An Algorithm based on Deep Learning for Intrusion Detection in IoT

Zoya Firdouse[1], K. Jayasree[2]

[1]PG Scholar, Department of Information and Technology, G. Narayanamma Institute of Technology and Science, India.
Email: *zoyafirdouse6[at]gmail.com*

[2]Assistant Professor, Department of Information and Technology, G. Narayanamma Institute of Technology and Science, India.
Email: *jayasree1231[at]gmail.com*

Abstract: ***In order to improve intrusion detection system (IDS) performance in Internet of Things (IoT) environments, an innovative framework is presented in this study. By leveraging the ToN-IoT telemetry dataset, which includes environmental sensor data such as temperature, pressure, and humidity, the framework aims to improve the classification and prediction of cyber attacks. This approach integrates machine learning and deep learning algorithms, including Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (KNN), Gradient Boosting, and Long Short-Term Memory (LSTM), to provide a comprehensive analysis. By combining telemetry data with traditional network data, the proposed framework offers a more holistic view of the system, thereby enhancing the accuracy of intrusion detection. This method helps in reducing false positives and improving contextual awareness, allowing for the differentiation between legitimate environmental changes and potential cyber threats. The results of our study emphasize the significance of combining a variety of data sources and cutting-edge algorithms to strengthen IoT systems against cyber threats by showing how the ToN-IoT telemetry dataset greatly enhances the detection and prediction of Cyberattacks.***

Keywords: Random forest (RF), Decision Tree (DT), Internet of Things (IoT), intrusion detection systems, *KNN, Gradient boost and LSTM.*

1. Introduction

The widespread use of Internet of Things (IoT) devices has resulted in unparalleled ease and effectiveness in a multitude of fields, such as environmental surveillance, automated homes, automation in industries, and healthcare. But there are also serious security flaws as a result of the quick integration of IoT devices into everyday life. IoT systems, due to their interconnected nature and the vast amounts of data they generate, have become prime targets for cyber-attacks. Intrusion Detection Systems (IDS) are critical for safeguarding these environments by identifying and mitigating potential threats.

IDS are vital components in network security architectures, designed to detect unauthorized access or anomalies that could indicate a cyber-attack. Traditional IDS approaches are broadly categorized into signature-based detection systems (SIDS) and anomaly-based detection systems (AIDS). SIDS rely on predefined patterns of known threats, making them efficient for detecting well-known attacks but less effective against novel threats. Conversely, AIDS models identify deviations from normal behavior, which allows them to detect previously unseen attacks but often at the cost of higher false positive rates.

Resource limitations, heterogeneity, and the dynamic nature of IoT devices are some of the distinctive features of IoT settings which pose specific challenges to traditional IDS implementations. For instance, It is challenging to implement computationally demanding security measures on IoT devices as they frequently have low processor and memory capacities. Furthermore, the creation of standardized security solutions is complicated by the diversity of IoT devices and the range of communication protocols employed. Additionally, IoT environments are highly dynamic, with devices frequently joining and leaving the network, which necessitates adaptable and scalable IDS solutions.

In order to tackle these issues, this research suggests a novel IDS framework that leverages the ToN-IoT telemetry dataset, which includes a rich array of environmental sensor data such as temperature, pressure, and humidity. By integrating telemetry data with traditional network data, The objective of the proposed framework is to improve IDS's contextual awareness so that it can differentiate between possible cyber threats and changes in the environment that are real. This all-encompassing strategy boosts intrusion detection accuracy while simultaneously reduces false positives by providing a more comprehensive understanding of the operational environment.

The proposed IDS framework employs a combination of machine learning (ML) and deep learning (DL) algorithms to analyze the telemetry data. The selected algorithms include Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (KNN), Gradient Boosting, and Long Short-Term Memory (LSTM). Each of these algorithms offers unique advantages in processing and analyzing large datasets. For example, RF and DT are known for their robustness and interpretability, While LSTM is useful for analysis of time series in telemetry data because it is especially good at identifying temporal relationships in sequential data.

The implementation of ML and DL approaches in IDS for IoT contexts has been the subject of several research. For instance, Ertam et al. (2017) investigated the effectiveness of several machine learning (ML) algorithms in identifying network intrusions and demonstrated how ensemble learning techniques in improving detection accuracy. Similarly, Jan et al. (2019) presented a lightweight IDS designed

Paper ID: SR24813173848 DOI: https://dx.doi.org/10.21275/SR24813173848

specifically for IoT contexts, highlighting the importance of efficient and extensible security solutions. However, there remains a significant gap in research concerning the integration of diverse data sources, such as telemetry data, to enhance the contextual awareness of IDS in IoT environments.

This study addresses these research gaps by presenting a comprehensive IDS framework that combines telemetry and network data using advanced ML and DL algorithms. The following are the study's main contributions:

- The advancement of a novel IDS framework that leverages the ToN-IoT telemetry dataset for improved intrusion detection.
- The integration of multiple ML and DL algorithms to enhance the accuracy and robustness of the IDS.
- The demonstration of the framework's effectiveness in reducing false positives and improving contextual awareness in IoT environments.

This deep learning-based approach not only enhances the detection capabilities of IDS in IoT networks but also addresses the challenges of scalability and adaptability. The details of the suggested algorithm are covered in detail in the sections that follows, its implementation, and performance evaluation, showcasing its superiority over traditional methods in ensuring robust security for IoT ecosystems.

2. Literature Survey

This paper introduces a novel IoT network testbed architecture designed to evaluate AI-based security applications. Utilizing the Service Orchestration (SO), the architecture supports Software-Defined Network (SDN), NSX vCloud NFV platform, AND Network Function Virtualization (NFV), to create dynamic testbed networks integrating edge, fog, and cloud tiers. These datasets include telemetry from IoT services, Linux and Windows systems, and network traffic data. The TON_IoT dataset is verified using various machine learning-based algorithms intrusion detection algorithms—Deep Neural Networks, Naive Bayes, Random Forest, and Gradient Boosting —which showed good detection accuracy. A comparative analysis with other network datasets highlights TON_IoT's ability to capture diverse legitimate and anomalous patterns, enhancing the validation of new AI-based security solutions. [1]

The internet has become a vital aspect of daily life, with the number of linked gadgets, notably Internet of Things (IoT), increasing rapidly. This surge has brought numerous security challenges that lack well-defined solutions. Several ways have been proposed to safeguard IoT networks, there is still a significant scope for improvement. One promising approach is the application of machine learning. This research explores multiple ML and DL strategies, utilizing standard datasets to enhance IoT security performance. Specifically, we developed a deep learning algorithm to detect denial-of-service (DoS) attacks. Implemented in Python using packages such as Seaborn, TensorFlow, and scikit-learn our findings show that a DL model can significantly improve accuracy, making IoT network attack mitigation more effective. [2]

Cybersecurity is vital for applications in cars, devices, homes, industrial systems, but IoT devices remain challenging to secure. This research explores deep learning-based intrusion detection methods, using GRUs, LSTMs, and CNNs to develop models. A standard IoT intrusion detection dataset evaluates these models. The study's empirical results are compared with existing approaches, showing that the proposed method achieves the highest accuracy in detecting IoT intrusions. [3]

Leveraging advanced information technology, Smart grids have become prime targets for cyber-attacks due to their reliance on IoT sensor devices that gather information from the grid and send it to the cloud. This vast cloud network, serving various smart infrastructures like homes and buildings, offers a substantial attack surface for cybercriminals. This study presents a comprehensive approach for detecting intrusions in IoT systems. Utilizing the IoTID20 dataset, which is derived from IoT infrastructure, the framework employs three advanced DL algorithms: a hybrid CNN-LSTM model, LSTM and CNN. To enhance system performance, PSO was applied for feature selection, reducing dataset dimensionality. The processed features were then classified using the deep learning models, achieving accuracies of 98.80% for CNN-LSTM, 99.82% for LSTM, and 96.60% for CNN. The proposed framework demonstrated superior performance on new datasets and will be implemented in our university's IoT environment. Comparative analysis showed that this system more effectively enhances IoT security against attacks, confirming its capability to detect real-world intrusions and improve IoT security. [4]

The IoT represents a new era where interconnected smart devices and sensors form a global network, influencing Each element of human action. Given its significant economic effect and widespread presence, IoT has become a prime target for cybercriminals, making cybersecurity a top priority. Traditional cybersecurity methods are becoming inadequate against the expansive IoT architecture and emerging threats. Deep learning provides intriguing possibilities for IoT intrusion detection due to its anomaly-based, data-driven methodology, which may discover new and unexpected attacks. [5]

The IoT has immense potential in applications like healthcare, defense, and power grids, making its security vital. Because of their limited resources and processing capacity, IoT networks are vulnerable to various attacks. Efficient safety actions, such as IDS, are necessary. This paper presents novel IDS that uses deep learning to classify traffic flow. By leveraging a newly published IoT dataset, we extract packet-level features and design a feed-forward neural network approach to multi-class and binary categorization of attacks. like information theft, reconnaissance, DDoS, and DoS. [6]

The rapid expansion of the IoT has increasingly attracted cybercriminals, as indicated by the increasing number of attacks against IoT devices and communication channels. Undetected attacks can cause severe service interruptions and financial losses, posing significant threats to identity protection. Real-time intrusion detection is critical for ensuring the profitability, security, and reliability of IoTenabled services. In this study, an inventive intrusion detection system (IDS) for Internet of Things devices is introduced. To detect fraudulent data aimed at linked Internet of Things devices, the system utilizes a four-layer deep Fully linked (FC) network architecture. Experimental performance analysis shows that the system effectively detects Sinkhole, Opportunistic Service, DDoS, Blackhole, and Wormhole attacks with an average accuracy of 93.74%. [7]

3. Proposed Method

The proposed method for developing the novel intrusion detection framework involves several key steps, including data collection, data preprocessing, feature engineering, model selection, and training. An comprehensive description of various procedures is given in this section, ensuring the reproducibility and transparency of the proposed approach.

a) Data Collection

The ToN-IoT telemetry dataset, which comprises an extensive collection of network and telemetry data, was employed in this investigation. The telemetry data encompasses environmental sensor readings such as temperature, pressure, and humidity, while the network data includes features relevant to network traffic and system logs.

b) Data Preprocessing

In order to get the dataset ready for machine learning and deep learning algorithms, data preparation is essential. The following procedures were done to preprocess the dataset::

- Data Cleaning: Missing values were handled by imputation or removal, and irrelevant features were dropped to ensure the quality of the dataset.
- Label Encoding: Label encoding was used to transform categorical variables—like temp_ condition—into numerical values that could be used with machine learning models.
- Feature Selection: Using statistical analysis and domain expertise, relevant characteristics were chosen to enhance model performance and minimize computing complexity.
- Data Splitting: An 80-20 split was used to divide the dataset into training and testing sets. This guarantees that the model's generalization skills are tested on a different sample of data after it has been trained on the first.

c) Feature Engineering

Through feature engineering, unstructured data is converted into useful characteristics that improve machine learning model performance. In this study, additional features were engineered from the telemetry data to capture temporal patterns and correlations between different sensors.

d) Model Selection

Several machine learning and deep learning algorithms were selected for their potential to effectively analyze and classify the telemetry and network data. The selected algorithms include:

- Random Forest (RF) is an ensemble learning approach noted for its resilience and interpretability.
- Decision Tree (DT): A simple yet powerful model that works well with categorical data.
- k-Nearest Neighbors (KNN): A non-parametric method effective for classification tasks.
- Gradient Boosting: An ensemble technique that builds models sequentially to reduce errors.
- Long Short-Term Memory (LSTM): A particular category of RNN that is compatible with time-series data due to its ability to capture temporal dependencies.

e) Model Training

The models were trained on the preprocessed training dataset. For machine learning models, standard training procedures were followed, including hyperparameter tuning through cross-validation. For the LSTM model, the data was reshaped to fit the input requirements of the LSTM layers.

Random Forest and Decision Tree: Trained using the scikit-learn library with hyperparameters tuning.
k-Nearest Neighbors: Trained with different values of k to find the optimal number of neighbors.
Gradient Boosting: Trained using the scikit-learn implementation with gradient boosting classifiers.
LSTM: Trained using the Keras library with appropriate input reshaping and one-hot encoding of target labels.

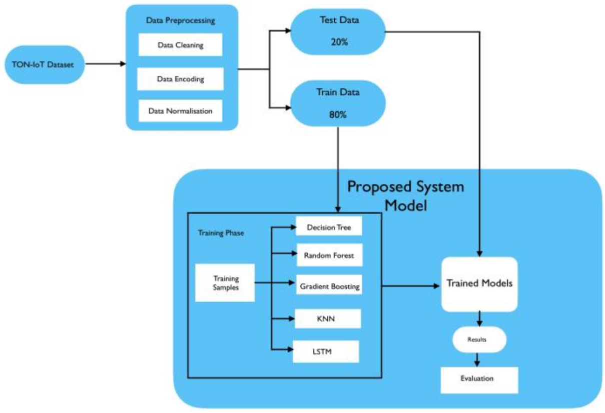


Figure 3.1: Proposed System Model

f) Evaluation Metrics

The models' performance was assessed through the use of standard metrics including F1-score, recall, accuracy, and precision. These metrics offer a thorough grasp of the model's capacity to accurately categorize both benign and malicious activity.

g) Experimental Setup

The experiments were carried out in a typical computer environment using sufficient computational resources to handle the training and evaluation of machine learning and deep learning models. The training process was monitored to prevent overfitting and ensure optimal performance.

By following this structured methodology, the proposed framework aims to leverage the ToN-IoT telemetry dataset effectively, showcasing notable advancements in the identification and forecasting of cyberattacks inside Internet of Things settings. The integration of diverse data sources and advanced algorithms underscores the importance of a holistic approach to securing IoT systems against emerging threats.

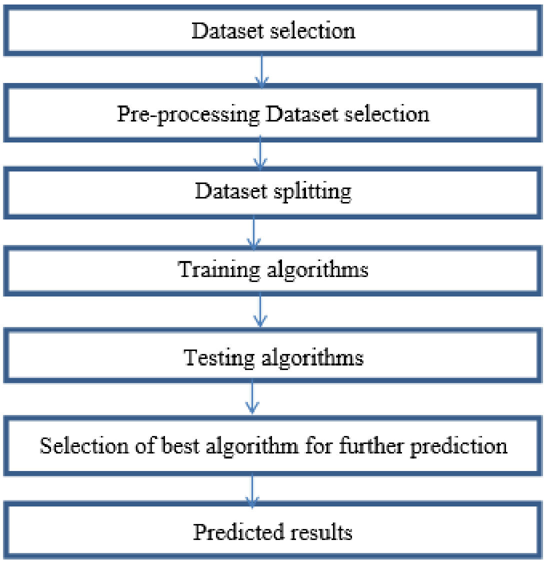


Figure 3.2: Flowchart of Proposed Methodology

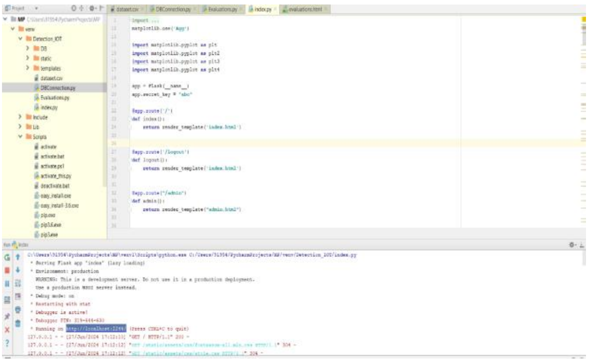
4. Result Analysis

The experimental results of the proposed Intrusion Detection System (IDS) framework for IoT environments highlight the effectiveness of telemetry data from IoT devices. This integration leverages the strengths of multiple machine learning and deep learning algorithms to enhance the detection and classification of cyber attacks. Here we discuss the detailed performance evaluation of each algorithm and the overall IDS framework.

Performance Evaluation

Several performance measures, including as accuracy, precision, recall, and F1-score, were used to assess the suggested IDS framework. These metrics give a thorough picture of how well the model can distinguish between benign and malicious behavior in an Internet of things environment.

Homepage is prepared using HTML/CSS coding.



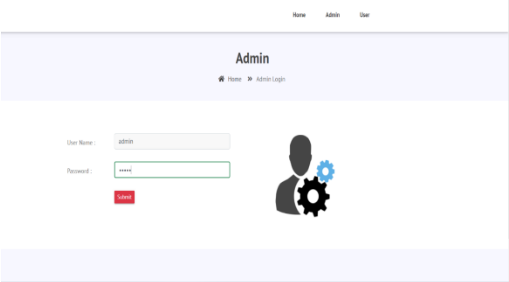
```
import ...
matplotlib.use('Agg')

import matplotlib.pyplot as plt
import matplotlib.pyplot as plt2
import matplotlib.pyplot as plt3
import matplotlib.pyplot as plt4

app = Flask(__name__)
app.secret_key = "abc"

@app.route('/')
def index():
    return render_template('index.html')


@app.route('/logout')
def logout():
    return render_template('index.html')


@app.route("/admin")
def admin():
    return render_template("admin.html")
```

```
C:\Users\31954\PycharmProjects\MP\venv1\Scripts\python.exe C:/Users/31954/PycharmProjects/MP/venv/Detection_IOT/index.py
 * Serving Flask app "index" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 319-644-630
 * Running on http://localhost:2246/ (Press CTRL+C to quit)
127.0.0.1 - - [27/Jun/2024 17:12:10] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [27/Jun/2024 17:12:12] "GET /static/assets/css/fontawesome-all.min.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Jun/2024 17:12:12] "GET /static/assets/css/style.css HTTP/1.1" 304 -
```

Figure 4.1: Importing algorithm

Figure 4.2: Intrusion detection IoT

Admin Login page is also has HTML/CSS code.



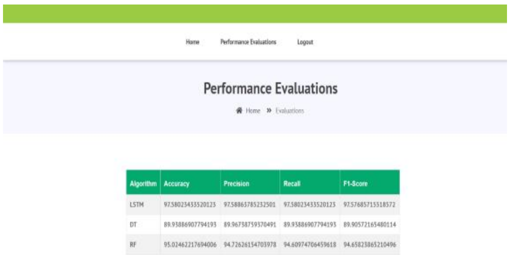
User Name : admin

Password : •••••

Submit

Figure 4.3: Admin login

Performance metrics figures and graphs from the experimental results illustrate the comparative performance of each algorithm. The graphs below clearly illustrate the advantages of LSTM in processing IoT telemetry data for intrusion detection by comparing the accuracy, precision, recall, and F1-score of various models.



Algorithm	Accuracy	Precision	Recall	F1-Score
LSTM	97.58023433520123	97.58863785232501	97.58023433520123	97.57685715518572
DT	89.93886907794193	89.96738759370491	89.93886907794193	89.90572165480114
RF	95.02462217694006	94.72626154703978	94.60974706459618	94.65823865210496

Figure 4.4: Performance Evaluations

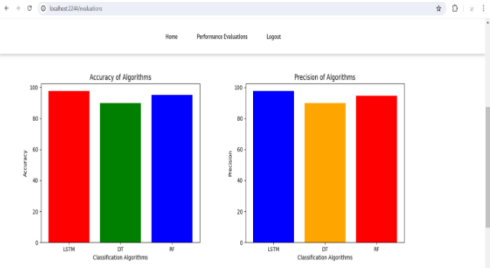


Figure 4.5: Accuracy and precision graph for algorithms

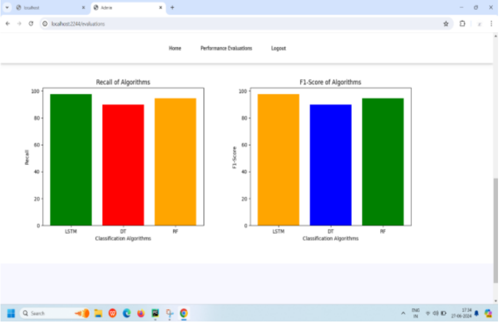
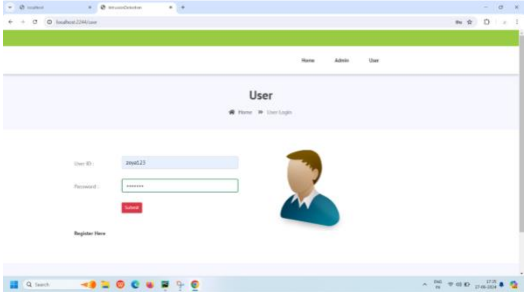


Figure 4.6: Recall and F1score graph for algorithms



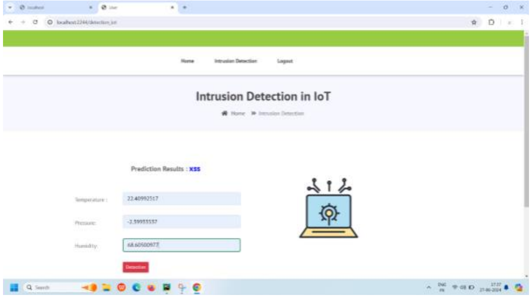
User

User ID : zoya123

Password : ••••••••

Figure 4.7: User login

Detection Page is also made up of HTML/CSS and it is used to test data for prediction of new instances.



Intrusion Detection in IoT

Prediction Results : XSS

Temperature : 22.40992517

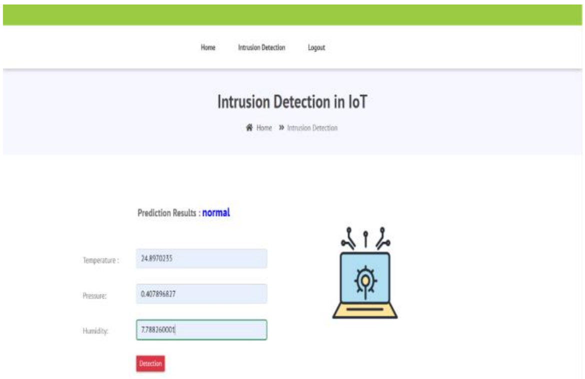
Pressure: -2.59933537

Humidity: 68.60500977

Figure 4.8: Prediction result XSS

Result 1: Specific Attack Type Detected

This intrusion detection is performed using the Long Short-Term Memory (LSTM) algorithm due to its superior performance in handling sequential data. The categorization result that was obtained provides information about the particular sort of attack that was discovered, including ransomware, DDoS, backdoors, XSS, scanning, injection, and password attacks. With respect to recognizing temporal patterns and differentiating between various forms of harmful behaviors in the Internet of Things, the LSTM model has been shown to be useful due to its high accuracy, precision, recall, and F1-score.



Intrusion Detection in IoT

Prediction Results : normal

Temperature : 24.8970235

Pressure: 0.407896827

Humidity: 7.788260001

Figure 4.9: Prediction result is normal

Result 2: Normal Class Detected
This intrusion detection is performed using the Long Short-Term Memory (LSTM) algorithm as it outperforms other methods in detecting anomalies over time. The categorization result for the provided input is 'Normal'. The LSTM model accurately identifies legitimate environmental changes, ensuring reliable differentiation from potential cyber threats.

The results obtained demonstrate the suggested IDS framework's stability and dependability, especially the LSTM algorithm, in improving intrusion detection within IoT environments by utilizing a comprehensive dataset that includes telemetry data. The multi-class classification capability of the LSTM model is especially valuable in providing detailed insights into the specific types of attacks, thereby enhancing the overall security and situational awareness of the IoT system.

5. Conclusion

In this paper, we proposed a novel framework to improve the Internet of Things (IoT) intrusion detection system (IDS) performance by leveraging the ToN-IoT telemetry dataset. By incorporating environmental sensor data alongside traditional network data, we aimed to provide a more comprehensive and accurate approach to detecting cyber attacks. The integration of various machine learning and deep learning algorithms, including Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (KNN), Gradient Boosting, and Long Short-Term Memory (LSTM), demonstrated significant improvements in the classification and prediction of attacks.According to our results, the LSTM model performed better than other algorithms because of its capacity to recognize temporal patterns in sequential data and learn from it, producing excellent accuracy, precision, recall, and F1-scores. This study provided a balanced and comprehensive detection capability, effectively reducing false positives and enhancing contextual awareness.

References

[1] Moustafa, Nour. "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets." *Sustainable Cities and Society* (2021): 102994. Public Access Here.

[2] Susilo, B.; Sari, R.F. Intrusion Detection in IoT Networks Using Deep Learning

Paper ID: SR24813173848 DOI: https://dx.doi.org/10.21275/SR24813173848

Algorithm. *Information* **2020**, *11*, 279. https://doi.org/10.3390/info11050279

[3] Banaamah, Alaa Mohammed, and Iftikhar Ahmad. 2022. "Intrusion Detection in IoT Using Deep Learning" *Sensors* 22, no. 21: 8417. https://doi.org/10.3390/s22218417

[4] Alkahtani, Hasan, and Theyazn HH Aldhyani. "Intrusion Detection System to Advance Internet of Things Infrastructure-Based Deep Learning Algorithms." *Complexity* 2021, no. 1 (2021): 5579851.

[5] Tsimenidis, S., Lagkas, T. & Rantos, K. Deep Learning in IoT Intrusion Detection. *J Netw Syst Manage* **30**, 8 (2022). https://doi.org/10.1007/s10922-021-09621-9

[6] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo and A. Robles-Kelly, "Deep Learning-Based Intrusion Detection for IoT Networks," *2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC)*, Kyoto, Japan, 2019, pp. 256-25609, doi: 10.1109/PRDC47002.2019.00056.

[7] Awajan, Albara. 2023. "A Novel Deep Learning-Based Intrusion Detection System for IoT Networks" *Computers* 12, no. 2: 34. https://doi.org/10.3390/computers12020034

[8] Booij, Tim M., Irina Chiscop, Erik Meeuwissen, Nour Moustafa, and Frank TH den Hartog. "ToN IoT-The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion datasets." *IEEE Internet of Things Journal* (2021). Public Access Here.

[9] Alsaedi, Abdullah, Nour Moustafa, Zahir Tari, Abdun Mahmood, and Adnan Anwar. "TON_IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven Intrusion Detection Systems." *IEEE Access* 8 (2020): 165130-165150.

[10] Moustafa, Nour, M. Keshk, E. Debie and H. Janicke, "Federated TON_IoT Windows Datasets for Evaluating AI-Based Security Applications," *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 848-855, doi: 10.1109/TrustCom50675.2020.00114. Public Access Here.

[11] Moustafa, Nour, M. Ahmed and S. Ahmed, "Data Analytics-Enabled Intrusion Detection: Evaluations of ToN_IoT Linux Datasets," *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 727-735, doi: 10.1109/TrustCom50675.2020.00100. Public Access Here.

[12] Moustafa, Nour. "New Generations of Internet of Things Datasets for Cybersecurity Applications based Machine Learning: TON_IoT Datasets." *Proceedings of the eResearch Australasia Conference, Brisbane, Australia*. 2019.

[13] Moustafa, Nour. "A systemic IoT-Fog-Cloud architecture for big-data analytics and cyber security systems: a review of fog computing." *arXiv preprint arXiv:1906.01055* (2019).

[14] Ashraf, Javed, Marwa Keshk, Nour Moustafa, Mohamed Abdel-Basset, Hasnat Khurshid, Asim D. Bakhshi, and Reham R. Mostafa. "IoTBoT-IDS: A Novel Statistical Learning-enabled Botnet Detection Framework for Protecting Networks of Smart Cities." *Sustainable Cities and Society* (2021): 103041.