# 5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network

Source: https://arxiv.org/abs/2212.01298

The paper "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network" addresses the increasing complexity and sophistication of 5G networks, which introduces new vulnerabilities. The introduction outlines the motivation for creating a new dataset for network intrusion detection specifically tailored for 5G networks, highlighting the technological advancements and security challenges unique to 5G. The paper then summarizes existing threat detecting mechanisms and why those mechanisms are inefficient in 5G networks, thus establishing the need for 5G-specific intrusion detection datasets. The paper than goes on to explain the methodology used to create the 5G-NIDD dataset and presents a detailed analysis of the dataset, including statistical information, the variety of included attack types, and the dataset's relevance for machine learning models. The authors evaluate the effectiveness of the dataset by testing various machine learning algorithms for intrusion detection, showcasing the dataset's utility in improving detection accuracy.

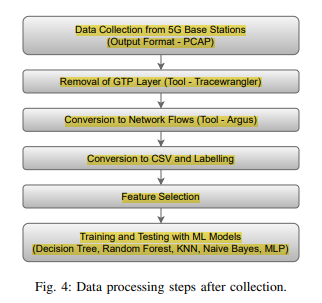
* As complexity of networks grow higher, threat surface also becomes broader
* Conventional reactive security approaches where the resolution action starts after the attack is detected, are insufficient against such intelligent attacks
* Some of the well-known publicly available datasets are outdated and have limited applicability in 5G network security research
* This article publishes 5G-NIDD, a network intrusion detection dataset generated from a real 5G test network
* The authors created 5G-NIDD, a labeled network intrusion detection dataset that contains nine intrusion types along with benign network traffic
* The authors also provide extensive evaluation of 5G-NIDD using multiple ML techniques, present the accuracy levels and the validity of the dataset in network intrusion detection
* The authors mainly evaluated few variants of DoS attacks and Port Scan attacks.

DoS Attacks

* ICMP Flood
* UDP Flood
* SYN Flood
* HTTP Flood
* Slowrate DoS

Port Scans

* SYN Scan
* TCP Connect Scan
* UDP Scan



The authors converted the flow data in argus format with 112 features into Comma Separated Values (CSV) format. The data collected at each attack session was available as separate CSV files.

Details on the analysis performed on the dataset using different ML models

ML Algorithms used:

* Decision Tree
* Random Forest
* K-Nearest Neighbor
* Naïve Bayes
* Multi Layer Perceptron

# Distributed Learning-Based Intrusion Detection in 5G and Beyond Networks

C. Park, K. Park, J. Song and J. Kim, "Distributed Learning-Based Intrusion Detection in 5G and Beyond Networks," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 490-495, doi: 10.1109/EuCNC/6GSummit58263.2023.10188312.

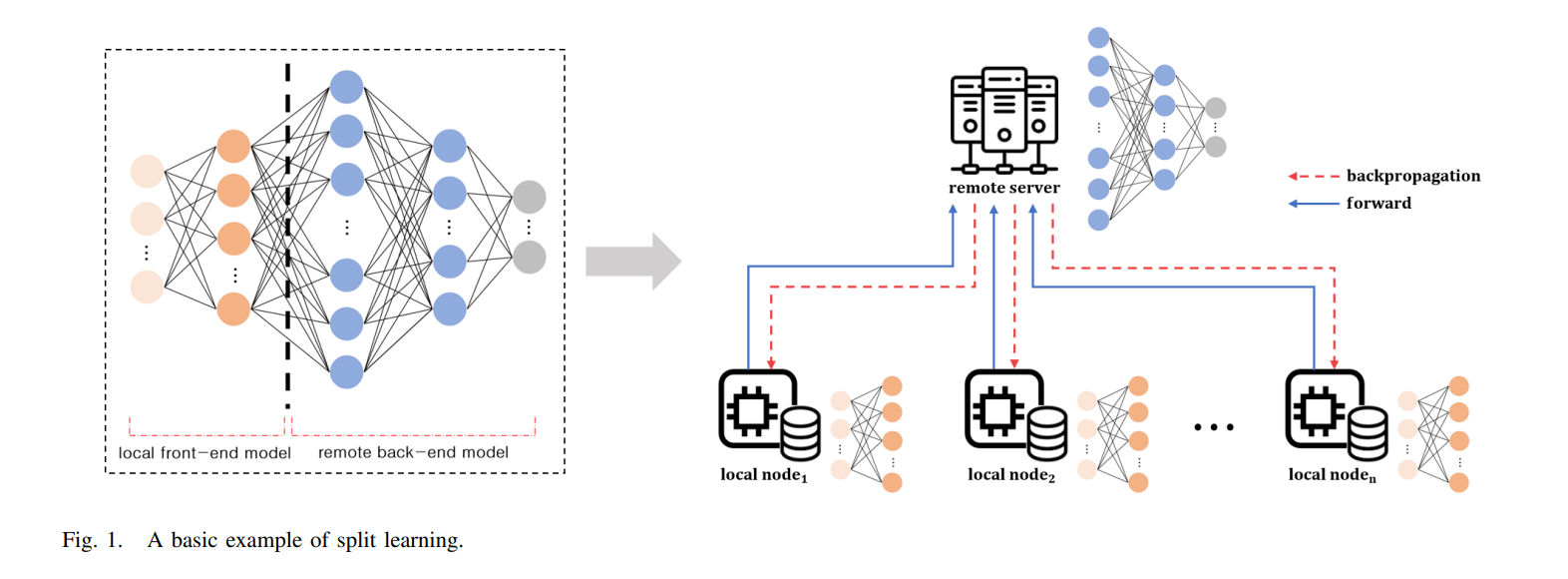
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10188312&isnumber=10188221>

* 5G-advanced and 6th generation mobile networks are evolving and are getting more complex. Similarly, cyberattacks have become more sophisticated and have rapidly increased in frequency.
* Many studies have been conducted on network intrusion detection systems (NIDS) to detect cyberattacks in advance.
* However, most of these studies concentrate on centralized environments and may not be suitable for deployment in distributed systems
* In this paper, authors propose a distributed learning-based network intrusion detection system that is capable of training data in a decentralized environment.

The authors use **Split Learning**, a learning system that which enables distributed learning that takes into account the different computing power of each node.

* allows for models to be trained in distributed systems with different computing resources
* The authors particularly focus on **SplitNN** model.

The experiments are conducted on **5G-NIDD dataset**.



Split learning is a machine learning technique that enables distributed learning in a decentralized environment while preserving privacy, and is designed to address the limitations of traditional Federated Learning, such as the uneven distribution of computational workloads.

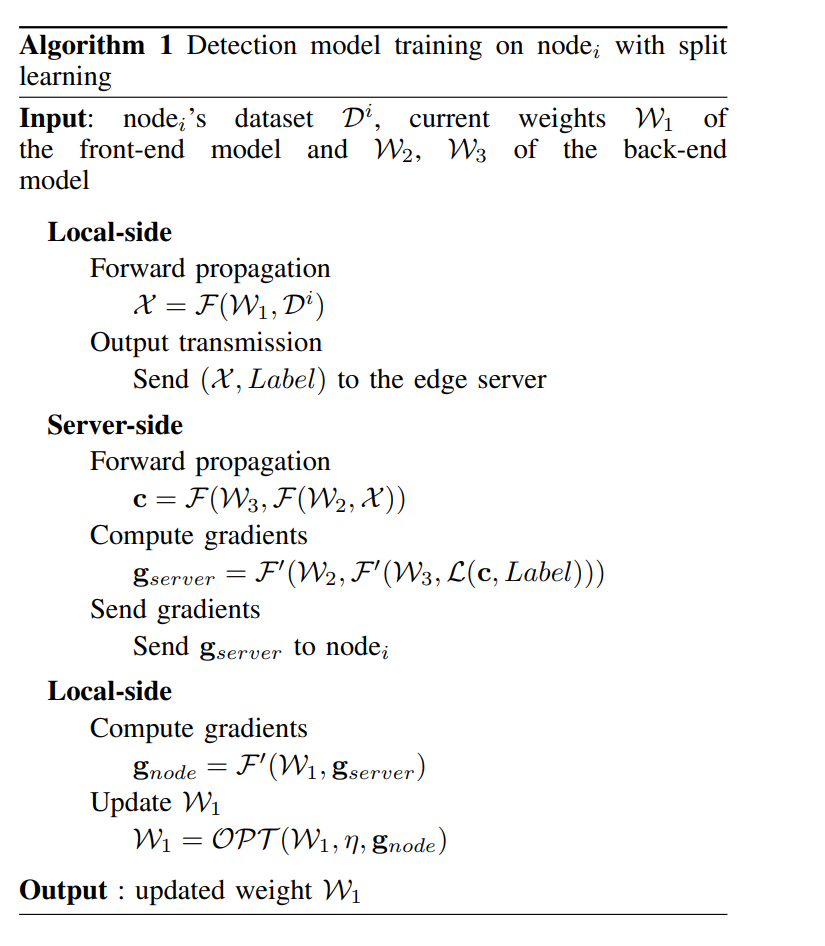
To know more about split learning:

<https://medium.com/@minhanh.dongnguyen/a-gentle-introduction-on-split-learning-959cfe513903#:~:text=In%20split%20learning%2C%20a%20deep,part%20in%20the%20collaborative%20training>.

<https://piyubasu.medium.com/split-learning-vs-federated-learning-and-their-use-cases-fe96d5c1394a>

Proposed methodology

For the predictive model, we combined a deep neural network (DNN) and a convolutional neural network (CNN) with split learning to build intrusion detection systems.



Results:

* in a distributed environment where the data occurs uniformly within the population, split learning approaches can have similar performance to centralized models
* In the case of the data imbalance scenario, the models showed that the performance of intrusion detection was sufficiently preserved when compared to the baseline and the ideal models
* we observed that the loss increases dramatically while learning on a particular node with a very small amount of data, but other nodes with sufficient data seemed to solve these problems immediately
* Through these experiments, authors confirmed that the split learning approach can be applied efficiently to detect network intrusion in a distributed mobile environment

# Securing 5G Networks with Federated Learning and GAN

Source: <https://thesis.unipd.it/handle/20.500.12608/46216>

The main objective of the thesis was to secure the access point of the 5G network in this federated learning environment. This was accomplished by placing an Intrusion Detection System at the endpoint which would classify the data as either benign or malicious. The effectiveness of this model was checked by simulating a malicious user and conducting certain adversarial attacks to determine if the model could defend against them. The study was conducted by performing two specific attacks i.e. Label-Flipping attack and Generative Adversarial Networks.

The author concludes that the federated learning environment cannot properly defend against adversarial machine learning.

The author, for his research, uses the 5G-NIDD dataset. This dataset is novel as a real network flow for 5G traffic did not exist prior to the release of this dataset. The two attacks that this dataset contains are the different variants of the DoS and port scan attacks. The author explains the data preprocessing steps. The categorical data in the 112 features were converted into numerical data through one hot encoding. For feature selection (to find relevant features) the author utilized these approaches, Pearson Correlation and ANOVA F-Scores.

