Class" with no attack.
    - These results suggest that the GAN does not introduce false positives for benign traffic, maintaining model accuracy for normal classifications even after synthetic data is added.

5. TESTING

5.1 TEST CASES

Table 1 Test Cases

Test Case ID	MODEL	Actual Output	Output before GAN	Output after GAN	Normal/Rare class
TC_1	KNN	DoS GoldenEye	DoS GoldenEye	DoS GoldenEye	Rare Class Attack: Yes
TC_2	KNN	DoS Hulk	DoS Hulk	DoS Hulk	Rare Class Attack: Yes
TC_3	KNN	Benign	Benign	Benign	Normal Class Attack: No
TC_4	KNN	DoS Slowloris	DoS Slowhttptest	DoS Slowloris	Rare Class Attack: Yes
TC_5	Random Forest	DoS GoldenEye	DoS GoldenEye	DoS GoldenEye	Rare Class Attack: Yes
TC_6	Random Forest	DoS GoldenEye	Benign	DoS GoldenEye	Rare Class Attack: Yes
TC_7	Random Forest	Benign	Benign	Benign	Normal Class Attack: No

TEST RESULTS

The test results of this table reveal the impact of using a GAN to enhance rare classes in an intrusion detection system. The cases cover various attacks and benign instances across different models (KNN and Random Forest) and provide insights into the effectiveness of the GAN-enhanced model.

Analysis of Test Cases

1. **Consistency in Detection (TC_1, TC_2, TC_5, TC_6):**
    - In these test cases, both KNN and Random Forest correctly identify the attack type both before and after the GAN enhancement. The outputs match the actual rare class attacks (e.g., "DoS GoldenEye" and "DoS Hulk").
    - The results demonstrate that the GAN-enhanced model successfully retains accurate detection for these attacks, indicating that the synthetic data

4.2.6. Evaluation Metrics

The enhanced IDS is evaluated based on several criteria to measure the effectiveness of GAN augmentation:

- **Recall for Rare Classes:** A higher recall indicates improved detection of rare attack types, signifying a reduction in missed detections.
- **Precision and F1-Score:** By balancing precision with recall, the system's accuracy in identifying rare attacks is evaluated.
- **False Positive and Negative Rates:** Reductions in these rates confirm that the GAN-enhanced model mitigates biases present in the original, unbalanced dataset.

4.2.7. Comparative Analysis

The project includes a comparison with baseline models to demonstrate the effectiveness of GAN augmentation:

- **Baseline Model:** The initial model trained on the unbalanced dataset serves as a reference.
- **Enhanced Model:** The model trained on the GAN-augmented dataset showcases the benefits of balancing data.

Results indicate that the GAN-based system achieves superior performance, especially in detecting rare attack classes. Lower false positives and negatives demonstrate the model's robustness in accurately identifying diverse attack types.

4.2.4. Dataset: CICIDS-2017

The CICIDS-2017 dataset is chosen for its comprehensive representation of real-world network traffic, containing approximately 3 million rows with detailed network flow records. From this dataset we have chosen the data collected on wednesday. It includes benign and malicious data, labeled with various attack types:

- **Benign Data:** Regular network traffic labeled as non-attack.
- **Attack Data Types:** The dataset includes multiple types of Denial of Service (DoS) attacks such as Hulk, Slowloris, GoldenEye, Heartbleed.

Each entry provides metadata such as source/destination IPs, ports, timestamps, and the nature of traffic. By leveraging the GAN to produce additional samples of rare attacks in this dataset, the model is expected to perform robustly against these diverse attack types.

4.2.5. System Architecture

The architecture for this enhanced IDS consists of the following components:

1. **Preprocessing and Data Balancing:** The CICIDS-2017 dataset is initially preprocessed, removing inconsistencies and splitting it into benign and attack classes. Using the GAN model, synthetic samples of rare attack classes are generated, creating an augmented dataset with balanced attack and benign samples.
2. **Model Training with Augmented Data:** KNN, Random Forest and LSVM are trained on the augmented dataset. Initially, models are trained on the unbalanced dataset for baseline comparison. After GAN-based augmentation, models are retrained to evaluate the effect of balanced data on detection accuracy.
3. **Performance Comparison and Evaluation:** To validate the effectiveness of the GAN-based augmentation, the model's performance on the unbalanced dataset is compared with its performance on the balanced dataset. Key performance metrics include accuracy, recall, precision, F1-score, and confusion matrix analysis.
4. **Deployment and Monitoring:** The trained model is then integrated into the IDS. Here, the IDS runs in real time, classifying traffic and detecting anomalies. Rare attack classes are detected with improved recall, lowering false negatives and allowing the system to flag critical threats accurately.

- **Host-based Intrusion Detection Systems (HIDS):** These monitor specific devices by analyzing file changes, traffic logs, and system configurations to detect anomalies.
- **Network-based Intrusion Detection Systems (NIDS):** These examine traffic across the network, spotting suspicious data patterns and alerting administrators. NIDS are strategically placed in network environments for comprehensive monitoring.

In this project, the focus is on Network-based Intrusion Detection Systems, leveraging data patterns across an entire network to recognize rare attacks that might otherwise go undetected.

4.2.2. Problem Statement

IDSs struggle to detect rare attacks accurately due to data imbalance—where benign data vastly outnumber attack data. This leads to overfitting towards common patterns, resulting in missed detection of rare, potentially damaging attack vectors. The challenge is magnified in systems that depend solely on conventional models, as they often fail to generalize well in cases where the occurrence of rare attacks is significantly low. By generating synthetic samples of rare attack data through GANs, this project aims to balance the dataset, enhancing IDS performance in terms of recall, precision, and false-positive reduction.

4.2.3. Proposed System

The solution involves integrating GANs into the IDS to address the class imbalance problem by generating synthetic samples of rare attack data, which are then combined with the existing dataset. The GAN architecture consists of:

- **Generator Network:** This network takes random noise as input and produces data that mimics the characteristics of rare attacks. By training the generator to create increasingly realistic attack patterns, it augments the minority class in the dataset.
- **Discriminator Network:** The discriminator distinguishes between real and synthetic samples, providing feedback to the generator to improve the quality of synthetic data iteratively.

Once the GAN-augmented dataset is created, KNN , RandomForest and LSVM are used to classify network traffic into attack and non-attack categories. This augmented model is expected to perform better, with a higher recall for rare classes and reduced false negatives.

4.1.2.4 Linear Support Vector Machine (LSVM)

Role in GAN: Similar to Random Forest, Linear SVMs aren't traditionally part of the GAN structure but can be employed alongside GANs to classify rare attacks. When GANs generate synthetic attack samples, SVMs can be used as a robust classifier to separate attack instances from normal data. The linear decision boundary of SVM can be beneficial in separating attack vectors, especially in high-dimensional spaces.

Decision Making: Linear SVM is a margin-based classifier that works by finding the hyperplane that best separates two classes (normal and attack). For rare class attack detection, SVM is particularly useful in scenarios where the classes are linearly separable, offering a simple yet effective method for decision-making. Linear SVMs can also handle high-dimensional data effectively, making them suitable for network security datasets with many features.

Enhancing Detection: SVMs perform well even when dealing with rare classes, especially with the right choice of kernel (e.g., linear). They are also resistant to overfitting in high-dimensional spaces, which helps improve detection accuracy. In rare attack detection, SVMs enhance the model's generalization ability, ensuring it does not misclassify new data instances, even when those attacks are underrepresented in the dataset.

4.2 OVERVIEW TECHNOLOGY

The objective of this project is to improve the detection of rare cyberattacks within intrusion detection systems (IDS) by using Generative Adversarial Networks (GANs). By addressing the imbalance between common benign data and rare attack data, the project aims to minimize false negatives and increase the accuracy of detecting uncommon, high-impact cyber threats. The system study covers the existing challenges in IDS, the proposed architecture, dataset details, and performance evaluation strategies.

4.2.1. Overview of Intrusion Detection Systems (IDS)

Intrusion Detection Systems are security mechanisms designed to monitor network or system activities for unusual behavior. They alert administrators to potential security breaches, malware, and policy violations, working as a "cyber watchdog" for networks. IDS systems fall into two main categories:

The GAN-generated attack instances, which aim to resemble real rare attacks, can be fed into a KNN classifier to further enhance the detection process.

Decision Making: KNN makes decisions based on the "neighbourhood" of a given data point, classifying it by considering the majority label of its nearest neighbours. In rare class attack detection, KNN can effectively identify whether a new data instance is similar to known attack patterns or normal behaviour. It is particularly useful when attacks have local structure (e.g., specific patterns in a subset of features), as it focuses on instance-based learning rather than a global model.

Enhancing Detection: KNN is particularly helpful in rare attack detection as it can handle complex and irregular data distributions, which is common in attack scenarios. By adjusting the value of K (number of neighbours), KNN can be fine-tuned to focus more on rare class instances, improving detection performance. Additionally, it can be adapted to deal with imbalanced data by assigning weights to neighbours based on distance, giving more importance to rare attack instances.

4.1.2.3 Random Forest

Role in GAN: Role in GAN: While Random Forests (RFs) are not typically part of the GAN architecture, they can be used as a standalone or supplementary model for rare class attack detection. RFs consist of an ensemble of decision trees that can capture complex relationships in data, providing a robust mechanism to classify rare attacks. When integrated with GAN-generated synthetic data, Random Forests can further improve classification accuracy by learning from a more balanced dataset.

Decision Making: Random Forests excel at handling imbalanced datasets due to their inherent bagging technique and voting mechanism. In rare attack detection, each decision tree learns different attack patterns, and the ensemble improves decision making by combining predictions from individual trees. This ensemble approach helps mitigate the risk of overfitting to the minority class, thus enhancing rare attack detection.

Enhancing Detection: RFs are highly interpretable and can capture non-linear relationships between features, which can be crucial in distinguishing between normal and rare attack patterns. Additionally, RF's ability to rank features (feature importance) helps highlight which aspects of network traffic or system logs are most indicative of an attack, thereby boosting detection capabilities.

- **Structure**: It often includes layers that upsample the input, with convolutional layers used in GANs for image data, enabling the generation of high-dimensional, detailed outputs.

2. Discriminator Network

The **Discriminator** (D) serves as a classifier, distinguishing between real data (from the original dataset) and fake data (produced by the generator). It takes both real and generated samples and learns to label them as either real (1) or fake (0).

- **Objective**: The discriminator's task is to accurately differentiate between real and generated data.
- **Training**: The discriminator is rewarded when it correctly identifies real vs. fake samples, and penalized when it fails. This feedback helps it become better at recognizing authentic data and identifying fake samples.
- **Structure**: The discriminator usually uses layers suitable for feature extraction, such as convolutional layers for image data, as it must learn to identify patterns that differentiate real and generated samples.

Adversarial Process

The GAN training process is adversarial: the generator aims to "fool" the discriminator, while the discriminator strives to correctly classify samples. This competition pushes both networks to improve until the generator creates data so realistic that the discriminator struggles to tell them apart.

In summary, the generator learns to mimic the data distribution, and the discriminator refines its ability to detect authenticity, resulting in high-quality synthetic data generated by the GAN.

4.1.2.2 K Nearest Neighbours Algorithm (KNN)

Role in GAN: While KNN is not directly used within GANs, it can be an effective post-processing tool for classifying data generated by GANs. KNN can be used to classify both real and synthetic attack data by evaluating its proximity to known data points in the feature space.

training ML models but also advances the field of intrusion detection by focusing on the reliable detection of minority attack classes, ultimately contributing to a more secure network environment.

4.1.2 MODELS

In our project on enhancing rare class attack detection using machine learning, we employed a range of models to optimize the detection of network intrusions and improve classification accuracy. The following is a comparative analysis of the models used:

4.1.2.1 Generative Adversarial Network(GAN)

A **Generative Adversarial Network (GAN)** is a type of deep learning model that generates new, synthetic data resembling a given dataset. Introduced by Ian Goodfellow in 2014, GANs consist of two main neural networks, a **generator** and a **discriminator**, which work against each other in a process called adversarial training. The interplay between these two networks leads to a refined generator that can create highly realistic data after several training iterations. GANs are commonly used in image generation, text synthesis, and enhancing rare class detection in areas like intrusion detection and medical imaging.

1. Generator Network

The **Generator** (G) is responsible for creating synthetic data, aiming to generate samples that resemble the real dataset. It starts by taking a random input, typically a noise vector zzz, and transforms it through a series of layers to produce an output similar to the target data (e.g., an image or sequence).

- **Objective**: The generator's goal is to create data that appears realistic enough to deceive the discriminator.
- **Training**: The generator's loss is derived from the discriminator's feedback. The generator is penalized when the discriminator correctly identifies the generated data as fake, prompting it to improve its output with each iteration.

5. **Heartbleed Port 444 (15:12 - 15:32 p.m.)**

   Heartbleed is a security vulnerability in the OpenSSL cryptography library carried out through the TLS heartbeat extension. Google Security first discovered this bug in 2014. However, it still is a security threat to many businesses and organizations. So it is crucial to understand how Heartbleed exploits vulnerable OpenSSL versions, its potential harm, and ways to protect your networks.

Relevance of the Dataset to the Project

The selected subset from the CICIDS-2017 dataset is integral to our project due to the nature of the attacks and their frequency within the dataset. Intrusion detection systems (IDS) often struggle with recognizing minority class attacks due to their rarity in typical network environments, which is a major limitation we aim to address. With the help of GAN models, we aim to generate synthetic samples of the minority class attacks in the dataset, specifically the DoS/DDoS types mentioned above. By augmenting these minority classes, we can create a balanced dataset that enhances the learning capability of conventional machine learning models, ultimately improving the IDS's detection accuracy for rare attack types.

Dataset Characteristics for Rare Class Augmentation

The CICIDS-2017 dataset includes detailed feature representation, which is essential for applying machine learning models that require nuanced patterns for classification. Given the imbalanced nature of this dataset, particularly in terms of rare attacks like Slowloris and Slowhttptest, GAN models play a crucial role. By training a GAN on this dataset's specific characteristics, our project leverages synthetic data generation to produce realistic, minor-class data that retains the complex, low-frequency patterns of these DoS attacks. This synthetic augmentation enhances the detection capabilities of the IDS, specifically for attacks that may otherwise go undetected due to insufficient training data.

In conclusion, the CICIDS-2017 dataset provides an ideal foundation for evaluating and enhancing intrusion detection methods due to its extensive attack coverage and realistic traffic simulation. By focusing on the Wednesday DoS/DDoS traffic subset, our project addresses the class imbalance issue that hinders traditional IDS performance, using GANs to generate robust, synthetic data for rare attacks. This approach not only improves the dataset's usability for

Subset Selection and Attack Patterns in Focus

In our project, we focus on data captured specifically on Wednesday of the dataset collection, which was designated for DoS and DDoS attack traffic. This subset includes four distinct DoS attack types: Slowloris, Slowhttptest, Hulk, and GoldenEye. These attacks are known for their unique intrusion techniques, each targeting different vulnerabilities in network infrastructure. Each of these attacks was recorded at specific time intervals, allowing us to isolate and analyze them separately. We believe that focusing on this subset of the CICIDS-2017 dataset provides the optimal context to evaluate our GAN model's effectiveness in augmenting and detecting rare attack instances, which is vital for the improvement of intrusion detection systems.

1. **DoS Slowloris (9:47 – 10:10 a.m.)**
   The Slowloris attack attempts to keep many connections open to the server by sending partial HTTP requests, forcing the server to maintain these open connections for prolonged periods and eventually exhausting its resources. This form of attack is complex to detect due to the minimal bandwidth it requires, which makes it appear like legitimate traffic under conventional detection methods.
2. **DoS Slowhttptest (10:14 – 10:35 a.m.)**
   Similar to Slowloris, Slowhttptest aims to exhaust server resources by sending partial requests. However, it employs different techniques to target application-layer vulnerabilities, making it challenging for traditional IDS to distinguish between benign and malicious traffic. The challenge with Slowhttptest data lies in its low traffic volume, which limits the availability of detectable patterns.
3. **DoS Hulk (10:43 – 11 a.m.)**
   Unlike Slowloris and Slowhttptest, Hulk is an aggressive DoS attack that bombards the server with large volumes of HTTP requests, overwhelming the server's capacity. This high-traffic attack pattern is easier to detect due to its volume but can be challenging to classify correctly in a highly imbalanced dataset where such aggressive attacks are rare.
4. **DoS GoldenEye (11:10 – 11:23 a.m.)**
   The GoldenEye attack targets vulnerabilities similar to those exploited by the Hulk attack but with a different payload structure. This variant demonstrates that even slightly altered patterns can produce unique traffic signatures, posing additional challenges in detecting minor classes of attacks.

4. IMPLEMENTATION

4.1. MODULES

4.1.1 DATASET:

The CICIDS-2017 dataset, developed by the Canadian Institute for Cybersecurity at the University of New Brunswick, is a comprehensive dataset specifically crafted to address current challenges in network intrusion detection research. This dataset is widely adopted in cybersecurity research due to its inclusion of diverse network traffic, reflecting realistic and modern-day attack patterns. For our project on "Enhancing Rare Class Detection Using GAN Model in Intrusion Detection Systems," we have selected a subset of this dataset with a targeted focus on rare and sophisticated attack types, specifically Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks. The CICIDS-2017 dataset's structure and data characteristics enable us to simulate realistic scenarios in detecting rare attack patterns, providing a valuable foundation for addressing the class imbalance issue typical in intrusion detection datasets.

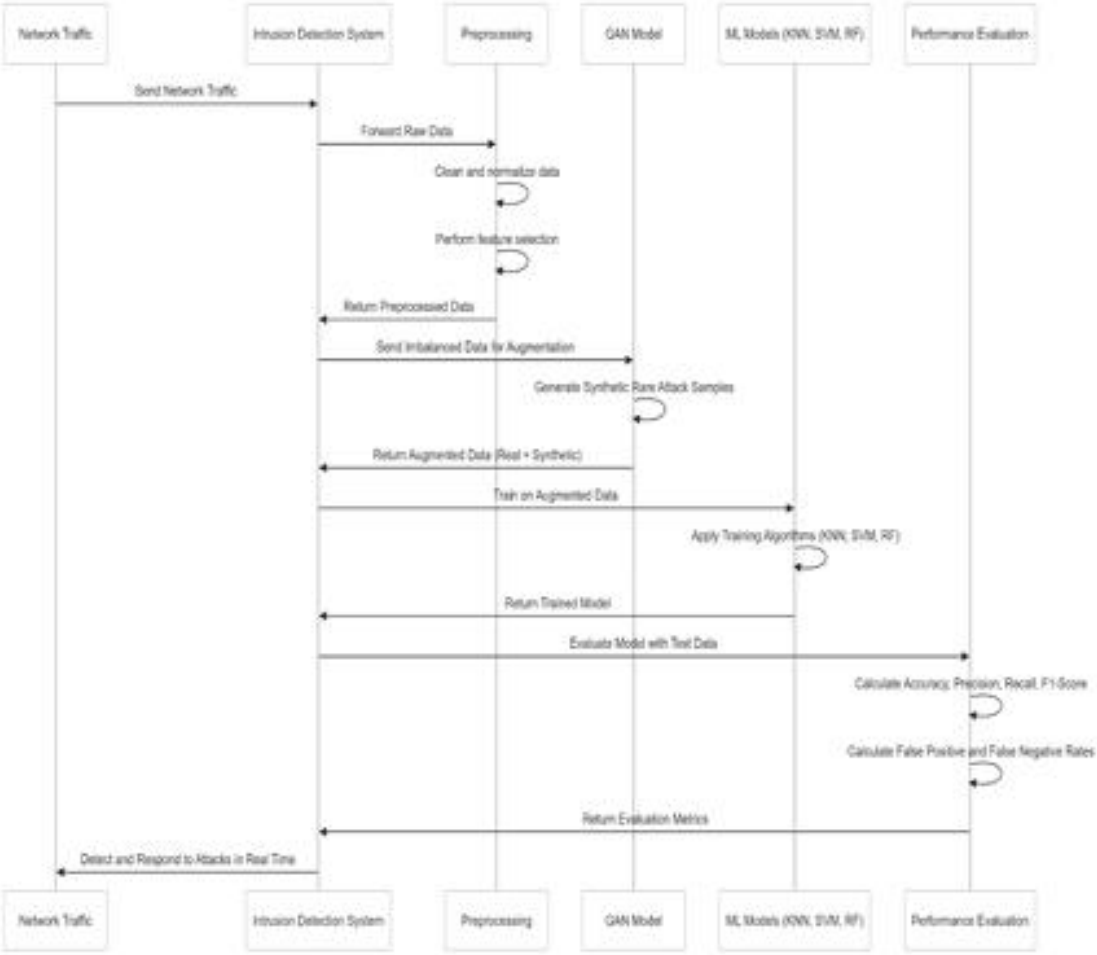
Dataset Overview and Significance for Intrusion Detection

The CICIDS-2017 dataset is designed to represent a real-world network traffic environment, covering a wide range of attack types and benign network behavior. It includes seven days of network traffic, each dedicated to different types of malicious and benign activities. This dataset is distinguished by its detailed traffic records, covering 80 features for each instance. The features include a range of network parameters such as flow duration, packet size, inter-packet intervals, and other critical metrics, all of which help in identifying unique network activity patterns for distinguishing between legitimate and malicious traffic. By capturing specific characteristics of traffic flow, CICIDS-2017 provides rich and extensive data for analyzing various cyber attacks, including the rare DoS and DDoS patterns we focus on.

- **Model**: Various machine learning models (KNN, SVM, Random Forest) are trained on the augmented dataset.
- **Evaluator**: Calculates the performance metrics, providing insights into the accuracy and reliability of the model in detecting rare attack classes.
- **IDS**: Integrates the trained model to detect attacks in real-time based on incoming network traffic.

- **Alert System to Security Analyst**: The analyst receives notifications from the Alert System for further investigation.

3.2.5. Sequence Diagram:



Explanation of the Sequence Diagram:

- **Preprocessor**: Cleans and normalizes the raw data, performs feature selection, and forwards the preprocessed data back to the IDS.
- **GAN**: Generates synthetic samples of rare attack data, combining them with the real data to form a balanced dataset.

Explanation of the Component Diagram

1. **Components**:
    - **Intrusion Detection System (IDS)**: The central component that orchestrates the flow of data through preprocessing, synthetic data generation, model training, and evaluation.
    - **Data Preprocessing Module**: Handles data cleaning, normalization, feature selection, and preparation for training.
    - **GAN Module**: Generates synthetic data samples for rare classes to balance the dataset.
    - **Machine Learning Models (MLModels)**: This includes various ML models (e.g., KNN, SVM, Random Forest) that are trained on the balanced dataset.
    - **Evaluation Module (Eval)**: Calculates performance metrics to assess the effectiveness of trained models.
    - **Network Data Source**: Provides raw network data for IDS analysis.
    - **Alert System**: Triggers and sends alerts to the Security Analyst upon rare attack detection.
    - **Security Analyst**: Receives alerts and reviews logs for potential threats.
2. **Relationships**:
    - **Network Data Source to IDS**: Network traffic is sent to the IDS for analysis.
    - **IDS to Data Preprocessing Module**: The IDS forwards raw data to the preprocessing module.
    - **Data Preprocessing to GAN Module**: If data imbalance is detected, the preprocessing module sends data to the GAN module to generate synthetic samples.
    - **GAN Module to Data Preprocessing Module**: The GAN returns synthetic data, which is combined with the original dataset.
    - **Data Preprocessing to Machine Learning Models**: The balanced dataset is sent to machine learning models for training.
    - **Machine Learning Models to Evaluation Module**: Trained models provide predictions to the evaluation module for metric calculations.
    - **IDS to Alert System**: If a rare attack is detected, the IDS sends an alert to the Alert System.

- **Evaluation**: This class calculates evaluation metrics and generates performance reports.

2. **Relationships**:
   - **IDS uses Preprocessor**: The IDS system uses the Preprocessor class to prepare data.
   - **IDS utilizes GAN**: The IDS relies on GAN to create synthetic data.
   - **IDS trains MachineLearningModel**: The IDS uses various machine learning models for training on the processed data.
   - **IDS assesses Evaluation**: The IDS system uses the Evaluation class to assess model performance.
   - **Inheritance**: KNN, SVM, and RandomForest inherit from Machine learning model , reflecting their specialized roles in training and predictions.

3.2.4 Component diagram:

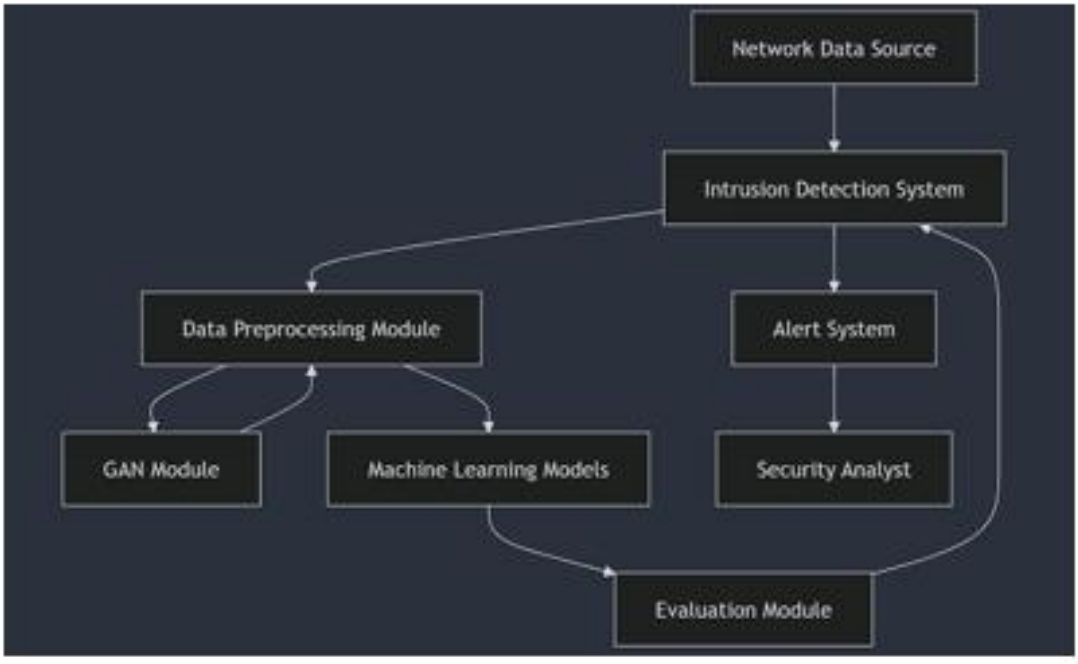


Figure 5 :Component Diagram

3.2.3. Class Diagram

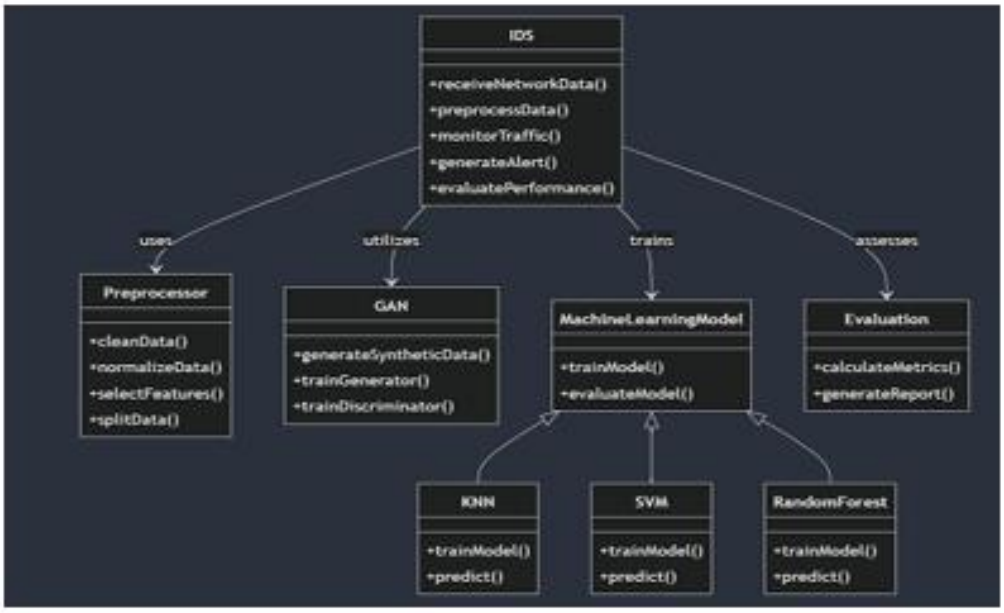


Figure 4 :Class Diagram

Explanation of the Class Diagram

1. **Classes**:
    - **IDS**: This is the main class representing the Intrusion Detection System. It has methods for receiving network data, preprocessing, monitoring traffic, generating alerts, and evaluating performance.
    - **Preprocessor**: This class handles data cleaning, normalization, feature selection, and splitting data into training and testing sets.
    - **GAN**: The GAN class generates synthetic data samples for rare attacks. It contains methods for training the generator and discriminator networks.
    - **MachineLearningModel**: This is an abstract class representing the machine learning model, with methods to train and evaluate models.
    - **KNN, SVM, RandomForest**: These subclasses inherit from the Machine Learning Model class, each representing a different type of machine learning model used in the IDS.

Explanation of the Diagram

- **Start**: The process begins when network traffic is received by the IDS.
- **Data Preprocessing**: Data cleaning, feature selection, and normalization steps are performed to prepare data for model training.
- **Data Imbalance Check**: Checks if there is an imbalance in the dataset (e.g., too few samples for rare attacks).
    - **If Imbalance Exists**: Synthetic data is generated using a GAN model to balance the dataset.
    - **If No Imbalance**: The existing data is used without augmentation.
- **Data Combination**: The real and synthetic data are combined to create a balanced dataset.
- **Split Data**: Data is split into training and testing sets.
- **Model Training**: Machine learning models (KNN, SVM, Random Forest) are trained on the training data.
- **Evaluation**: The model's performance is evaluated on the test data using metrics such as accuracy, precision, recall, F1-score, and false positive/negative rates.
- **Report Generation**: Generates a performance report for analysis.
- **Deployment Decision**: Decides whether to deploy the trained model.
    - **If Yes**: The model is deployed for real-time monitoring.
- **Real-Time Monitoring**: The deployed model monitors network traffic.
- **Attack Detection and Alerts**: If a rare attack is detected, an alert is generated, and the Security Analyst is notified.
- **Review Alert Logs**: The Security Analyst reviews logs and assesses the detected threats.
- **End**: The process completes once alerts are reviewed.

This activity diagram captures the flow of data preprocessing, GAN-based augmentation, model training, evaluation, and real-time deployment in the IDS for enhanced rare attack detection.

3.2.2. Activity Diagram

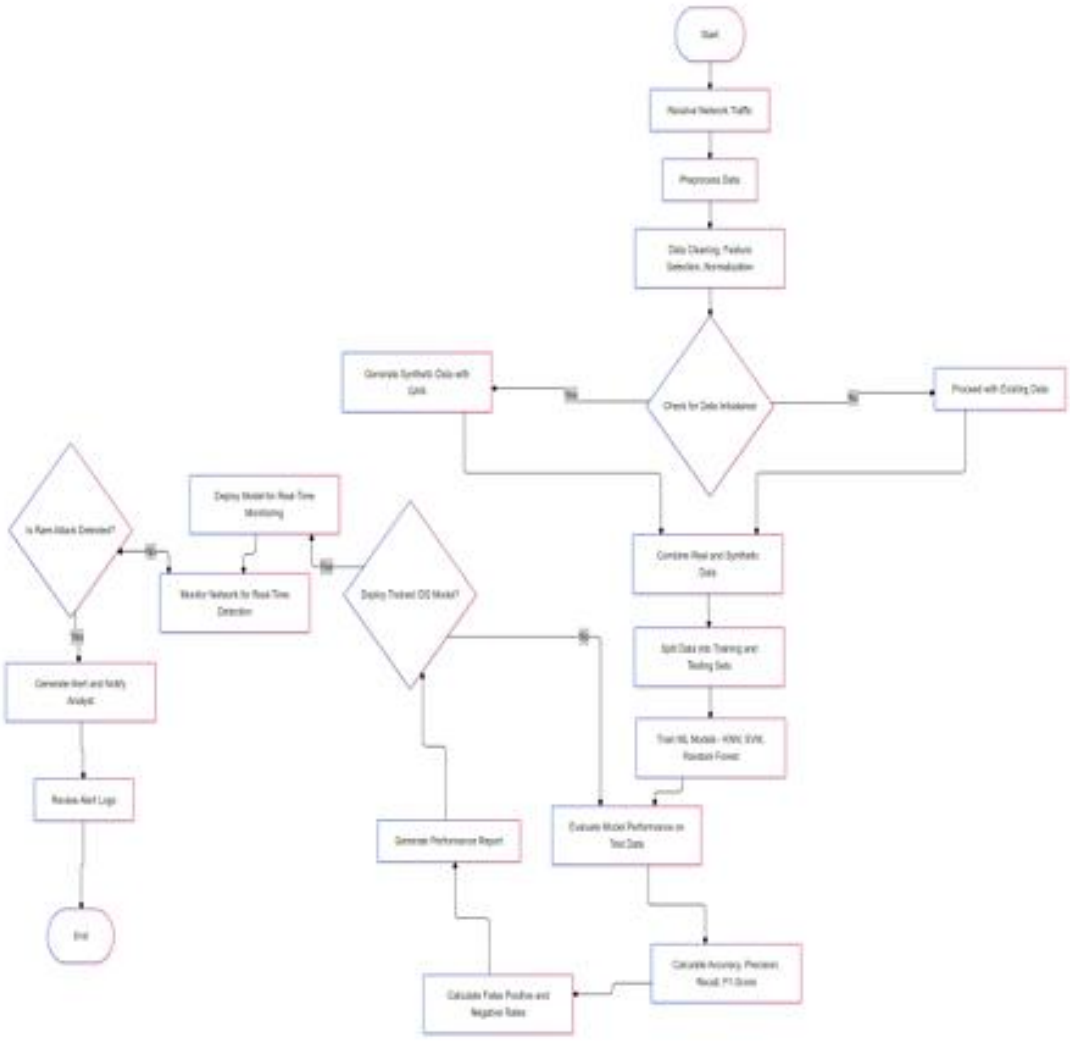


Figure 3 Activity Diagram

Explanation of the Diagram

- **Actors**:
  - **Administrator**: Configures IDS settings and deploys the trained IDS model.
  - **Network Traffic**: Represents the incoming network data being monitored by the system.
  - **Security Analyst**: Monitors and evaluates model performance and reviews alert logs.
- **Use Cases**:
  - **Configure IDS Settings**: Allows the administrator to set up initial IDS configurations.
  - **Send Network Data**: Network data is sent to the IDS for analysis.
  - **Preprocess Data**: Raw data is cleaned, normalized, and prepared for modeling.
  - **Generate Synthetic Data with GAN**: A GAN model creates synthetic data for rare attack classes.
  - **Augment Dataset**: Combines real and synthetic data for training.
  - **Train Machine Learning Models**: Machine learning models (KNN, SVM, Random Forest) are trained on the augmented dataset.
  - **Evaluate Model Performance**: The model's performance is evaluated using metrics.
  - **Generate Evaluation Report**: Summarizes model performance.
  - **Deploy IDS Model**: Deploys the trained model for real-time network monitoring.
  - **Monitor Network for Attacks**: IDS monitors network traffic to detect suspicious activity.
  - **Alert on Rare Attack Detection**: The system alerts the analyst when a rare attack is detected.
  - **Review Alert Logs**: The analyst reviews alerts and logs for further action.

This flowchart-style use case diagram captures the roles and interactions between the actors and main functionalities within the system.

3.1.2.3 RAM:

- Adequate RAM, typically 16 GB or more, is essential for handling large datasets and training complex models.

3.1.2.4 Storage:

- Sufficient storage space for datasets and model checkpoints. SSDs are preferable for faster data access.

3.1.2.5 Internet Connection:

- A reliable internet connection is necessary for downloading datasets, libraries, and collaborating with team members if applicable.

3.2 UML DIAGRAMS

3.2.1. Use Case Diagram

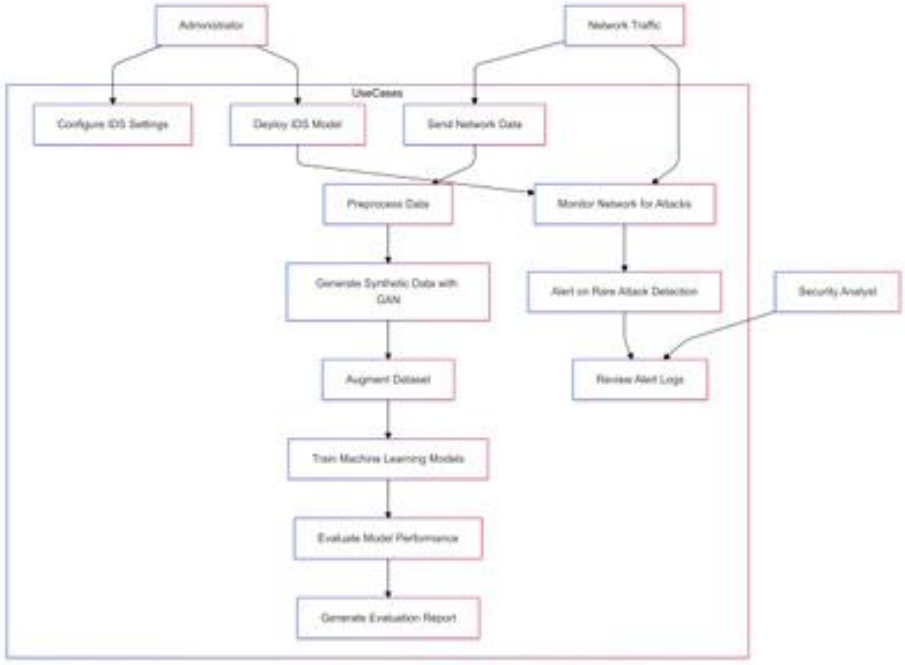


Figure 2 Use Case Diagram

3. DESIGN

3.1 SYSTEM REQUIREMENTS

3.1.1 Software Requirements:

3.1.1.1 Programming Languages:

- **Python:** Python is the predominant language for machine learning and data analysis. You'll need Python for implementing your ML models, data preprocessing, and analysis.

3.1.1.2 Machine Learning Libraries:

- **TensorFlow or PyTorch:** These deep learning frameworks are necessary if you plan to work with neural networks, in working with balanced and unbalanced dataset.

3.1.1.3 Data Analysis and Visualization:

- **Pandas and Numpy:** For data manipulation and analysis.
- **Matplotlib and Seaborn:** For data visualization.

3.1.1.4 Jupyter Notebook:

- Jupyter Notebook are great for interactive data exploration and model development.

3.1.2 Hardware Requirements:

3.1.2.1 Computing Power:

- A reasonably powerful computer with a multi-core CPU is necessary for training ML models, especially if you're working with large datasets or complex deep learning models.

3.1.2.2 GPU :

- Having access to a GPU (Graphics Processing Unit) can significantly speed up training deep learning models.

real and synthetic data, the IDS models gain a more balanced view, improving sensitivity to rare attacks.

The project employs several machine learning algorithms to find the most effective approach for rare class detection. These include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest, each providing unique strengths in handling outlier detection, binary classification, and generalization, respectively. Performance metrics such as accuracy, precision, recall, F1-score, False Positive Rate (FPR), and False Negative Rate (FNR) are used to evaluate each model's effectiveness. Improvements in FPR and FNR indicate the system's enhanced reliability, showing fewer false alarms and missed detections.

This GAN-enhanced IDS approach promises improved detection rates for rare classes, creating a more robust system capable of identifying both common and uncommon threats in real-world networks. By reducing vulnerability to underrepresented cyber threats, this project aims to strengthen an organization's overall security. Future work may explore deploying this system across various network environments and refining the GAN architecture to further boost detection capabilities.

random forests to achieve best results. The paper proposes a novel intrusion detection method, integrating hybrid sampling and deep hierarchical networks. Utilizing OSS and SMOTE, it constructs a balanced dataset for efficient model training. Evaluation on NSL-KDD and UNSW-NB15 datasets shows superior performance compared to other classifiers in accuracy, precision, and recall rate.

2.1.10 A Deep Learning Approach for Network Intrusion Detection System [10]

Proper feature selections from the network traffic dataset for anomaly detection is difficult,Unavailability of labelled traffic dataset from real networks for developing an Network Intrusion Detection System(NIDS).Developing a flexible and efficient NIDS for unforeseen and unpredictable attacks,to show high detection accuracy with less false-alarm rates.Methods used are Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naive-Bayesian (NB), Random Forests (RF), and Self-Organized Maps (SOM). This paper proposed a deep learning approach for developing a Network Intrusion Detection System (NIDS), utilizing sparse autoencoder and softmax regression, and evaluated its performance on the NSL-KDD dataset, achieving notable accuracy in anomaly detection.

2.2 SYSTEM STUDY

This project aims to enhance the detection of rare attack classes in Intrusion Detection Systems (IDS) using the CICIDS 2017 dataset, with a focus on Wednesday's network traffic data, which includes a mix of both benign and rare malicious activities. Due to the sparse occurrence of these attacks amidst predominantly normal traffic, detecting rare attack patterns can be challenging. To address this, the project involves data preprocessing, GAN-based data augmentation, and multiple machine learning models.

The data preprocessing phase is essential to prepare the CICIDS 2017 dataset for model training. This involves data cleaning, feature selection, and normalization, which together ensure the dataset is high-quality and machine-learning-ready. The data is then split into training and testing sets to enable model evaluation on unseen data. Given the class imbalance problem—where benign samples far outnumber rare attack samples—the project utilizes Generative Adversarial Networks (GANs) for data augmentation. By generating synthetic data points resembling rare attack instances, GANs help to balance the dataset. This augmentation process involves two networks: the generator, which learns to create synthetic rare attack samples, and the discriminator, which evaluates the samples' authenticity. By training on both

conventional and state-of-the-art techniques on the NSL-KDD dataset. However, lacking detailed explanation of statistical methods, limited dataset representation, and absence of real-world testing raise concerns about generalizability and effectiveness in practical cybersecurity scenarios.The paper focuses on enhancing intrusion detection accuracy with DLNID, utilizing Bi-LSTM, attention mechanism, ADASYN for imbalanced data, and a modified stacked autoencoder for dimensionality reduction. Achieves superior accuracy (90.73%) and F1 score (89.65%) on NSL-KDD dataset.The paper's outlines delineates DLNID, leveraging ADASYN for data imbalance, stacked autoencoder for downscaling, and channel attention with Bi-LSTM for network structure, achieving superior performance on KDDTest+. DLNID promises improved classification for network intrusion detection.

2.1.8 GIDS: GAN BASED INTRUSION DETECTION SYSTEM FOR IN-VEHICLE NETWORK [8]

As V2X technology enables interactions with vehicles and everything from outside such as Vehicles, Infrastructure, security threats on Electronic Control Units(ECU) of vehicles become higher.Methods used are Generative Adversarial Nets (GAN), Design of Neural Networks,Convolutional NeuralNetwork (CNN), Deep Neural Network (DNN),Evaluating One-hot-vector Encoding.Develop a security system to mitigate the various risks of the vehicle using Intrusion Detection System (IDS) for in-vehicle network is required to protect all of the ECUs and y to detect and respond known and unknown attacks of today.Intrusion Detection System (IDS) enhances performance,enables detection of unknown attacks solely with normal data,achieves high accuracies,demonstrating its potential as a real-time, scalable, and secure IDS solution for diverse vehicle types.

2.1.9 Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network [9]

The proposed method faces challenges including limited generalization beyond NSL-KDD and UNSW-NB15 datasets, potentially inadequate accuracy for robust intrusion detection, risks of complexity and overfitting with deep hierarchical networks, interpretability issues, and computational resource demands.The paper aims to mitigate data imbalance in intrusion detection by proposing a hybrid sampling method integrated with a deep hierarchical network. This approach aims to create a balanced dataset, facilitating improved model learning and reducing training time. The main focus of the paper is to propose a hybrid intrusion detection method, KNN-RF, aimed at improving the detection rates of abnormal behaviors, including Probe, U2R, and R2L attacks. This method combines KNN outlier detection and multilevel

(ADASYN) for data balancing, the Multi-Layer Perceptron (MLP) classifier achieves high accuracy and F1 scores, outperforming previous algorithms.The paper presents a method using deep learning techniques to enhance intrusion detection systems (IDS) for recognizing and blocking unauthorized access. It focuses on improving detection rates for specific attacks like R2L and U2R, achieving high accuracy, and outperforming previous models. Future research aims to address remaining challenges and enhance interpretability.

2.1.6 A Deep Learning Model for Network Intrusion Detection with Imbalanced Data [6]

The paper lacks detailed analysis of the dataset's representativeness, potential biases, and generalizability. It doesn't thoroughly discuss the impact of data augmentation methods on performance, nor does it address the potential limitations of the chosen algorithms. Additionally, there's a lack of comparative analysis with other approaches, potentially limiting the paper's applicability and robustness.The objective of this paper is to address the low detection accuracy in network intrusion detection systems by proposing a deep learning model named DLNID (Deep Learning Model for Network Intrusion Detection). It combines an attention mechanism and Bi-LSTM (Bidirectional Long Short-Term Memory) network, utilizes ADASYN (Adaptive Synthetic Sampling) for imbalanced data, and achieves superior accuracy (90.73%) and F1 score (89.65%) on NSL-KDD dataset.The paper proposes DLNID (Deep Learning Model for Network Intrusion Detection), utilizing Bi-LSTM and attention mechanism, with ADASYN for imbalanced data. Achieves 90.73% accuracy and 89.65% F1 score on NSL-KDD dataset.The paper presents DLNID (Deep Learning Model for Network Intrusion Detection), employing ADASYN for data imbalance, stacked autoencoder for downscaling, and channel attention with Bi-LSTM for network structure. Achieves 90.73% accuracy and 89.65% F1 score, outperforming other models, promising for network intrusion detection development.

2.1.7 A novel statistical analysis and autoencoder driven intelligent intrusion detection approach [7]

The methodolog-y introduces a statistical analysis and autoencoder-driven intrusion detection

the challenges and limitations faced in integrating machine learning into IIoT security mechanisms, particularly in addressing imbalanced dataset problems.The objective of this paper is to investigate the necessity of integrating machine learning into the security mechanisms of the Industrial Internet of Things (IIoT). It aims to identify the reasons for this integration, assess the current shortcomings in machine learning performance for IIoT security, and explore the challenges and real-world considerations associated with this integration.The study employs an experimental approach to address the objectives. It involves analyzing the fundamental differences between IIoT and regular IT networks, examining the vulnerabilities and security requirements specific to IIoT systems. The paper then investigates the performance of machine learning algorithms in meeting these security needs. An IIoT testbed, mimicking a real industrial plant, is utilized as a proof of concept to demonstrate the feasibility and effectiveness of integrating machine learning into IIoT security mechanisms.The cyber-security of the IIoT devices is critical. Intrusion detection is the main security concern in these applications. Machine learning solutions and big data analytics have been widely used to ensure a secure platform in these systems. However, when it comes to a real-world scenario and applying these algorithms practically, they sometimes fall short. The main focus of this paper was studying imbalanced dataset problems and show in which extend the machine learning algorithms are able to help.

2.1.5 Machine Learning-Based Adaptive Synthetic Sampling Technique for Intrusion Detection [5]

The approach presents advancements in enhancing intrusion detection systems (IDS) using deep learning. However, issues include dataset biases, limited attack type coverage, overfitting risks, interpretability challenges, scalability concerns, susceptibility to adversarial attacks, and complexities in multi-class classification extension. Resolving these is vital for robust network security solutions.The paper presents a novel deep learning-based intrusion detection system tailored for IoT networks, aiming to overcome the shortcomings of traditional security methods. By combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and attention mechanisms, the model effectively detects traffic anomalies. Adaptive synthetic sampling (ADASYN) is utilized to address imbalanced data issues. Results show strong performance metrics such as accuracy, precision, recall, and F1 scores, along with high Area Under the Curve (AUC) values.The paper proposes a deep learning-based intrusion detection system for IoT networks, integrating Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), and attention mechanisms. Utilizing adaptive synthetic sampling

of the proposed approach.This paper introduces the GAN-CNN-BiLSTM model to improve intrusion detection. GAN expands the abnormal class size, while CNN-BiLSTM handles feature extraction and classification. Evaluation on the CICIDS 2017 dataset shows GAN's effectiveness in addressing class imbalance, with CNN-BiLSTM outperforming other models in accuracy. Future work will focus on enhancing the model with larger and diverse datasets, optimizing intrusion detection, and validating its real-world applicability. The GAN-CNN-BiLSTM model strengthens intrusion detection, aiding in identifying and defending against network attacks.

2.1.3 GAN-based imbalanced data intrusion detection system [3]

The reliance on GANs to mitigate data imbalance in network intrusion detection may introduce challenges. Synthetic data quality, architecture choice, and integration into existing systems pose potential issues. Thorough validation and consideration of computational overhead and detection accuracy are crucial for real-world applicability and reliability.The main goal of the study is to tackle the problem of imbalanced data in network intrusion detection. Data imbalance occurs when some classes of data (e.g., different types of network attacks) are significantly underrepresented compared to others, leading to biased models that perform poorly on less frequent classes.The study utilizes GANs, which are typically known for their ability to generate realistic synthetic data. This approach allows the model to create additional examples of underrepresented classes, thereby balancing the training dataset. After generating the balanced dataset using GANs, the study employs a Random Forest classifier to evaluate the detection performance. Random Forest is chosen likely due to its robustness and effectiveness in handling varied types of data and its capability to reduce overfitting, which is crucial in a scenario where synthetic data is used.This study demonstrates the potential of using GANs to mitigate the effects of data imbalance in network intrusion detection systems. By generating synthetic data to balance the dataset, the proposed method enhances the learning process, resulting in more accurate and reliable detection capabilities. This could be particularly valuable in cybersecurity, where detecting rare but dangerous threats is crucial.

2.1.4 Effect of Imbalanced Datasets on Security of Industrial IoT Using Machine Learning [4]

The research gap lies in the discrepancy between the theoretical effectiveness of machine learning for IIoT security and its practical application. While machine learning and big data analytics are touted as solutions for IIoT security, real-world implementation often falls short, particularly in handling imbalanced datasets. The study aims to bridge this gap by investigating

2. LITERATURE SURVEY

2.1 RELATED WORK

We referenced 10 distinct research papers on implementing GAN in Network Intrusion Detection Systems(NIDS).

2.1.1 Network Intrusion Detection method based on GAN Model [1]

The existing network intrusion detection methods have less label samples in the training process, and the detection accuracy is not high.In this research paper, existing network intrusion detection methods have less labels for training, to overcome this issue, here this paper implements the GAN Model by using the Adversarial idea contained in GAN. This model enhances the original training dataset by generating samples for expanding the original label sample set.The research paper employs a Generative Adversarial Network (GAN) model to address limited labelled data in network intrusion detection. GAN generates additional samples to expand the label sample set. Experimental results demonstrate superior detection accuracy and performance over other methods, proving its effectiveness in distinguishing false data and its practical applicability.The experimental results show that the proposed GAN model is better than other methods in terms of detection, accuracy and other performance indicators. It is proved that the method proposed in this paper has better Network Intrusion Detection (NID) ability, can effectively detect false data and distinguish it from real data, and has good practicability and generalisation ability.

2.1.2 A Method for Network Intrusion Detection Based on GAN-CNN-BiLSTM [2]

The problem is due to large amount of data learning, the performance will be degraded due to data imbalance. In view of the serious imbalance of network traffic data sets at present, this paper proposes to process data expansion with GAN to solve data imbalance and detect network intrusion in combination with CNN and BiLSTM.This study aims to enhance network intrusion detection by addressing data imbalance using Generative Adversarial Networks (GAN). Deep learning methods, though accurate, suffer from performance degradation with imbalanced data. The objective is to combine GAN with Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks for improved detection efficiency.The method involves preprocessing imbalanced network traffic data with GAN to generate synthetic data, mitigating imbalance issues. A hybrid GAN-CNN-BiLSTM model is trained and tested using the processed data. Performance evaluation against traditional models like SVM, DBN, CNN, and BiLSTM showcases the superior efficiency and accuracy

model with sufficient exposure to both rare and common events, enabling it to accurately identify rare classes in real-world scenarios.

Finally, the trained model is evaluated on the testing dataset to measure its performance. This testing phase is critical to confirm that the model is not only proficient at detecting rare events within the training set but also generalizes well to new data. By assessing accuracy, precision, and recall, the process aims to validate the model's ability to detect rare but significant events accurately.

In summary, this process enhances the detection of rare classes by using GANs to address data imbalance and training ML models on a more representative dataset. This method improves the model's sensitivity to rare events, ultimately increasing its effectiveness in real-world applications where detecting these events is crucial.

The process of enhancing rare class detection in a dataset using Generative Adversarial Networks (GANs) and machine learning (ML) models is a structured approach aimed at overcoming the common data imbalance challenge in many applications, such as Intrusion Detection Systems (IDS). Typically, rare events in datasets are underrepresented, which limits the model's ability to learn to identify these infrequent but often critical events. This process uses GANs to generate synthetic samples of the rare class, thus balancing the training data and improving model performance.

The process begins with the normalization of the dataset. Normalization is essential to ensure that all data features have a consistent scale, preventing any single feature from disproportionately influencing the model. Once normalized, the dataset is divided into a training set and a testing set. This division allows for an effective evaluation of model performance on unseen data, ensuring that the model is not merely memorizing the training data but generalizing well to new instances.

Within the training set, the data is analyzed to separate the rare class samples from the majority (non-rare) class samples. The rare class samples, representing the underrepresented events that we aim to detect, are isolated for focused enhancement. This separation is critical, as it allows the GAN to concentrate on generating synthetic samples that specifically resemble the rare class. Meanwhile, the remaining data, consisting of more frequent or "normal" events, is designated as the non-rare class.

Next, the GAN model is employed to generate synthetic data for the rare class. GANs are a powerful tool for this purpose because they consist of two adversarial networks—the generator and the discriminator—that work together to create realistic data. The generator produces new data samples, while the discriminator evaluates whether each sample is real or synthetic, prompting the generator to improve iteratively. This adversarial process ensures that the generated rare class samples are highly realistic and resemble actual rare class instances, effectively "filling in" the training data where genuine rare samples are lacking.

With the synthetic rare samples in place, the training set is now balanced, containing both the original data and newly generated rare class data. This balanced training set is then used to train various ML models like Random Forest , KNN , LSVM. The goal here is to equip the

Drawbacks of Existing System:

However, these traditional approaches face significant challenges when it comes to detecting rare attacks:

1. **Data Imbalance**: In most real-world scenarios, the number of normal instances far outweighs the number of attack instances, particularly for rare types of attacks. This imbalance leads to IDS models being biased toward the majority class (normal behavior), resulting in poor detection rates for the rare but critical attacks.

2. **Limited Generalization**: Traditional IDS methods, particularly those relying on signatures, have limited ability to generalize to new or previously unseen attack patterns. Anomaly detection methods, while more adaptable, often struggle with distinguishing between rare attacks and benign anomalies.

3. **High False Positive Rate**: Anomaly-based systems tend to flag unusual behavior as potential threats, which can lead to a high rate of false positives. This creates an additional burden on cybersecurity teams who must shift through numerous alerts to identify genuine threats.

1.2 PROPOSED SYSTEM:

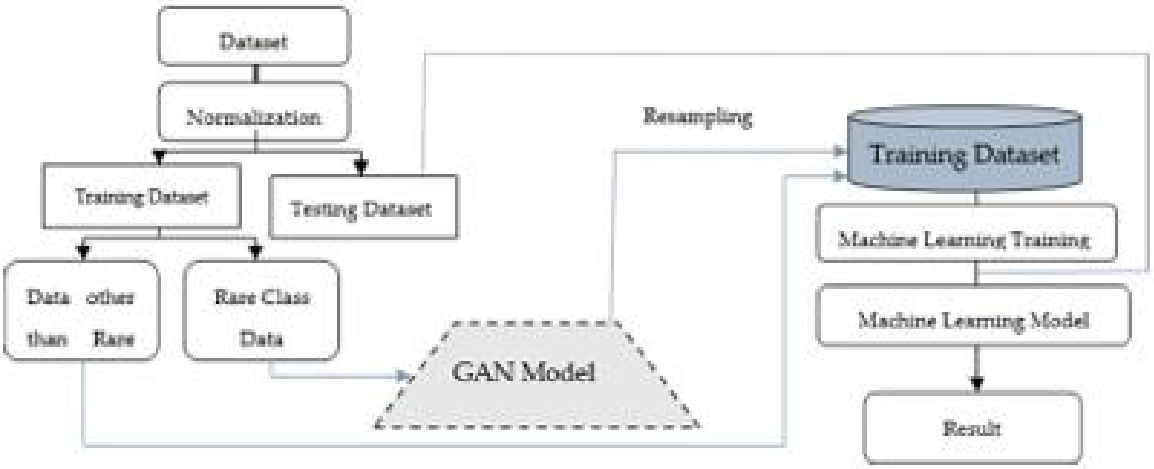


Figure 1 :Block Diagram

By comparing these models, we aim to identify not only the most effective algorithm but also the specific circumstances under which each model is most useful.

The ultimate goal of our project is to create a more resilient IDS that can protect digital networks from both common and rare cyber threats. By addressing the data imbalance problem and evaluating various ML algorithms, we hope to provide a framework for developing more accurate and reliable IDS that can adapt to the rapidly changing landscape of cyber threats. If successful, our approach could be transformative for organizations in all sectors, offering a more effective defense mechanism against rare, sophisticated cyberattacks. A strengthened IDS will allow organizations to detect and respond to rare threats proactively, reducing the risk of undetected attacks and enhancing the overall security of their digital infrastructure.

In summary, our project leverages GANs to balance IDS datasets and improve rare attack detection by generating realistic synthetic samples of underrepresented threats. By integrating this approach with a comparative analysis of ML algorithms, we aim to enhance IDS accuracy and adaptability, making a valuable contribution to cybersecurity. Through rigorous testing on real-world data, we expect to demonstrate the feasibility and effectiveness of our approach, paving the way for the next generation of IDS capable of addressing the challenges posed by rare, sophisticated cyber threats.

1.1 EXISTING SYSTEM:

The existing system for detecting cyberattacks in Intrusion Detection Systems (IDS) typically relies on traditional methods such as signature-based detection and anomaly-based detection:

1. **Signature-Based Detection**: This method uses predefined patterns or signatures of known threats to detect attacks. It is highly effective for identifying known threats but struggles with new or rare attacks that do not match existing signatures.
2. **Anomaly-Based Detection**: Anomaly detection systems monitor network traffic and system behavior to identify deviations from the normal. This method can detect previously unknown threats but often generates false positives, particularly when the anomaly is a benign variation rather than a malicious attack.

decision trees, is known for its robustness against overfitting, which makes it well-suited for high-dimensional data like IDS logs.

Comparing these models allows us to identify the most effective approach for detecting rare attacks within the balanced dataset. By testing each algorithm on both balanced and unbalanced data, we assess their strengths and weaknesses under realistic conditions, giving us insight into which methods work best for rare threat detection. Through rigorous testing on real-world data, we aim to demonstrate that a balanced dataset improves detection accuracy and enhances the IDS's ability to generalize, allowing it to detect both common and rare threats more effectively.

Beyond improving IDS performance, our project contributes to the broader cybersecurity landscape by addressing a key weakness in traditional detection systems. Standard IDS often rely on supervised learning techniques that train models on labeled data to recognize known attack patterns. However, the sheer diversity of today's cyber threats makes it challenging for supervised learning to keep up, especially for rare and novel attacks that deviate from established patterns. By incorporating GANs into the training process, we take a semi-supervised approach that allows the IDS to "learn" from both real and generated data. This hybrid approach bridges the gap between supervised and unsupervised learning, enhancing the IDS's capability to detect complex, previously unseen attacks.

The use of GANs in generating rare attack data could have a significant impact on the future of IDS and cybersecurity as a whole. Traditional data augmentation methods often fail to produce realistic samples, especially for complex threats like APTs. In contrast, GANs create synthetic data that is nearly indistinguishable from real attack data, providing IDS with an additional layer of training that enhances its detection capabilities. This is particularly valuable for organizations that may not have access to comprehensive datasets, as GANs allow them to generate synthetic data that improves dataset representativeness. As a result, IDS becomes more accessible and effective for smaller organizations that may lack the resources to collect extensive labeled data.

1. INTRODUCTION

Intrusion Detection Systems (IDS) are vital tools in safeguarding digital networks from the ever-evolving landscape of cyber threats. As cyberattacks grow more sophisticated, one of the biggest challenges for IDS is identifying rare but highly impactful attacks. These uncommon threats—such as advanced persistent threats (APTs) and zero-day exploits—are particularly difficult to detect because they are infrequent and operate with subtle, complex signatures. Traditional IDS, which are often trained on datasets populated with common, well-known threats, struggle to identify these rare attacks. When these attacks evade detection, they can infiltrate networks and bypass security protocols, leading to data breaches and network disruptions that carry serious consequences for organizations.

Our project addresses this challenge by tackling a critical weakness in traditional IDS: the data imbalance problem and data scarcity. In most IDS training datasets, frequent attacks are overrepresented, while rare attacks appear too infrequently to be learned effectively by the model. Without a balanced dataset, machine learning (ML) models within IDS struggle to generalize to rare attack scenarios, weakening their ability to respond effectively to these less common but damaging threats and which leads to false output.

To bridge this gap, we're using Generative Adversarial Networks (GANs) to generate synthetic samples of rare attacks. GANs, a type of ML model designed for data generation, consist of two networks that work against each other: the generator, which creates new samples, and the discriminator, which tries to distinguish between real and synthetic samples. This adversarial process improves the quality of synthetic data over time, allowing GANs to produce samples that closely resemble real-world attack data. By integrating GAN-generated samples into the IDS training process, we create a more balanced dataset that enhances the IDS's ability to detect rare attacks.

Once we've balanced the dataset with GAN-generated samples, we evaluate different ML algorithms to determine the optimal models for detecting rare attacks in both balanced and unbalanced datasets. We test a range of algorithms, including K-Nearest Neighbors (KNN), Random Forest, and LSVM chosen for their distinct capabilities in classifying patterns within network data by differentiating normal data and attack data. For instance, KNN is a straightforward but powerful tool for anomaly detection while random forest, an ensemble of

LIST OF TABLES

TABLE OF CONTENTS

ABSTRACT

Intrusion detection systems (IDS) are essential for securing modern networks against unauthorized access and cyber threats. Traditional IDS models often rely on machine learning (ML) algorithms to detect anomalies within network traffic data. However, a key limitation in these systems is the unbalanced nature of real-world data, where the majority of traffic is benign or "normal" and only a small fraction represents actual intrusions. This data imbalance leads to poor detection performance for rare intrusion types, as conventional ML models are biased towards the majority class, resulting in high false negatives for rare but critical classes of attacks.

This project addresses the class imbalance problem in IDS using a Generative Adversarial Network (GAN) model to enhance detection performance for rare intrusion classes. GANs, which consist of a generator and discriminator network in a competitive framework, are capable of generating realistic synthetic data. By training a GAN on the minority (rare) classes, we can generate synthetic samples to augment the training dataset, thus balancing the distribution across classes. The synthetic samples generated by the GAN complement the original dataset, allowing for a more robust ML model that can effectively identify rare intrusions.

The experimental setup includes a baseline model trained on the original unbalanced dataset and a proposed model trained on the GAN-augmented balanced dataset. The models are evaluated using Random Forest and K-Nearest Neighbors (KNN) as the classifier due to its simplicity and effectiveness in high-dimensional intrusion detection contexts. Performance is measured using standard metrics such as accuracy, precision, recall, F1-score, and the confusion matrix, with a particular emphasis on the metrics for the rare classes. Notably, the classification report isolates these rare classes, allowing for a detailed comparison of each model's ability to correctly identify intrusion types.

This project demonstrates the potential of GAN-based data augmentation to improve class balance and detection performance for rare intrusions in IDS. The proposed approach can be extended to other domains where data imbalance impacts classification efficacy, paving the way for more balanced and effective ML solutions in anomaly detection tasks.

DECLARATION

This is to certify that our project titled "**Enhancing Rare Class Attack Detection in Intrusion Detection System**" submitted to Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology in complete fulfilment of the requirement for the award of Bachelor of Technology in CSE- (Artificial Intelligence and Machine Learning) is a bonafide report to the work carried out by us under the guidance and supervision of Naga Durga Saile.K, Assistant Professor, Department of CSE-(AIML & IoT), Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology. To the best of our knowledge, this has not been submitted in any form to another University/Institute for an award of any degree/diploma.

Burugula Keerthana
21071A6673
Dept. of CSE (AIML&IoT)

Manga Kaveri
21071A66A2
Dept. of CSE (AIML&IoT)

Shaik Farooq
21071A66C1
Dept. of CSE (AIML&IoT))

Sheelam Saipriya
21071A66C3
Dept. of CSE (AIML&IoT)

Thummakunta Trishank
21071A66C6
Dept. of CSE (AIML&IoT)

Estd. 1995

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

An Autonomous, ISO 9001:2015 & QS I Gauge Diamond Rated Institute, Accredited by NAAC with 'A++' Grade
NBA Accreditation for B.Tech. CE, EEE, ME, ECE, CSE, EIE, IT Programmes
Approved by AICTE, New Delhi, Affiliated to JNTUH, NIRF 135th Rank in Engineering Category
Recognized as "College with Potential for Excellence" by UGC
Vignana Jyothi Nagar, Pragathi Nagar, Nizampet (S.O), Hyderabad – 500 090, TS, India.
Telephone No: 040-2304 2758/59/60, Fax: 040-23042761

CERTIFICATE

This is to certify that **Burugula Keerthana (21071A6673), Manga Kaveri (21071A66A2), Shaik Farooq (21071A66C1), Sheelam Saipriya (21071A66C3), Thummakunta Trishank (21071A66C6)** have successfully completed their Mini project work at CSE-(AIML & IoT) Department of VNRVJIET, Hyderabad entitled **"Enhancing Rare Class Attack Detection in Intrusion Detection Systems"** in partial fulfilment of the requirements for the award of B. Tech degree during the academic year 2024-2025.

This work is carried out under my supervision and has not been submitted to any other University/Institute for award of any degree/diploma.

GUIDE
Naga Durga Saile. K
Assistant Professor

Head of the Department
Dr. Sagar Yeruva
Associate Professor

Enhancing Rare Class Attack Detection in Intrusion Detection Systems

A Mini Project report submitted
in the partial fulfilment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science & Engineering

(Artificial Intelligence and Machine Learning)

by

21071A6673	Burugula Keerthana
21071A66A2	Manga Kaveri
21071A66C1	Shaik Farooq
21071A66C3	Sheelam Saipriya
21071A66C6	Thummakunta Trishank

Under the Guidance of

Naga Durga Saile.K

Assistant Professor

CSE – AIML & IoT, VNR VJIET

Submitted to

DEPARTMENT OF

CSE- (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING & INTERNET OF THINGS)

Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology,

Hyderabad, Telangana

November 2024

Hawk Tuah HAWK

#2494

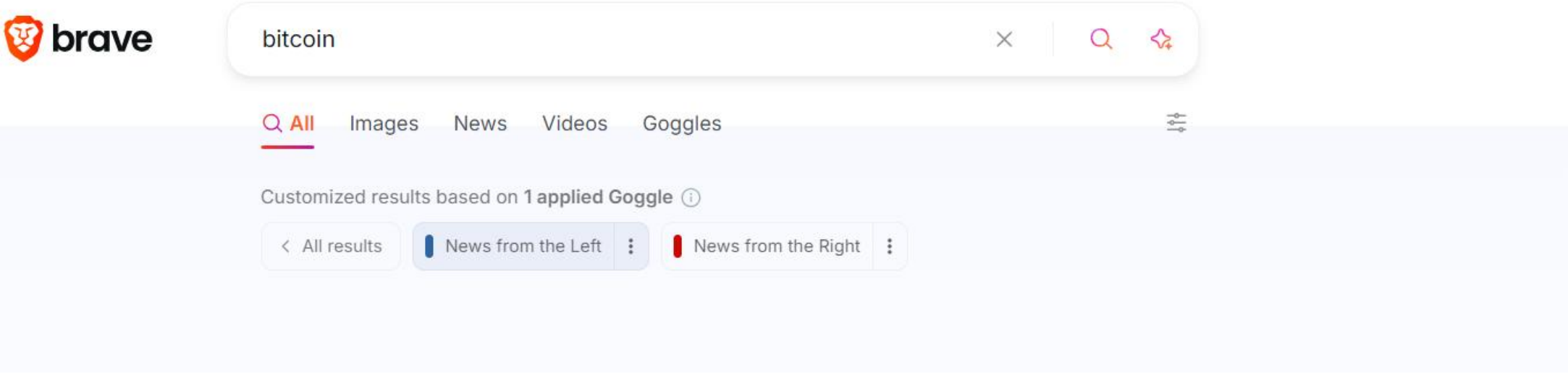
838

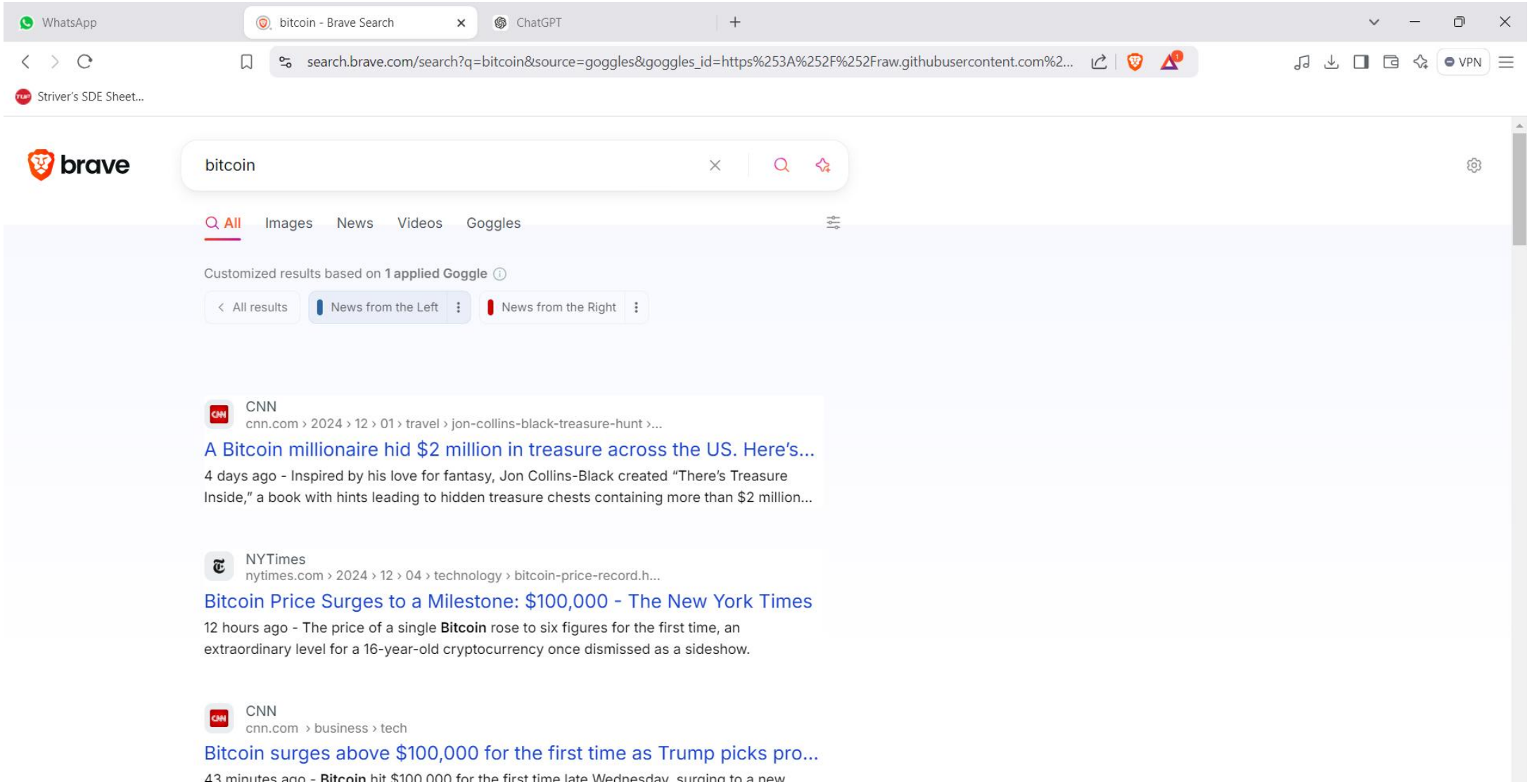
$0.0001478

▼ 29.25% (7d)

Market cap	Volume (24h)
$142.29K ▼ 25.44%	$130.23K ▲ 868.73%
FDV	**Vol/Mkt Cap (24h)**
$147.79K	91.20%
Total supply	**Max. supply**
1000M HAWK	1000M HAWK

Circulating supply


brave
bitcoin
All
Images
News
Videos
Goggles
Customized results based on 1 applied Goggle
All results
News from the Left
News from the Right


WhatsApp
bitcoin - Brave Search
ChatGPT
search.brave.com/search?q=bitcoin&source=goggles&goggles_id=https%253A%252F%252Fraw.githubusercontent.com%2...
VPN
Striver's SDE Sheet...
brave
bitcoin
All
Images
News
Videos
Goggles
Customized results based on 1 applied Goggle
All results
News from the Left
News from the Right
CNN
cnn.com › 2024 › 12 › 01 › travel › jon-collins-black-treasure-hunt ›...
A Bitcoin millionaire hid $2 million in treasure across the US. Here's...
4 days ago - Inspired by his love for fantasy, Jon Collins-Black created "There's Treasure Inside," a book with hints leading to hidden treasure chests containing more than $2 million...
NYTimes
nytimes.com › 2024 › 12 › 04 › technology › bitcoin-price-record.h...
Bitcoin Price Surges to a Milestone: $100,000 - The New York Times
12 hours ago - The price of a single Bitcoin rose to six figures for the first time, an extraordinary level for a 16-year-old cryptocurrency once dismissed as a sideshow.
CNN
cnn.com › business › tech
Bitcoin surges above $100,000 for the first time as Trump picks pro...
43 minutes ago - Bitcoin hit $100,000 for the first time late Wednesday, surging to a new

Application of biogas

Biogas is used as cooking fuel.

Biogas is mental light gas burner for lighting purpose.

Biogas is used for water heating.