Version number and datePresenter, company name and addressTitle and subtitleDecorative sidebar

Table of Contents

[Project Description 2](#_heading=h.gjdgxs)

[Overview: 2](#_heading=h.30j0zll)

[Key Requirements:](#_heading=h.1fob9te) 3

[Deliverables:](#_heading=h.3znysh7) 4

[Acronyms and Abbreviations :](#_heading=h.2et92p0) 5

[Design and Architecture](#_heading=h.tyjcwt) 6

[Implementation Strategy](#_heading=h.3dy6vkm) 10

[High-level Work Breakdown Structure :](#_heading=h.1t3h5sf) 10

[Schedule / Timeline :](#_heading=h.4d34og8) 18

[Required Hardware :](#_heading=h.2s8eyo1) 19

[Required Software :](#_heading=h.17dp8vu) 19

[Third Party Content / Libraries :](#_heading=h.3rdcrjn) 20

[Quality:](#_heading=h.26in1rg) 21

[Other Special Considerations:](#_heading=h.lnxbz9) 21

[Process](#_heading=h.35nkun2) 22

[Process Description and justification](#_heading=h.1ksv4uv) 22

[Tools :](#_heading=h.44sinio) 22

[Roles and Responsibilities:](#_heading=h.2jxsxqh) 23

[Location of Project Artifacts:](#_heading=h.z337ya) 24

[Sponsor Communications:](#_heading=h.3j2qqm3) 24

# **Project Description**

## **Overview:**

* **Statement of work:** Our topic for the capstone project is “ Detecting intrusion in softwarized 5G networks using machine learning”. Our sponsor, Dr. Abdallah Moubayed, assigned our team a research paper, “5D-NIDD: A comprehensive network Intrusion detection dataset generated over 5G Wireless network”, which provides a detailed exposition of the creation, characteristics, and analysis of a novel dataset, 5G-NIDD, developed for network intrusion detection research in 5G environments. The dataset is distinguished by its generation from a real, operational 5G test network, offering a unique and realistic benchmark for evaluating AI/ML-based security solutions. This dataset addresses the scarcity of comprehensive, current, and realistic datasets for network security research, especially in the context of the emerging 5G networks, characterized by their complexity, scale, and diverse service offerings.
* **Final Deliverables**: Our object is to perform various data processing techniques, such as preprocessing the 5G-NIDD dataset by removing all the null and zero columns to improve the accuracy and performance of the machine learning models. With this preprocessed data, we must apply various correlation techniques to extract the relevant features and normalize the dataset with extracted features. After this, we feed our processed dataset to the machine-learning models and analyze their performance and accuracy. We have to achieve a better result than what is presented in the assigned paper.
* **Business Domain:** The business domain of our research is network security within the context of 5G wireless technology. We must address the critical need for sophisticated and proactive security mechanisms to identify and mitigate intelligent attacks on the complex and heterogeneous infrastructure of 5G networks. The increasing reliance on 5G technology across various sectors, including critical infrastructure, healthcare, and IoT, underscores the importance of this research domain
* **Technical Platform:** As our work is primarily based on a dataset, 5G-NIDD, which is generated from a real 5G test network located in Oulu, Finland, which is part of the 5G Test Network Finland (5GTN). This network provides a realistic environment for generating network traffic, both benign and malicious, under controlled conditions. The dataset includes network flows derived from various attack scenarios, such as Distributed Denial of Service (DDoS) attacks, ICMP floods, UDP floods, and HTTP floods, among others, alongside normal traffic. Tools and technologies include Argus for flow data generation, Tracewrangler for data processing, and various attack simulation tools like hping3 and GoldenEye. As per our requirement, we only have to utilize the generated dataset and feed it to the machine learning models after processing it.
* **Motivation of work:** Building upon the research described in "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network," our motivation of work focuses on several vital enhancements and expansions to further the field of network security, particularly within the rapidly evolving 5G and future 6G environments such as the expansion of threat models, dynamic and real-time data labeling, incorporation of AI/ML for active defense mechanisms, distributed detection and federated learning, testbed and dataset as a service and research beyond 5G security research.

## **Key Requirements:**

* **Functional Requirements:**
* Data Preprocessing: The solution must be capable of preprocessing the 5G-NIDD dataset by
  + Identifying and removing all null and zero-value columns to ensure data quality.
  + Applying correlation techniques to identify and extract relevant features from the dataset.
  + Normalizing the dataset based on the extracted features prepares it for machine learning analysis.
* Machine Learning Model Integration: Implement machine learning models that can be trained on the preprocessed 5G-NIDD dataset to detect network intrusions.
  + The models must support classification tasks to distinguish between benign and malicious network traffic.
  + The solution should enable model performance evaluation using accuracy, precision, recall, and F1 score metrics.
* Performance Analysis: The system must include functionality to analyze and report the performance of the machine learning models.
  + Comparison against baseline results presented in the assigned research paper to determine improvement.
  + Detailed model performance analysis under various attack scenarios mentioned in the dataset, such as DDoS, ICMP floods, UDP floods, and HTTP floods.
* **Technical Requirement:**
* Dataset Handling: The technical platform must support efficient handling of the 5G-NIDD dataset, including:
  + Secure storage and access mechanisms for the dataset.
  + Capability to process large volumes of network flow data generated from an actual 5G test network.
* Machine Learning Tools and Libraries: Utilize appropriate machine learning libraries and frameworks compatible with the dataset's characteristics and the project's goals.
  + The chosen libraries must support the preprocessing techniques, feature extraction, normalization, and model evaluation criteria outlined in the functional requirements.
* Scalability and Extensibility: Design the solution with scalability in mind to accommodate future enhancements, such as
  + Expanding the dataset with additional attack scenarios or more complex network environments.
  + Integrating additional machine learning models or deep learning approaches for enhanced detection capabilities.
* **Performance Requirement:**
  + Accuracy Improvement: The machine learning models must achieve a higher accuracy in detecting network intrusions than the baseline presented in the research paper.
  + Specific accuracy targets may be set based on preliminary analysis and the performance of existing models documented in the research paper.
  + Processing Time: The preprocessing and analysis processes should be optimized for efficiency.
  + Define maximum acceptable processing times for data preprocessing and model training phases.
  + Resource Utilization: Ensure that the solution is optimized for minimal and efficient use of computational resources, considering the complexity and volume of the 5G-NIDD dataset.

## **Deliverables:**

* Sprint 1:
  + Develop an in-depth understanding of the assigned research paper, “5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network.”
  + Understanding the reference papers on 5G and 6G intrusion detection systems will help you understand the requirements to improve the performance and accuracy of machine learning models.
* Sprint 2:
  + Perform Data Preprocessing on the 5G-NIDD dataset, such as finding correlation coefficient, input sampling, and data normalization.
  + Apply boosting algorithms to the processed data and analyze the accuracy.
* Sprint 3:
  + Exploration of Feature selection on the provided 5G-NIDD dataset.
  + Improve Boosting model performance by feeding processed data into the boosting algorithm.
  + Implement ANN (Artificial Neural Networks) based model to analyze the accuracy and performance.
  + Implement a GAN (Generative Adversarial Networks) based model to analyze the accuracy and performance.
* Sprint 4 (till date):
  + Implement feature selection and input sampling techniques on the dataset.
  + Explore and Implement various boosting algorithms using multi-classification.
  + Implement a GAN-based model by applying multi-classification.
  + Implement ANN models using TensorFlow and modify the metrics.
* Final Deliverables:
  + This is a comprehensive report detailing the data preprocessing steps, feature extraction, and normalization techniques, machine learning models used, their performance analysis, and a comparison with the baseline results.
  + A software solution that includes all the code for data preprocessing, machine learning model implementation, performance analysis, and documentation for setup, configuration, and usage.

## 

## **Acronyms and Abbreviations:**

| **Glossary of Abbreviations and Acronyms** | **Definition** |
| --- | --- |
| 5G-NIDD | 5th Generation Network Intrusion Detection System |
| 5GTN | 5G Test Network |
| AdaBoost | Adaptive Boosting |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Networks |
| CatBoost | Categorical Boosting |
| DoS | Denial of Service |
| GAN | Generative Adversarial Networks |
| IDS | Intrusion Detection System |
| ML | Machine Learning |
| ReLU | Rectified Linear Unit |
| XgBoost | Extreme Gradient Boosting |

Table 1: Acronyms and Abbreviations

# 

# 