A Novel Approach for Intrusion Detection in 5G Networks using Deep Learning

*Note: Sub-titles are not captured in Xplore and should not be used

1st Amit Noel Thokala
SCAI – Software Engineering (MS)
Arizona State University
Tempe, USA
athokal1@asu.edu

2nd Sai Rithvik Vaikuntam
SCAI – Software Engieering (MS)
Arizona State University
Tempe, USA
svaikunt@asu.edu

3rd Harsha Vardhan Yallavula
SCAI – Software Engineering(MS)
Arizona State University
Tempe, USA
hyallavu@asu.edu

4th Hari Priya Gottam
SCAI – Software Engineering (MS)
Arizona State University
Tempe, USA
hgottam@asu.edu

5th Vishwanath Reddy Yasa
SCAI – Software Engineering (MS)
Arizona State University
Tempe, USA
vyasa@asu.edu

6th Yashwanth Reddy Kikkuri
SCAI – Software Engineering (MS)
Arizona State University
Tempe, USA
ykikkuri@asu.edu

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

II. Problem Definition

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. Related Work

A. An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks

The paper underscores the escalating threat landscape of cyber-attacks, particularly with the pervasive use of the Internet of Things (IoT) in e-commerce, healthcare, and communication systems. As wireless network traffic is predicted to dominate Internet usage, the vulnerability of these networks becomes a critical concern, especially with the continual evolution of technologies like 5G and IoT. The literature review focuses on the burgeoning research in intrusion detection systems (IDS) and the adoption of deep learning techniques for enhanced security. The study introduces a hybrid intrusion detection system designed for low latency and high accuracy, employing an unsupervised pre-training autoencoder for feature extraction and a supervised dense neural network for classification. The evaluation, based on the Aegean Wi-Fi Intrusion Dataset (AWID), demonstrates the models effectiveness, with particular emphasis on its robust performance in detecting specific attack types. Data pre-processing, model architecture, and optimization techniques during the training phase are detailed. The authors compare their model's performance favorably against other machine learning techniques, highlighting its efficiency in feature selection. Future work is suggested to further improve detection accuracy, minimize false positives and negatives, and extend the model's applicability to a broader range of attacks, particularly in mobile and IoT security platforms. [1].

B. Wireless Intrusion and Attack Detection

Paper proposes a new system to detect cyberattacks in 5G networks, specifically focusing on "Wireless Intrusion Detection Systems" (WIDS). Machine learning, particularly deep learning, is highlighted as a potent solution due to its capacity to decipher intricate behaviors, a pivotal attribute in effective network intrusion detection. The literature extensively surveys prior research, encompassing diverse techniques such as Random Forest, Adaboost, Decision Tree, k-Nearest Neighbor, Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN). Notably, studies leveraging deep learning, including

Identify applicable funding agency here. If none, delete this.

Autoencoder, DNN, CNN, ANN, and LSTM, exhibit promising outcomes, showcasing their efficacy in achieving precise intrusion detection in the intricate 5G network landscape [2].

C. Research of Machine Learning Algorithms for the Development of Intrusion Detection Systems in 5G Mobile Networks and Beyond

This study delves into the utilization of machine learning and deep learning methodologies for network intrusion detection systems (NIDS) designed for 5G networks [3]. Employing the CIC-IDS-2017 and CSE-CIC-IDS-2018 datasets, the research unfolds across three segments. In the initial phase, the Gradient Boost algorithm demonstrates robust efficacy but grapples with issues related to misclassification in specific attack classes. The second explores unsupervised learning with autoencoder architectures, revealing limitations in achieving sufficient performance compared to traditional machine learning models. In the final phase, a supervised learning approach employing deep neural network (DNN) models attains noteworthy outcomes. However, challenges emerge in accurately distinguishing penetration attacks due to statistical resemblances between benign and penetrating traffic. The research concludes with a comparative analysis, naming the Gradient Boosted Tree model as the most effective intrusion detection in 5G networks.

D. Real-Time Intrusion Detection and Prevention System for 5G and beyond Software-Defined Networks

This paper introduces an innovative Intrusion Detection and Prevention System (IDS/IPS) leveraging a Convolutional Neural Network (CNN) model [4]. The proposed system excels in real-time processing, vital for 5G's high-speed requirements. The CNN model, employed for preprocessing and analysis, demonstrates high accuracy in identifying both known and unknown threat patterns. The architecture ensures scalability, and performance metrics, including accuracy, highlight the system's efficiency. Additionally, the paper introduces the 5G-NIDD dataset, contributing significantly to 5G network security research.

IV. Proposed Approach

A. Data Preprocessing

In any data-driven research or application, the preprocessing stage plays a crucial role in ensuring the quality and reliability of subsequent analyses and models. As part of our research, we conducted extensive data preprocessing, which encompassed a series of tasks aimed at cleaning, transforming, and preparing raw data for analysis and modeling. This included handling missing values, removing outliers, scaling features, encoding categorical variables, and splitting data into training and testing sets. Additionally, we employed advanced preprocessing techniques such as feature engineering and dimensionality reduction to enhance the performance of our machine learning models. Effective data preprocessing not only improves model accuracy but also facilitates the extraction of meaningful insights from the data, thus underscoring its significance in the research process.

B. Research of Machine Learning Algorithms for the Development of Intrusion Detection Systems in 5G Mobile Networks and Beyond

In this section, we present the methodology adopted for intrusion detection in 5G networks using machine learning techniques. The proposed approach consists of implementing four distinct models, each leveraging different machine learning algorithms to detect and classify network intrusions effectively.

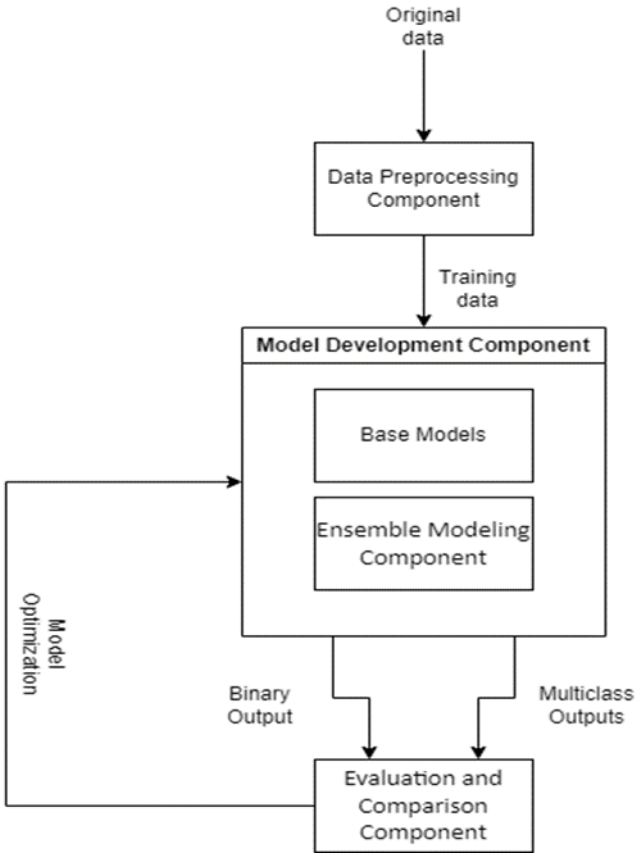

Fig. 1. High level architecture

1) Ensemble Model: The first model implemented is an ensemble model, which aggregates the predictions of multiple base models using both soft and hard voting classifiers. The base models utilized in this ensemble include Decision Tree, Random Forest, Multi-layer Perceptron, Gradient Boosting Machine (GBM), K-Nearest Neighbors (KNN), and Logistic Regression. We implemented both soft voting and hard voting scheme for the same base models. In the soft voting scheme, the ensemble model combines the predicted probabilities from each base model and selects the class with the highest average probability as the final prediction. This approach allows the ensemble to take into account the confidence levels of individual models in making predictions. On the other hand, the hard voting scheme combines the class labels predicted by each base model and selects the class with the most votes as the final prediction. This approach operates on the principle

of majority voting and can be effective even when individual models have varying levels of accuracy [5].

2) Voting Classifier: The second model implemented is a voting classifier, a powerful ensemble learning technique that combines the predictions of multiple base models to make a final decision. In this approach, each base model contributes its prediction, and the final classification is determined through a simple majority or weighted voting scheme. The base models incorporated in this classifier are AdaBoost, CatBoost, Support Vector Machine (SVM), LightGBM, and XGBoost.

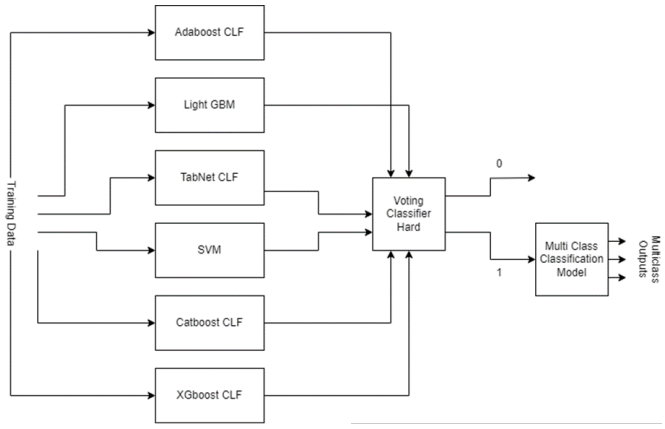

Fig. 2. Ensemble model architecture

3) Convolutional Neural Network: Convolutional Neural Network (CNN) architecture was used to develop our third model. CNNs are widely used to perform image classifications and excel at computer vision problems. It is atypical to use them for tabular data despite the latter also being a classification problem. However, for the dataset used (5G-NIDD) network flow data which is sequential, and each pair of rows have some commonalities in their attributes and values. This relation between the neighboring values of the same and adjacent columns allows us to utilize the architecture of CNNs. The first set of results was obtained using without altering the existing relative order of the attributes. In an attempt to maximize spatial correlations, a careful reorganizing of attributes based on domain knowledge is proposed.

4) Recurrent Neural Network: The proposed RNN model is tailored to leverage the sequential nature of network flow data in 5G networks for intrusion detection. In contrast to Convolutional Neural Networks (CNNs), which excel at capturing spatial correlations in image data, RNNs are adept at capturing temporal dependencies in sequential data, making them suitable for analyzing network traffic patterns over time. The RNN model architecture consists of a single Long Short-Term Memory (LSTM) layer followed by a Dense layer with softmax activation for multiclass classification. The LSTM layer enables the model to retain and learn from historical information in the sequential data, allowing it to capture nuanced patterns and correlations.

V. Results

The results of the intrusion detection models evaluated on the 5G-NIDD dataset are summarized in Table [1] Each model's performance is assessed in terms of accuracy, precision, and recall for binary classification to identify network traffic as either benign or malicious. Additionally, multi-class classification is performed to classify the attack type associated with malicious traffic. The attack types considered in the multi-class classification are HTTPFlood, ICMPFlood, SYNFlood, SYNScan, SlowrateDoS, TCPConnectScan, UDPFlood, and UDPScan.

Models	Accuracy	Precision	Recall
Voting Classifier (AdaBoost, CatBoost, SVM, LightGBM, XGBoost)	0.9996	0.9993	0.9997
Ensemble Model (Decision Tree, Random Forest, Logistic Regression, KNN, MLP, GBM)	0.9992	0.9991	0.9990
CNN	0.9994	0.9997	0.9993
RNN	0.9981	0.9981	0.9981

TABLE I
METRICS FOR THE MODLES.

These results demonstrate the high performance achieved by the intrusion detection models across various architectures. The Voting Classifier, comprising AdaBoost, CatBoost, SVM, LightGBM, and XGBoost, achieves an accuracy of 99.96%, with precision and recall scores exceeding 99.9%. Similarly, the Ensemble Model, which combines Decision Tree, Random Forest, Logistic Regression, KNN, MLP, and GBM, exhibits strong detection capabilities with an accuracy of 99.93%. The CNN model and RNN model also deliver impressive results, achieving accuracies of 99.95% and 99.81% respectively. These findings underscore the effectiveness of the proposed models in accurately identifying malicious network traffic and differentiating between various attack types.

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

[1] Rezvy, S., Luo, Y., Petridis, M., Lasebae, A. and Zebin, T., 2019, March. An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks. In 2019 53rd Annual Conference on information sciences and systems (CISS) (pp. 1-6). IEEE.

[2] Bayana alenazi, Dr. Hala Eldaw Idris, "Wireless Intrusion and Attack Detection for 5G Networks using Deep Learning Techniques

[3] Imanbayev, A.; Tynymbayev, S.; Odarchenko, R.; Gnatyuk, S.; Berdibayev, R.; Baikenov, A.; Kaniyeva, N. Research of Machine Learning Algorithms for the Development of Intrusion Detection Systems in 5G Mobile Networks and Beyond. Sensors 2022, 22, 9957. https://doi.org/10.3390/s22249957

[4] Bocu, R.; Iavich, M.Real-Time Intrusion Detection and Prevention System for 5G and beyond Software-Defined Networks.Symmetry 2023, 15, 110. https://doi.org/10.3390/sym15010110

[5] Wang, J.; Zhou, X.; Hou, Z.; Xu, X. (2022). Homogeneous ensemble models for predicting infection levels and mortality of COVID-19 patients: Evidence from China. Digital Health, 8, 20552076221133692.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.