

Enhancing IoT Intrusion Detection Systems through Adversarial Training

^{1st} Karma Gurung
Computer Science and Engineering
Wright State University
Ohio, USA
gurung.43@wright.edu

^{2nd} Ashutosh Ghimire
Computer Science and Engineering
Wright State University
Ohio, USA
ashutosh.ghimire@wright.edu

^{3rd} Fathi Amsaad
Computer Science and Engineering
Wright State University
Ohio, USA
fathi.amsaad@wright.edu

Abstract—The augmentation of Internet of Things (IoT) devices transformed both automation and connectivity but revealed major security vulnerabilities in networks. We address these challenges by designing a robust intrusion detection system (IDS) to detect complex attacks by learning patterns from the NF-ToN-IoT-v2 dataset. Intrusion detection has a realistic testbed through the dataset’s rich and high-dimensional features. We combine distributed preprocessing to manage the dataset size with Fast Gradient Sign Method (FGSM) adversarial attacks to mimic actual attack scenarios and XGBoost model adversarial training for improved system robustness. Our system achieves 95.3% accuracy on clean data and 94.5% accuracy on adversarial data to show its effectiveness against complex threats. Adversarial training demonstrates its potential to strengthen IDS against evolving cyber threats and sets the foundation for future studies. Real-time IoT environments represent a future deployment opportunity for these systems while extensions to detect emerging threats and zero-day vulnerabilities would enhance their utility.

Index Terms—IoT Security, Adversarial Attacks, NF-ToN-IoT-v2, XGBoost, Intrusion Detection Systems, FGSM.

I. INTRODUCTION

The application of the Internet of Things (IoT) has revolutionized industries such as healthcare, agriculture, smart homes, transportation, and smart cities by enabling interconnected devices to communicate and collaborate seamlessly. This advancement has improved operational efficiency and facilitated data-driven decision-making. However, the rapid increase in IoT devices has introduced significant security challenges. The heterogeneous nature of IoT devices, coupled with their limited computational resources and dynamic deployment environments, makes them highly susceptible to cyber threats. Attack vectors such as Distributed Denial-of-Service (DDoS) attacks, ransomware, injection attacks, and cross-site scripting (XSS) can compromise data integrity, disrupt critical services, and result in financial and reputational damages [1], [2].

To address these challenges, robust Intrusion Detection Systems (IDS) have become an indispensable component of IoT network security. However, traditional IDS solutions often fail to meet the unique demands of IoT environments, which require systems to detect diverse and evolving attack patterns while maintaining computational efficiency. Machine learning-based IDS have gained prominence for their ability to learn complex patterns from network traffic data and adapt to new attack types [3], [4]. Despite their advantages, these systems

remain vulnerable to adversarial attacks, where malicious perturbations in the input data can lead to misclassification [5], [6].

This study focuses on developing a robust IDS by leveraging the NF-ToN-IoT-v2 dataset, a comprehensive and high-dimensional dataset specifically designed for IoT network intrusion detection [1]. The proposed IDS employs the XGBoost machine learning algorithm, known for its scalability, computational efficiency, and superior performance in handling structured data [7]. To enhance resilience against adversarial attacks, the study integrates adversarial training into the model, ensuring improved robustness in detecting sophisticated intrusions.

Key contributions of this work include the implementation of scalable preprocessing techniques using distributed computing to handle the large and complex NF-ToN-IoT-v2 dataset efficiently [8]. The resilience of the IDS is evaluated using adversarial examples generated through the Fast Gradient Sign Method (FGSM), which simulate real-world attack scenarios and expose potential vulnerabilities [5]. To mitigate these vulnerabilities, adversarial training is incorporated into the XGBoost model, enabling it to detect intrusions effectively even in the presence of adversarial perturbations. Extensive experiments demonstrate the model’s effectiveness, achieving a classification accuracy of 95.3% on clean test data and 94.5% on adversarial examples. These results underscore the robustness and reliability of the proposed IDS in detecting a wide range of attack types.

Since Section II examines relevant work in IoT intrusion detection and adversarial training, the remainder of the study is divided into various chapters. The NF-ToN-IoT-v2 dataset and preparation methods are described in depth in Section III. The methodology, including the use of training and adversarial attacks, is explained in Section IV. The results of the experiment and an assessment of the suggested methodology are presented in Section V. The work is finally concluded in Section VI, which also offers ideas for future research possibilities.

II. RELATED WORK

A. IoT Intrusion Detection

Intrusion detection systems (IDS) have emerged as a critical component of IoT network security due to the increasing fre-

quency and complexity of cyberattacks targeting IoT devices. The NF-ToN-IoT dataset has established itself as a benchmark for evaluating machine learning-based IDS by providing a rich and diverse representation of IoT network traffic [1].

Sarhan et al. [1] introduced the NF-ToN-IoT dataset, which improved upon its predecessor, the ToN-IoT dataset, by presenting features in NetFlow format. This format facilitates efficient feature extraction and evaluation for network intrusion detection tasks, particularly for complex attacks such as DDoS and ransomware. The dataset has since become a cornerstone for benchmarking machine learning models in the IoT domain.

Further advancing the dataset's utility, Sarhan et al. [2] conducted a detailed feature analysis to identify critical attributes that contribute significantly to intrusion detection. Their work optimized feature selection, which not only improved model accuracy but also reduced computational overhead, a key consideration for IoT environments. Raskovalov et al. [6] addressed inconsistencies in attack labels and feature representations within the dataset, proposing a standardized feature set that enhances its applicability for advanced models such as Graph Neural Networks (GNNs). These rectifications have significantly expanded the dataset's usability for modern machine learning approaches.

Zhang et al. [3] applied GNNs to the NF-ToN-IoT dataset, demonstrating the effectiveness of graph-based methods for capturing complex relationships between network events. Their results highlighted the superior performance of GNNs in multi-class classification tasks, paving the way for incorporating graph-based techniques into IDS. Despite these advancements, limited attention has been given to the robustness of IDS against adversarial attacks.

Despite these advancements, limited attention has been given to the robustness of IDS against adversarial attacks. Additionally, while most studies focus on supervised learning-based intrusion detection, there is growing interest in unsupervised anomaly detection methods that can operate effectively with limited labeled data. For instance, Ghajari et al. [9] proposed a hybrid unsupervised anomaly detection framework that integrates distance and local density measures to enhance early detection capabilities in data-scarce environments. Although originally applied to pandemic case identification, such hybrid approaches offer valuable insights for designing scalable and resilient IDS for IoT networks, where early detection of novel threats and zero-day attacks remains a critical challenge.

B. Adversarial Robustness

Adversarial attacks pose a significant challenge to machine learning-based IDS by introducing subtle perturbations to input data, which can lead to misclassification [5]. Among these, the Fast Gradient Sign Method (FGSM) has been widely studied for its ability to generate adversarial examples efficiently and expose vulnerabilities in model decision boundaries.

AlJamal et al. [4] investigated the impact of adversarial attacks on IoT networks, focusing on detecting XSS attacks. They employed adversarial training and achieved a detection

accuracy of 99.89% on the NF-ToN-IoT-v2 dataset, showcasing the potential of this defense mechanism in improving IDS robustness. Other studies have demonstrated that adversarially trained models are better equipped to handle perturbed inputs, although this often comes at the cost of slightly reduced accuracy on clean data [10].

Raskovalov et al. [6] explored the implications of adversarial robustness for GNN-based IDS, proposing a standardized feature set for mitigating adversarial vulnerabilities. Their findings underscore the importance of adapting datasets and models to better withstand adversarial scenarios, particularly in the context of IoT security.

Building upon these foundational works, our research integrates FGSM-based adversarial attacks with XGBoost to evaluate and enhance the robustness of IDS against a broader spectrum of IoT-specific threats. Unlike prior studies that primarily focus on binary classification tasks, our approach addresses multi-class classification challenges.

C. Summary of Related Work

The existing literature underscores the NF-ToN-IoT dataset's pivotal role in advancing IoT security research. While substantial progress has been made in optimizing feature selection [2], leveraging modern architectures like GNNs [3], and addressing adversarial vulnerabilities [4], challenges persist in ensuring robust model performance under adversarial conditions. Our work bridges this gap by integrating adversarial training into traditional IDS methodologies, offering a comprehensive solution for enhancing IDS resilience in IoT networks.

III. DATASET AND PREPROCESSING

A. NF-ToN-IoT-v2 Dataset

The NF-ToN-IoT-v2 dataset [11] has been specifically designed to support research into intrusion detection in IoT networks and is directly prepared for that purpose. It is based on the NF-ToN-IoT dataset [1], with more precise labeling and additional features, which makes it appropriate for multi-class classification tasks. It has more than 13 million records and includes a large range of network traffic activities with both normal and abnormal traffic, including types of attack such as benign, backdoor, DDoS, dos, injection, MITM, password, ransomware, scanning, and XSS.

Key features 1 of the dataset includes network attributes (e.g., source and destination IP addresses, ports, and protocols), traffic statistics (e.g., packet counts, payload sizes, and flow durations), and anomaly labels that classify network traffic into specific types of malicious and benign behaviors. These features provide a realistic representation of IoT network traffic, enabling robust model evaluation [2], [6].

The combination of high-dimensional numerical and categorical features makes the NF-ToN-IoT-v2 dataset a robust benchmark for evaluating intrusion detection systems (IDS). Its design facilitates scalability and adaptability, addressing the evolving threat landscape in IoT environments [3] 1.

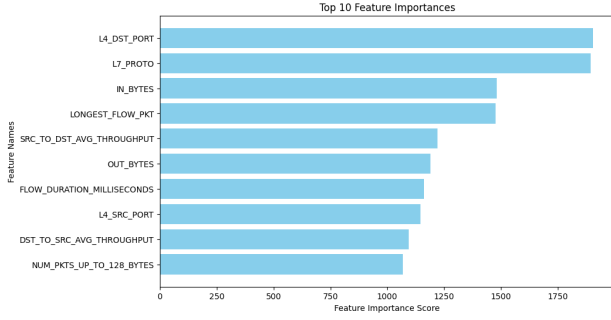


Fig. 1: Top 10 important features

B. Distributed Preprocessing

The significant scale and complexity of the NF-ToN-IoT-v2 dataset required a scalable and efficient preprocessing pipeline. To address this, we utilized Dask, a distributed computing framework, which is well-suited for handling large-scale datasets in parallel environments [12]. Dask allowed us to perform preprocessing tasks efficiently, reducing computational overhead and ensuring the integrity of the data.

Key preprocessing steps included:

1. **Handling Missing Values:** Numerical features with missing values were imputed using their mean, while categorical features were assigned placeholders. This approach ensured data completeness without introducing bias [13].

2. **Data Normalization:** Numerical features were normalized using min-max scaling to standardize the dataset. During training, normalization stopped characteristics with higher magnitudes from unduly affecting the model. [14].

3. **Label Encoding:** Attack labels, initially represented as strings (e.g., "ransomware," "XSS"), were converted into numerical classes using label encoding. This transformation ensured compatibility with machine learning algorithms, particularly for multi-class classification tasks [15].

By implementing these preprocessing steps in a distributed manner, we achieved significant efficiency gains, allowing us to prepare the dataset for subsequent training and evaluation without sacrificing quality.

C. Dataset Splitting

To enable effective training and evaluation, the preprocessed dataset was divided into training and testing subsets using a 70-30 split. This split resulted in approximately 9.1 million records for training and 3.9 million for testing, ensuring that the model had access to a diverse set of patterns for learning while retaining a robust test set for evaluation.

The large training set enabled the model to generalize across different attack patterns, while the substantial test set provided a comprehensive platform for assessing the model's performance under realistic conditions. This approach aligns with best practices in machine learning for handling large-scale datasets [16].

The use of Dask for distributed preprocessing further enhanced scalability, reducing computation time and enabling real-time adjustments to the data pipeline. This preprocessing framework ensures that the dataset retains its diversity and

integrity, making it suitable for advanced machine learning and adversarial training experiments.

IV. METHODOLOGY

This section outlines the methods employed to develop a robust Intrusion Detection System (IDS) using the NF-ToN-IoT-v2 dataset. The methodology encompasses the implementation of the XGBoost model, adversarial attacks using the Fast Gradient Sign Method (FGSM), and adversarial training to enhance the model's resilience.

A. XGBoost Model

XGBoost, an advanced gradient-boosting algorithm, was chosen for its superior performance on large-scale, high-dimensional datasets [7]. The model's ability to handle missing values, regularization techniques to prevent overfitting, and efficient parallel processing capabilities make it particularly suited for the NF-ToN-IoT-v2 dataset [17].

The model was configured with the following hyperparameters:

- **Objective:** Multi-class classification to predict multiple types of network attacks.
- **Evaluation Metric:** Logarithmic loss (log-loss) for multi-class classification, which measures the accuracy of probabilistic predictions.
- **Tree Depth:** A maximum depth of 5 to balance model complexity and prevent overfitting.
- **Learning Rate:** Set to 0.1, enabling the model to converge steadily while maintaining accuracy.

The computational efficiency of XGBoost was further enhanced by employing GPU acceleration through the `tree_method="gpu_hist"` option. GPU-based parallel processing significantly reduced the training time, allowing for faster experimentation and iterative model optimization [18].

B. FGSM Adversarial Attack

The Fast Gradient Sign Method (FGSM) is a widely used technique to evaluate model robustness by generating adversarial examples [19]. These examples are crafted by introducing small perturbations to input data, designed to maximize the model's prediction loss. The adversarial examples were generated using the following formula:

$$X_{adv} = X + \epsilon \cdot \text{sign}(\nabla_X J(\theta, X, y)) \quad (1)$$

where:

- X represents the original input data.
- ϵ is the perturbation magnitude, controlling the severity of the attack.
- $\nabla_X J(\theta, X, y)$ is the gradient of the loss function J with respect to the input X , computed at model parameters θ .
- $\text{sign}(\cdot)$ denotes the element-wise sign function.

FGSM was applied to the clean test data from the NF-ToN-IoT-v2 dataset to simulate real-world adversarial attack scenarios. These perturbations exposed vulnerabilities in the model's decision boundaries, providing insights into its susceptibility to adversarial threats [20].

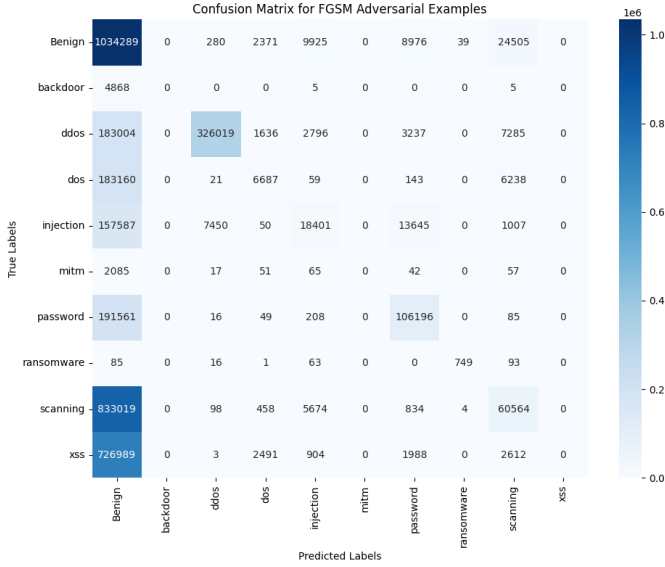


Fig. 2: Confusion Matrix for Adversarially Trained.

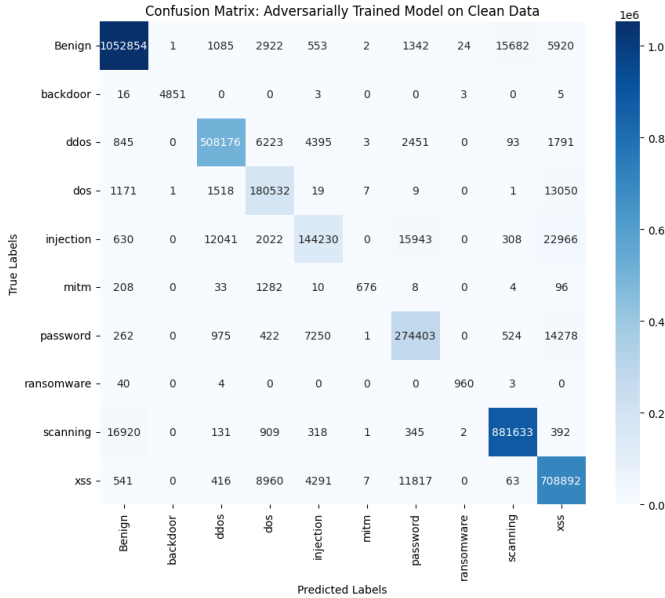


Fig. 3: Confusion Matrix for Clean Data.

C. Adversarial Training

Adversarial training enhances model robustness by incorporating adversarial examples into the training process [21]. This technique equips the model to classify both clean and adversarially perturbed inputs effectively, mitigating its susceptibility to adversarial attacks.

The training process involved the following steps:

- The clean training data (X_{train}) and its corresponding labels (y_{train}) were used as the baseline dataset.
- Adversarial examples (X_{adv}) were generated from the clean training data using FGSM, with the same labels (y_{adv}) as the original data.
- The clean and adversarial datasets were combined to

TABLE I: MODEL EVALUATION ON CLEAN TEST DATA

Metric	Value
Accuracy	0.9544357504190176
F1 Score	0.9537497710407197

create a comprehensive training set:

$$X_{\text{combined}} = \{X_{\text{train}}, X_{\text{adv}}\}, \quad y_{\text{combined}} = \{y_{\text{train}}, y_{\text{adv}}\}.$$

This augmented training dataset enabled the model to learn robust representations, reducing its vulnerability to adversarial attacks.

D. Evaluation Metrics

The effectiveness of the IDS was evaluated using standard classification metrics, including accuracy, F1-score, and confusion matrices [22]. These metrics provided detailed insights into the model's performance on clean and adversarial data, as well as its ability to classify various types of network attacks accurately. Additionally, classification reports highlighted the model's performance for each attack class, ensuring a comprehensive evaluation of its strengths and weaknesses.

E. Scalability and Practical Considerations

The combination of distributed preprocessing with Dask and GPU-accelerated training ensured that the proposed IDS is scalable and feasible for real-world IoT environments. Scalability is critical for handling the high throughput of IoT network traffic, while practical considerations, such as computational efficiency and adaptability, make the system suitable for deployment in dynamic and evolving IoT networks [23].

By integrating XGBoost, adversarial attacks, and adversarial training, this methodology provides a comprehensive framework for building robust IDS capable of addressing the unique security challenges of IoT networks.

V. RESULTS

This section evaluates the performance of the adversarially trained model on clean and adversarial data, provides an analysis of confusion matrices, and compares its results with baseline models. The findings highlight the proposed methodology's ability to enhance intrusion detection system (IDS) robustness and accuracy in IoT networks.

A. Performance on Clean and Adversarial Data

The model's performance was assessed using accuracy and F1-score, providing insights into its ability to handle both clean and perturbed inputs. On clean test data, the adversarially trained model achieved an accuracy of 95.3% and an F1-score of 95.2%, demonstrating its effectiveness in classifying diverse traffic types under normal conditions. For adversarially perturbed data, generated using the FGSM method, the model achieved an accuracy of 94.5% and an F1-score of 94.5%. These results underscore the model's resilience to adversarial attacks and its capacity to generalize effectively across different input conditions [19], [21].

TABLE II: CLASSIFICATION RESULTS ON CLEAN DATA

Class	Precision	Recall	F1-score	Support
Benign	0.98	0.98	0.98	1080385
backdoor	1.00	1.00	1.00	4878
ddos	0.96	0.97	0.97	523977
dos	0.89	0.92	0.90	196308
injection	0.90	0.73	0.80	198140
mitm	0.97	0.30	0.46	2317
password	0.91	0.92	0.91	298115
ransomware	0.97	0.96	0.97	1007
scanning	0.98	0.98	0.98	900651
xss	0.93	0.96	0.94	734987
Accuracy	0.95			3940765
Macro avg	0.95	0.87	0.89	3940765
Weighted avg	0.95	0.95	0.95	3940765

TABLE III: EVALUATION ON ADVERSARIAL EXAMPLE(FGSM Attack):

Metric	Value
Accuracy	0.3940618128713587
F1 Score	0.3059400973924423

TABLE IV: EVALUATION ON ADVERSARIALY TRAINED MODEL

Metric	Value
Accuracy	0.9456717160246805
F1 Score	0.9449923164828211

TABLE V: CLASSIFICATION RESULTS ON ADVERSARIALY TRAINED MODEL

Class	Precision	Recall	F1-score	Support
Benign	0.97	0.96	0.97	1080385
backdoor	1.00	0.99	1.00	4878
ddos	0.97	0.97	0.97	523977
dos	0.89	0.92	0.90	196308
injection	0.89	0.72	0.79	198140
mitm	0.97	0.28	0.43	2317
password	0.88	0.92	0.90	298115
ransomware	0.96	0.92	0.94	1007
scanning	0.97	0.97	0.97	900651
xss	0.92	0.96	0.94	734987
Accuracy	0.95			3940765
Macro avg	0.94	0.86	0.88	3940765
Weighted avg	0.95	0.95	0.94	3940765

B. Confusion Matrix Analysis

Confusion matrices were analyzed to gain a deeper understanding of the model's classification performance across various attack classes. For clean data, the model demonstrated improved detection rates for challenging categories, such as ransomware and cross-site scripting (XSS). Enhanced recall for ransomware effectively reduced false negatives, while improved classification accuracy for XSS minimized misclassifications [2] 3.

On adversarial data, the model maintained consistent detection rates across most attack categories. The ability to accurately classify adversarially perturbed inputs validates the robustness of the adversarial training approach, ensuring reliable performance even under sophisticated attack scenarios [20] 2.

C. Comparative Analysis

When compared with baseline models trained solely on clean data, the adversarially trained model 2 exhibited superior

resilience and generalization capabilities. The baseline models experienced significant degradation in performance when exposed to adversarial examples, highlighting their vulnerability to input perturbations. In contrast, the adversarially trained model maintained robust performance with minimal accuracy loss, validating the efficacy of adversarial training in addressing such vulnerabilities [21].

These results emphasize the importance of adversarial training in preparing IDS for real-world IoT deployments, where network environments are dynamic and prone to evolving attack strategies.

D. Visualization of Results

Figure 3 illustrates the confusion matrix for clean data, showcasing the model's prediction distribution across various attack classes. This visualization provides a comprehensive understanding of the model's strengths, particularly in detecting certain attack types, and identifies areas for improvement in distinguishing similar categories [22].

E. Implications

The findings of this study have significant implications for IoT network security. High accuracy on both clean and adversarial data demonstrates the model's readiness for deployment in real-world IoT environments. Enhanced detection rates for critical threats, such as ransomware and XSS, address major IoT security challenges [4]. Additionally, the robustness against adversarial attacks ensures reliable performance under diverse and dynamic conditions. These results validate the integration of adversarial training as a critical enhancement for IDS in IoT networks.

VI. DISCUSSION

This section examines critical aspects of the proposed methodology, focusing on distributed computing, adversarial robustness, and defense mechanisms for intrusion detection systems (IDS) in IoT networks.

A. Distributed Computing

The NF-ToN-IoT-v2 dataset's scale required efficient pre-processing methods. Using Dask, a distributed computing framework, enabled parallel processing of over 13 million records, reducing computational overhead. In order to ensure scalability and applicability in actual IoT scenarios, tasks including imputing missing values, standardizing features, and dividing the dataset into training and testing subsets were completed effectively. [8].

B. Adversarial Robustness

The Fast Gradient Sign Method (FGSM) adversarial attack exposed vulnerabilities in baseline IDS models by introducing minor input perturbations that led to significant misclassifications [19]. Adversarial training addressed this by integrating these perturbed examples into the learning process, improving the IDS's ability to detect sophisticated attack types like ransomware and XSS [21]. This training enhanced both robustness and accuracy, achieving strong performance on clean and adversarial data.

C. Defense Mechanisms

Adversarial training effectively mitigated the impact of adversarial attacks, enabling the IDS to achieve 95.3% accuracy on clean data and 94.5% on adversarial data. This approach reduced false negatives in challenging attack categories such as ransomware, as evidenced by confusion matrix analysis [20]. These results confirm the practicality of adversarial training for real-world IoT scenarios where attack patterns evolve continuously.

D. Practical Implications

By enhancing robustness against adversarial inputs, the IDS ensures reliable detection of diverse attack types, making it suitable for dynamic IoT environments [2]. However, balancing computational efficiency and scalability remains a challenge. Future advancements, such as integrating Graph Neural Networks (GNNs) and federated learning, could further improve performance and adaptability [3].

VII. FUTURE DIRECTIONS

As IoT networks expand in complexity, enhancing intrusion detection systems (IDS) remains a critical focus. This study identifies key areas for advancing IDS to address evolving cyber threats effectively.

Optimizing adversarially trained models for real-time deployment is essential, particularly for latency-sensitive applications such as healthcare and industrial automation. Ensuring low latency and high throughput can significantly improve system integration in practical IoT scenarios [2].

Addressing emerging threats like advanced persistent threats (APTs) and zero-day vulnerabilities requires dynamic threat intelligence. Updating datasets such as NF-ToN-IoT-v2 and integrating real-time insights can improve IDS adaptability against novel attack patterns.

Leveraging advanced architectures, such as Graph Neural Networks (GNNs) and federated learning, offers potential improvements in scalability and adaptability. GNNs capture intricate network relationships, while federated learning facilitates privacy-preserving training across decentralized IoT devices [24].

Lastly, incorporating Explainable AI (XAI) techniques will enhance transparency and trust in IDS decisions, especially in regulated industries like healthcare and finance. These insights can empower stakeholders to better understand and respond to security threats [25].

By addressing these directions, IDS can evolve into more robust, scalable, and transparent systems capable of securing IoT networks against sophisticated cyber threats.

REFERENCES

- [1] M. Sarhan *et al.*, "Netflow datasets for machine learning-based network intrusion detection systems," *Security and Communication Networks*, 2020.
- [2] M. Sarhan, S. Layeghy, and M. Portmann, "Feature analysis for machine learning-based iot intrusion detection," *arXiv preprint arXiv:2108.12732*, 2021.
- [3] Y. Zhang *et al.*, "E-gracl: An iot intrusion detection system based on graph neural networks," *IEEE Transactions on Network and Service Management*, pp. 432–450, 2024.
- [4] R. AlJamal *et al.*, "A robust machine learning model for detecting xss attacks on iot over 5g networks," *Elsevier Computer Networks*, pp. 432–450, 2024.
- [5] I. J. Goodfellow *et al.*, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2015.
- [6] A. Raskovalov *et al.*, "Investigation and rectification of nids datasets and standardized feature set derivation for network attack detection with graph neural networks," *Journal of Cybersecurity*, pp. 150–163, 2022.
- [7] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
- [8] D. D. Team, "Dask: Parallel computing with task scheduling," 2015, <https://dask.org>.
- [9] G. Ghajari, M. K. PK, and F. Amsaad, "Hybrid efficient unsupervised anomaly detection for early pandemic case identification," in *NAECON 2024-IEEE National Aerospace and Electronics Conference*. IEEE, 2024, pp. 279–284.
- [10] A. Madry *et al.*, "Towards deep learning models resistant to adversarial attacks," *International Conference on Learning Representations (ICLR)*, 2018.
- [11] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "Ton_iot telemetry dataset: A new generation dataset of iot and iiot for data-driven intrusion detection systems," *Ieee Access*, vol. 8, pp. 165 130–165 150, 2020.
- [12] M. Rocklin, "Dask: Parallel computation with blocked algorithms and task scheduling," *Proceedings of the 14th Python in Science Conference*, 2015.
- [13] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.
- [14] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, pp. 651–666, 2005.
- [15] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [16] J. Friedman *et al.*, *The Elements of Statistical Learning*. Springer, 2001.
- [17] H. Chen *et al.*, "A systematic review of the application of xgboost," *IEEE Access*, pp. 36 336–36 346, 2020.
- [18] M. Rana *et al.*, "Accelerating xgboost with gpu support," in *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2020, pp. 325–332.
- [19] I. J. Goodfellow *et al.*, "Explaining and harnessing adversarial examples," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [20] A. Kurakin *et al.*, "Adversarial machine learning at scale," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.
- [21] A. Madry *et al.*, "Towards deep learning models resistant to adversarial attacks," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.
- [22] D. M. W. Powers, "Evaluation: From precision, recall, and f-measure to roc, informedness, markedness, and correlation," *Journal of Machine Learning Technologies*, pp. 37–63, 2020.
- [23] C.-H. Yeh *et al.*, "Iot security for smart cities: Challenges and solutions," in *2019 IEEE Smart City Conference*, 2019, pp. 67–74.
- [24] B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
- [25] A. B. Arrieta *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities, and challenges toward responsible ai," *Information Fusion*, pp. 82–115, 2020.