Development of the Testbed for Testing Deep Learning Based IDS System for 5G Network

Roman Odarchenko [1,3], Azamat Imanbayev[2], *and* Alla Pinchuk[1]

[1]*National Aviation University, Lyubomyra Gusara Ave 1, Kyiv, 03058, Ukraine*
[2]*Kazakh National University, al-Farabi Street 71, Almaty, 050040, Kazakhstan*
[3]*Bundleslab KFT, Vali u, 4. 4. em. 2. ajto, Budapest, 1117, Hungary*

Abstract

In modern conditions, due to the huge emerging landscape of new cyber threats, global risks and given the most widespread connection of heterogeneous devices to the network via cellular communication networks, the issues of ensuring the necessary level of cybersecurity in these networks are becoming a priority that needs to be addressed as soon as possible. Therefore, in this research, the main attention is paid to the development of IDS for 5G networks based on artificial intelligence. In particular, this work is devoted to the development of the concept of the system, the analysis of existing datasets for its training and the development of the most appropriate test architecture of 5G network or testing the trained AI-based IDS for 5G networks.

To build a test network, various options for using open-source solutions were analyzed in detail, among which preference was given to OpenAirInterface. For this architecture, the integration of the developed IDS will be the easiest and most expedient. Also, it will be relatively easy to generate the necessary types of attacks on the network and the corresponding traffic analysis.

Keywords 1

5G, 6G, IDS, Artificial Intelligence, Deep Learning, Machine learning, datasets, testbed

1. Introduction

Compared to the present, in 2025, we will be using 13 times more data. It is anticipated that by 2025, there will be 21 billion devices connected to the Internet, in contrast to the current 7 billion [1]. Many of these new devices will manage and control our homes, urban infrastructure, transportation, and other aspects of our lives. Beyond 2030, wireless applications will necessitate significantly higher data rates (up to 1 Tbps), extremely low end-to-end latency (<1 ms), and exceedingly high end-to-end reliability (99.99999%) [2]. However, this remarkable digital evolution is only achievable through the advent of the new generation of 5G and 6G mobile networks. Following the introduction of 5G technology, academia and industry have begun exploring the potential of 6th generation (6G) wireless network technology. The progression of mobile network technologies, particularly with the emergence of new 5G services and architectures, presents novel challenges in the realm of security and user privacy protection. The threat landscape is rapidly evolving, and attacks can originate from any point of connection. Smart Cities, which will inevitably be constructed on the foundation of next-generation cellular networks, are particularly vulnerable as they constitute critical infrastructure. Moreover, the aim is to extend the capabilities of mobile communication beyond what previous generations could achieve. Several potential technologies are predicted to underpin 6G networks, including both existing and future technologies such as post-quantum cryptography, artificial intelligence (AI), machine learning (ML), advanced edge computing, molecular communication, THz, visible light

MoMLeT+DS 2023: 5th International Workshop on Modern Machine Learning Technologies and Data Science, June 3, 2023, Lviv, Ukraine
EMAIL: odarchenko.r.s@ukr.net (R. Odarchenko); imanbaevazamat@gmail.com (A. Imanbayev); pinchuk.ad87@gmail.com (A. Pinchuk)
ORCID: 0000-0002-7130-1375 (R. Odarchenko); 0000-0003-3719-4091 (A. Imanbayev); 0000-0003-3567-0445 (A. Pinchuk)

communication (VLC), and distributed ledger (DL) technologies like blockchain [3]. From a security and privacy standpoint, these advancements necessitate a reevaluation of conventional security practices. Authentication, encryption, access control, communications, and malicious activity detection must meet the heightened demands of future networks. Additionally, new approaches to security are required to ensure reliability and privacy. The escalating need for information security in cellular networks, owing to the proliferation and diversification of cyber attacks, serves as a prerequisite for such efforts.

2. Background analysis

In recent years, there has been an increasing focus on research regarding testbeds for simulating cyberattacks within the cybersecurity community. Numerous cyber range solutions have been proposed, including NCR [4], DETERLab [5], SimSpace [6], EDURange [7], CYRA [8], KYPO [9], and CyRIS [10]. Many of these initiatives have centered around the development and integration of models, tools, and methodologies for defining simulation rules, as well as offering practical guidance for conducting attack simulations [11]. While early cyber range systems were physical, more recent proposals have shifted towards virtual environments, which reduces costs and enhances flexibility [12].

However, to the best of our knowledge, there have been no studies on cyberspace systems specifically designed for 5G cybersecurity testing. Our objective is to propose a system that provides a fully virtualized 5G network. To overcome the limitations of existing cyberspaces, our research suggests a new testbed capable of simulating a comprehensive 5G network within a virtual environment. This approach allows for simplified configuration without the need for complex hardware components.

One of the shortcomings of existing cyberspaces is the lack of publicly available datasets and mechanisms for generating high-precision synthetic data. Given that 5G networks face various security challenges and threats, several papers have proposed using ML/DL methods to automatically detect malicious network traffic. However, current proposals require significant improvements in terms of real-time detection and analysis of potential threats [13].

Furthermore, current proposals largely overlook the integration of ML models in a fully automated manner and the verification of their functionality in operational environments [14]. We propose the use of Intrusion Detection Systems (IDS) for 5G mobile networks, which can monitor traffic in real-time and identify abnormal patterns.

3. Problem statement

Currently, commercial 5G networks are being rapidly deployed in many countries worldwide, while research and development efforts for enhancing cellular networks towards 6G are ongoing. In the present landscape, with the emergence of numerous new cyber threats and global risks, coupled with the extensive connectivity of diverse devices through cellular communication networks, ensuring a sufficient level of cybersecurity in these networks has become a crucial and pressing task.

Hence, the primary objective of this paper is to enhance the security systems of previous generations of cellular networks, their individual components, and the mechanisms for detecting cyber incidents. To accomplish these objectives, the following tasks need to be undertaken:

- development of the IA-based IDS for 5G concept;
- development and deployment of the 5G testbed for testing the IDS;
- development of appropriate software and its testing on a real cellular network.

4. Datasets for traffic analysis

The greatest challenge lies in obtaining reliable access to attack detection systems. While the necessary data can be acquired through network monitoring, most datasets are not publicly

disclosed due to security and privacy concerns. Moreover, gathering information online can be a costly endeavor. As a result, developers seek to manage their networks or systems using the available datasets. Malowidzki et al. [13] and Haider et al. [15] highlight the complexity of the issue surrounding the lack of attack analysis datasets and the associated requirements for compilation. Additionally, these datasets are utilized to assess the accuracy of identifying attacks. The quality of the data directly impacts the results generated by the network intrusion detection system. In recent years, the cybersecurity community has made efforts to address this challenge, leading to the publication of several sets of intrusion detection data. This section will explore commonly used datasets for Intrusion Detection Systems (IDS), taking into account the advantages and disadvantages of existing datasets.

4.1. Overview of existing datasets for Traffic Analysis in Mobile Networks

Common properties serve as evaluation criteria to enable a meaningful comparison of statistics (FAIR Concepts [16]). This is because each task or scenario is associated with a specific set of data records. For instance, the ISCX dataset [17] is chosen to emphasize labeling, while the UGR'16 dataset [18] is selected for its ability to capture long-term effects. Figure 1 illustrates four principles that are closely linked to the aforementioned concept:

	Name	Description
F	Findable	1.(meta)data are assigned A globally unique and eternally persistent identifier. 2.(meta)data are registered or index in a searchable resource. 3.Metadata specify the data identifier.
A	Accessible	1.(meta)data are retrievable by their identifier using a standardized communication protocol. 2.The protocol is open, 3 end universally implementable. 3.The protocol allows for an authentication and authorization procedure, where necessary. 4.Metadata are accessible, even when the data are no longer available.
I	Interoperable	1.(meta)data use a formal, accessible, shared and broadly accessible language for knowledge representation. 2.(meta)data use vocabularies that follow FAIR principles. 3.Metadata include qualified references to other data.
R	Reusable	1.(meta)data save plurality of accurate and relevant attributes. 2.(meta)data are released with the clear and accessible data usage license. 3.(meta)data are associated with their provenance. 4.(meta)data meet domain relevant community standards.

Figure 1: FAIR

Given this concept, this article describes a collection of attack detection data. For the purpose of this work, several new and popular datasets were selected (see Fig. 2).

Abbreviation	Dataset Name
KDD-CUP'99	KDD Cup99
DARPA99	DARPA 1999 TP Dump Files
NSL-KDD	NSL KDD Dataset
ISX2012	Information Security Center of Excellence 2012 evaluation dataset
ADFA2013	ADFA Linux Dataset (ADFA-LD) and ADFA Windows Dataset (ADFA-WD)
UNSW-NB15	University of New South Wales IDS Dataset
CIC-IDS2017	Canadian Institute for Cybersecurity IDS dataset
CSE-CIC-IDS2018	A Realistic Cyber Defense Dataset by Communications Security Establishment (CSE) and The Canadian Institute for Cybersecurity

Figure 2: Network Intrusion Detection System datasets

4.1.1. KDD-CUP'99

KDD'99 was created in 1999, in which the functions are divided into four groups, such as:

- basic functions
- content features
- time-based traffic features
- host-based traffic features

Although the number of records is quite large, amounting to 4,898,430 records, four types of attacks (DoS attacks, R2L, U2R, and Probe) dominate the dataset (see Fig. 3).

Category	Whole KDD	Corrected KDD	10% KDD
DoS	3 883 370	229 853	391 458
Probe	41 102	4 166	4 107
R2L	1 126	16 347	1 126
U2R	52	70	52
Normal	972 780	60 593	97 277
Total	4 898 430	311 029	494 020

Figure 3: Number of samples in KDD'99

Many studies and projects have been carried out during our lifetime. However, there are two main issues with this: duplicate records and obsolete records (i.e. no new attacks are included), which means that each trained model is subject to a partial learning algorithm. Therefore, the algorithm does not detect rare attacks in the record.

4.1.2. NSL-KDD

To solve the problem with the KDD-Cup'99 dataset, a new dataset was created called NSL-KDD. In fact, it is almost identical to its predecessor, except for a few advantages. Firstly, the issue of duplicate copies of the training dataset is addressed, avoiding biased results. Secondly, there are no duplicate instances in the test data set, which allowed the researchers to make more practical applications without random sampling.

Since this dataset is formed from the previous dataset, it also has 42 functions about different connections. However, this set is not suitable for the network IDS model because it does not have public data. Number of samples in NSL-KDD can see Fig. 4.

Category	KDDTrain+20%	KDD Train+	KDDTest+
DoS	9 234 (37%)	45 927 (37%)	7 458 (33%)
Probe	2 289 (9.16%)	11 656 (9.11%)	2 421 (11%)
R2L	209 (0.8%)	995 (0.85%)	2 654 (12.1%)
U2R	11 (0.04%)	52 (0.04%)	200 (0.9%)
Normal	13 449 (53%)	67 343 (53%)	9711 (43%)
Total	25 192	125 973	22 544

Figure 4: Number of samples in NSL-KDD

4.1.3. UNSW-NB15

The UNSW-NB15 dataset was created in 2015 at the University of New South Wales. There are over 2.5 million records. The UNSW-NB15 database contains 49 network connectivity indicators that can be divided into 6 groups:

- Basic features;
- Time features;

– Content features;
– Connection features;
– Additional features;
– Two features for the class label.

A complete list of attacks in the dataset and their quantities is given in Fig. 5. Each entry contains information about which of the ten classes of the connection belongs to, regular connections or one of nine different types of attacks. UNSW-NB15 introduces new IDS datasets and is used in several recent studies.

№	Type	Quantity	Description
1	Normal	2 218 761	Natural Transaction Data
2	Fuzzers	24,246	Attempt to cause suspension of a program or network by feeding it randomly generated data.
3	Analysis	2,677	Contains various port scanning, spam and HTML file intrusion attacks.
4	Backdoors	2,329	A technique in which the system's security mechanism is bypassed unnoticed to access a computer or its data.
5	DoS	16,353	A malicious attempt to make a server or network resource inaccessible to users, usually a temporary interruption or suspension of the services of a host connected to the Internet.
6	Exploits	44,525	The attacker knows about security problems in the system and uses these vulnerabilities for his own purposes.
7	Generic	215,481	The technique works against all block ciphers (with a given block and key size), regardless of the structure of the block cipher.
8	Reconnaissance	13,987	Contains all types of attacks that collect information about the network.
9	Shellcode	1,511	A small piece of code used as a payload when exploiting software vulnerabilities.
10	Worms	174	The attacker replicates himself to spread to other computers. Often it uses a computer network for distribution, relying on security failures on the target computer access it.

Figure 5: Types of attacks in UNSW-NB15

4.1.4. CICIDS2017

CICIDS2017 is a new data collection developed by the Canadian Institute for Cyber Security. The dataset includes the latest network attacks, but also meets all the criteria for enhanced specific attacks in the ISCX2012 dataset [19]. Since the launch of CICIDS2017, this database has attracted researchers to analyse and develop new models and algorithms. However, the best detection models need to detect all types of attacks, thus traffic data should be combined throughout the day to create a single data set that uses IDS to produce a typical IDS.

The data source contains an ML file with 8 CSV files. These files contain information about the types of attacks over a five-day period, including normal traffic. In some files, binary classification is convenient, in others you need to create multi-class definition templates. The files containing the CICIDS-2017 data set are presented in Fig. 6.

№	File name	Contained attacks
1	Monday-WorkingHours.pcap_ISCX.csv	Benign
2	Tuesday-WorkingHours.pcap_ISCX.csv	Benign, FTP-Patator, SSH-Patator
3	Wednesday-WorkingHours.pcap_ISCX.csv	Benign, DoS GoldenEye, DoS Hulk, DoS Slowhttptest, DoS Slowloris, Heartbleed
4	Thursday-WorkingHours-MorningWebAttacks.pcap_ISCX.csv	Benign, Web Attack -Brute Force, Web Attack -SQL Injection, Web Attack -XSS
5	Thursday-WorkingHours-AfternoonInfiltration.pcap_ISCX.csv	Benign, Infiltration
6	Friday-WorkingHours.pcap_ISCX.csv	Benign, Bot
7	Friday-WorkingHours-AfternoonPortScan.pcap_ISCX.csv	Benign, PortScan
8	Friday-WorkingHours-AfternoonDDos.pcap_ISCX.csv	Benign, DDoS

Figure 6: CICIDS2017 dataset files

It offers different types of attacks based on a 2016 McAfee report and is publicly available. Complete data set format with 2,830,743 instances with 15 class tokens (1 normal and 14 attacks) and 79 features (78 features and 1 for attack type tokens).

4.1.5. CSE-CIC-IDS2018

CSE-CIC-IDS2018 is the latest version of the CSE Intrusion Detection dataset, which collected 10 days of network traffic. It has been extended due to the criteria used to create CIC-IDS201.

It has a similar structure to CICIDS2017 but is built around a large network of simulated users and attacking machines. The main purpose of this dataset is to train and predict models to detect insecure traffic based on anomalies. As with many network attack data sets, the data sets have class imbalances. Data-level algorithms or methods can be used to eliminate this problem. The attack

scenarios with percentage distribution are shown in Fig. 7 below.

Class Label	Quantity	Volume
Benign	2 856 035	63.111%
Bot	286 191	6.324%
DoS	1 289 544	28.497%
Brute Force	513	0.011%
Infiltration	93 063	2.056%
SQL injection	53	0.001%
Total	4 525 399	100%

Figure 7: CSE-CIC-IDS2018 attack scenarios

4.2. Issues with IDS benchmark datasets

As mentioned earlier, there are a sufficient number of available datasets for training and predicting network intrusion detection systems. However, only a limited number of these datasets contain relevant types of attacks and features that are practical for implementing models. This section discusses the main challenges faced by researchers working on intelligent intrusion detection systems.

First and foremost, collecting reliable research data is extremely difficult. Technology evolves rapidly, with security threats and new attacks constantly emerging. Consequently, datasets quickly lose their relevance and value within the cybersecurity community. Another challenge is data integrity. Researchers need to incorporate not just CSV files, but also audit logs and raw network data. Audit logs provide valuable information about cyberattacks, while raw data enhances threat detection capabilities.

The evolving landscape of attacks poses another challenge. As technology advances, hackers adapt their attack methods to current systems and software, leading to the emergence of new and deprecated attack types. This creates a perpetual cycle. To address this, researchers can either employ new datasets or utilize dataset generators that simulate hacker behavior and create appropriate attacks.

Furthermore, the generated datasets must be as realistic as possible to be applicable in practical network environments. This means including normal traffic from various end-user workstations and servers. Otherwise, the trained model may not be suitable for a specific computer network. Privacy considerations also come into play when working with datasets. While an organization's computer networks are the most trusted sources of data, they are often unwilling to share their audit logs or network logs due to privacy policies. As a result, researchers mostly rely on popular datasets, which are modeled data rather than real network traffic data.

The need for labeling is crucial in both supervised and unsupervised learning approaches. Labels are essential for calculating the accuracy of the employed algorithms. Experts typically gather secure network activity in cyberspaces before launching attacks on network traffic. They first establish normal traffic patterns and then introduce attacks. Some experts inject attacks into normal traffic, while others manually label the data, which is a more laborious process.

Additionally, the establishment of dataset criteria plays an important role. Markus Ring [20] discusses common aspects of dataset descriptions and classifies them into five categories.

Finally, for a dataset to be valuable, it must gain broad acceptance within the research community. Without this support, the dataset may only be utilized in a limited number of research projects.

5. AI-based IDS for 5G concept

Currently, the implementation of intelligent analytics is essential across various wireless networks, ranging from local networks to remote clouds. Network traffic prediction and estimation

are critical for network operations and management, including congestion management, routing, resource allocation, service level agreement management, and other network responsibilities [21]. Therefore, the utilization of machine learning (ML) and artificial intelligence (AI) will play a significant role. In a hierarchical order, we can consider the following:

- Artificial intelligence serves as a crucial component in comprehending vast amounts of data. It finds application in various areas, such as data preparation, data retrieval, data flow visualization, geospatial tracking, and real-time tracking.
- Machine learning tackles the challenge of dealing with large volumes of data in 5G networks. It utilizes specialized algorithms that enable computers to learn and adapt. As the size of data on the network grows due to connected sensors, traditional methods of tracking and identifying patterns become insufficient. Machine learning surpasses conventional data analysis approaches by analyzing data from multiple sources and establishing logical connections among them.

Future opportunities for 5G networks encompass reliable analysis, network optimization, and improved efficiency in business solutions. Automation of the anomaly detection (AD) process is another potential benefit, as it can significantly contribute to operational and management systems. Consequently, this reduces the number of false positives and enhances the understanding of the underlying cause of anomalies. Figure 8 illustrates the automated anomaly analysis process.

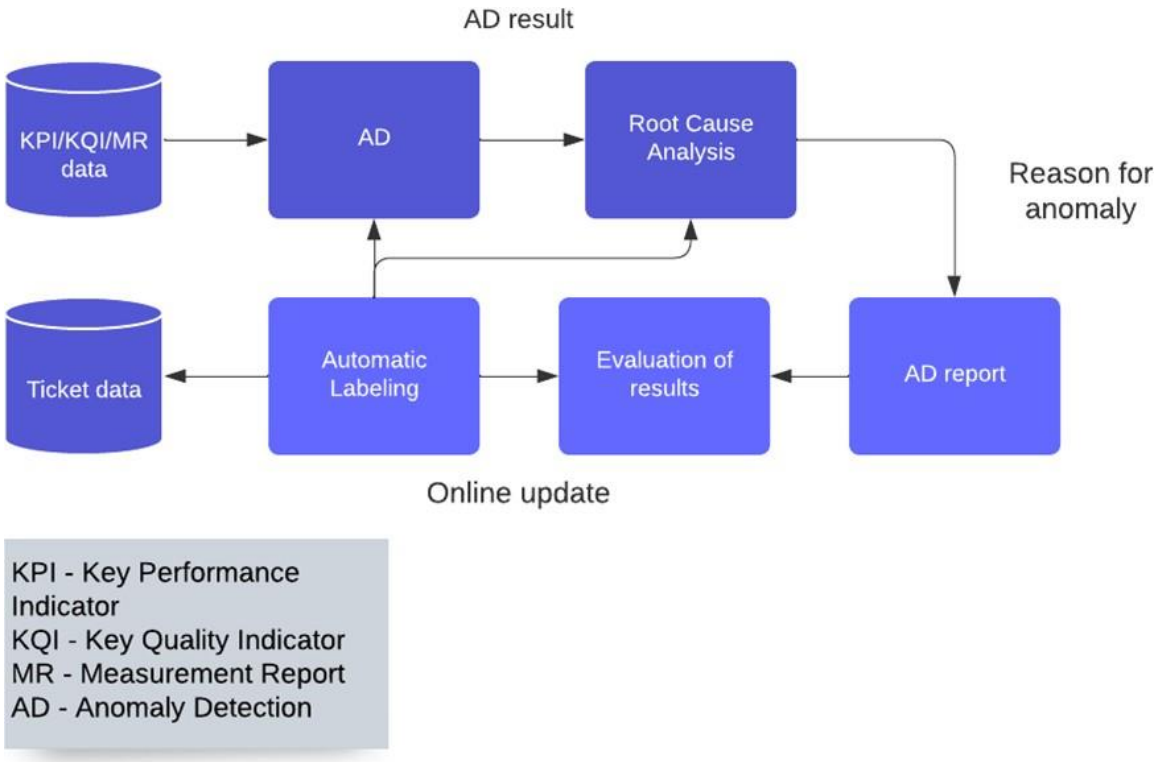

Figure 8: Working process for automatic anomaly analysis

In this regard, it is necessary to develop new models and methods of Internet traffic in 5G networks, which will become the main types of networks connecting existing networks. Since each IoT application is characterized by individual network traffic parameters and the principle of interaction between physical and virtual Internet objects, it is necessary to develop models and methods for the interaction of IoT applications in 5G networks.

Despite significant progress, the 5G specification provides mobile operators with some guidance on how to ensure that their 5G networks truly support AI/ML. Much remains to be done to further simplify AI/ML-enabled networks and implement the core concepts of AI and ML as underlying network structures, including:

- Individual system of collecting detailed data;

- Demonstration of RAN capabilities for optimizing user networks and services;
- RAN programming, RAN service architecture supporting AI/ML;
- Data transfer and analysis/implementation of AI/ML to enable effective innovation;
- Open datasets in wireless networks to accelerate the development of algorithms and new AI applications in wireless networks.

As networks evolve beyond 5G, artificial intelligence could become an integral part of the overall blueprint for a holistic approach to managing this complex system.

The basic concept of using a network access discovery system, developed using ML and DL methods, includes the following three main steps, as shown in Fig. 9.

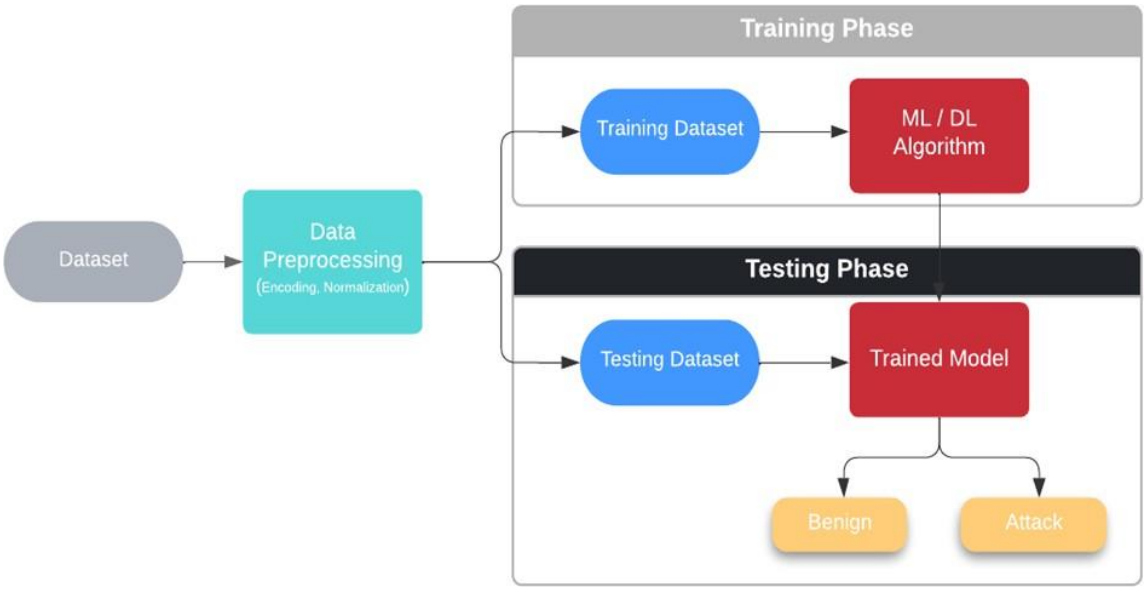

Figure 9: NIDS methodology based on machine learning/deep learning

This figure shows relatively recent work on NIDS using machine and deep learning. After a brief analysis of the works [13, 22-24], it can be seen that in lots of them traditional machine learning algorithms have been used, and deep learning is still in its early stages of development. Although NIDS has been extensively studied, the most significant changes have only occurred in the content of the data set, which contains information about attack patterns.

Machine learning algorithms such as K-Nearest Neighbor (CNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Means Clustering, Fast Learning Network have already been used by others.

For example, in the research paper [25] presented six different machine learning models (Decision Tree, Random Forest, K Nearest Neighbors, Adaboost, Gradient Boosting, and Linear Discriminant Analysis) using one of the latest datasets, namely CSE-CIC-IDS2018. SMOTE was used by multiplying the data from the minority group to reduce unbalanced factors. Overall, their studies were able to improve the accuracy of the model from 4.01% to 30.59%.

At the same time, Yao et al. [26] proposed a multilevel structure of the IDS model called Multilevel Semi-Supervised ML (MSML) which also uses the RF model. The main idea is to redirect to the next model if it is not checked. The experimental results showed the advantage of the model in detecting attacks even on small samples in the dataset.

The next work is devoted to the ANN-based intrusion detection model. This algorithm is different in that it can perform non-linear simulations using large amounts of data. However, this model has drawbacks: it takes a long time, slows down the learning process, and makes the solution inefficient. To solve this problem, Huang et al. [27] suggested using the Extreme Learning Machine (ELM) as it has a direct connection to one of the hidden layers which analytically determines the output weights . The author's research became the basis for the work of other researchers such as Li et al. [28] suggested the idea of using the Fast Learning Network. The problem was that this algorithm was based on a parallel connection of a multi-layer neural network and a single-layer

forward neural network.

Ali and others [29] continued the Fast Learning Network model idea in the KDD Cup'99 dataset using a particle swarm optimization known as PSO-FLN. After comparing the performance of their model with other optimization algorithms, they concluded that their model was superior, showing an increase in the number of neurons in the hidden layer. Despite good results, this model suffered from low accuracy for lower attack classes. Also, some deep learning models have been built due to their efficiency and autonomy of learning important features of the dataset. A notable example is the research paper by Naseer et al. [30], who compared different DL and ML models. As a result, Deep CNN and LSTM outperformed the rest.

A lot of work has been done with AutoEncoder (AE) in intrusion detection systems. For reference, this is one of the most popular deep learning methods as it matches the best features of the dataset. There are several subtypes of AE such as Stacked, Sparse and Variational AE. Shone et al. [31] proposed an IDS based on the deep AE method and ML RF. Their work was successful in terms of computation and time, using only the AE coding part to make it work asymmetrically. Experiments were performed on two datasets such as KDD Cup '99 and NSL-KDD. Compared to Alrawashdeh et al. [32] and their deep trust network models, the author's model was better, although not useful for detecting R2L and U2R attacks. This was due to the selection of a dataset that did not contain such cases.

Yang et al. proposed using the Stacked Sparse Autoencoder (SSAE) to extract high-level feature representations from intrusive behavioral information [33]. As a result, they found that high dimensional sparse features are more discriminatory for intrusion behavior than previous methods, and the base classification process is greatly accelerated by using high dimensional sparse features. While this model provides adequate detection rates for U2R and R2L attacks, it is still lower than other dataset classes. The same methodology was followed by the authors of [34] using AE and SVM. The results show an overall performance improvement over other DL and ML models.

A review of the literature shows that more work is needed to characterize the features of network attacks. After all, by defining a common pattern, it is possible to provide high accuracy for all attacks in the dataset. Many also use outdated datasets that do not have the new attack spectrum. In addition, research using class imbalance techniques to prevent infrequent attacks on datasets is limited.

So, it is critical to implement an effective security system to protect the 5G network from these threats. One such security system is the Intrusion Detection System (IDS).

An IDS is a software or hardware system that monitors network traffic for signs of malicious activity or policy violations. The IDS is designed to detect and alert network administrators to any suspicious behavior that could compromise the integrity, confidentiality, or availability of a network. Identifiers may be implemented in various parts of the network such as user equipment (UE), radio access network (RAN), and core network (CN).

The implementation of IDS in the 5G network is of paramount importance due to the significant increase in the number of connected devices and the amount of data that is transmitted over the network. With the advent of the Internet of Things (IoT), a huge number of devices with different levels of security are connected to the network, making it vulnerable to attacks. In addition, the 5G network is expected to support critical applications such as autonomous vehicles, remote healthcare, and industrial automation that require a high level of security and reliability.

IDS can help mitigate the security risks associated with 5G by providing real-time detection and alerting of malicious activity. Identifiers can detect various types of attacks, including distributed denial of service (DDoS), malware, and intrusion attempts. With IDS, network administrators can quickly respond to security incidents and take appropriate action to prevent further damage.

In conclusion, the implementation of IDS is critical to the security and reliability of the 5G network. As the number of connected devices and the amount of data transferred over the network increases, security risks also increase. An IDS can help detect and mitigate these risks by providing real-time detection and alerting of malicious activity. Therefore, it is important to prioritize the implementation of IDS in the 5G network to ensure the security and reliability of the network.

Software-defined security can be used to create an intelligent core network intrusion detection system, given that two key components of the 5G network - the RAN and the core network - are fully virtualized and defined by software. This makes it possible to develop an automated security

system in which copies of the traffic from the reverse connection and the core network are sent to the SDS for analysis, as described in [35]. It is important to note that traffic copies do not have any impact on network performance during the analysis phase. However, before determining whether the traffic is anomalous or not, pre-processing is required to ensure that the data is suitable for use with machine learning or deep learning models. Anomalies can then be analyzed using appropriate algorithms and the results stored in the Policy Manager database. These results are then sent to the VNF manager, which updates the IDS module. The time it takes to process the model plays a vital role in the presentation of the end results and determines when the template should be run to keep the module's policies up to date. Using this method, you can automate detection, update the database of attacks, and take the necessary actions to protect your network from intrusions.

Physically, this concept can be implemented in the form of a Cyber incident response platform for 5G cellular networks, which will receive data on the security status, cyber idents directly from various network nodes (Fig. 3).

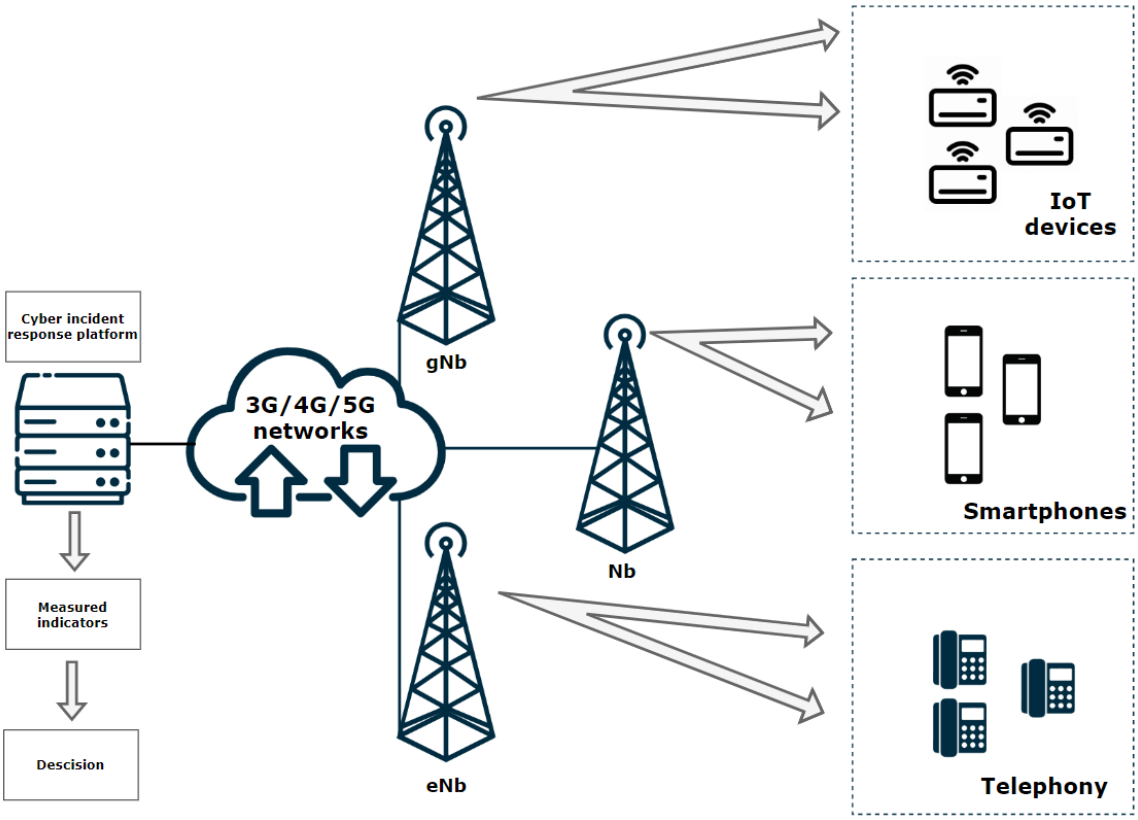

Figure 10: Place of IDS in 5G cellular networks

IDS presented on Figure 10, was described in more details in [13]. It was only trained, but not tested in real environment. That is why it was decided to deploy 5G network testbed for conducting all the necessary tests.

6. 5G network testbed development and deployment

Now let's move on to the development of a test bench for revalidation of the IDS in 5G measure (Fig. 10).

Also, some logical blocks can be used for dermal use of them in the same tools, which were examined during the research.

First of all, the general scheme of the network was development. Thus, the network contains the following components: CORE (network core), RAN (Radio Access Network), MEC (Multi-access Edge Computing), and UE (User Equipment) (Fig.11).

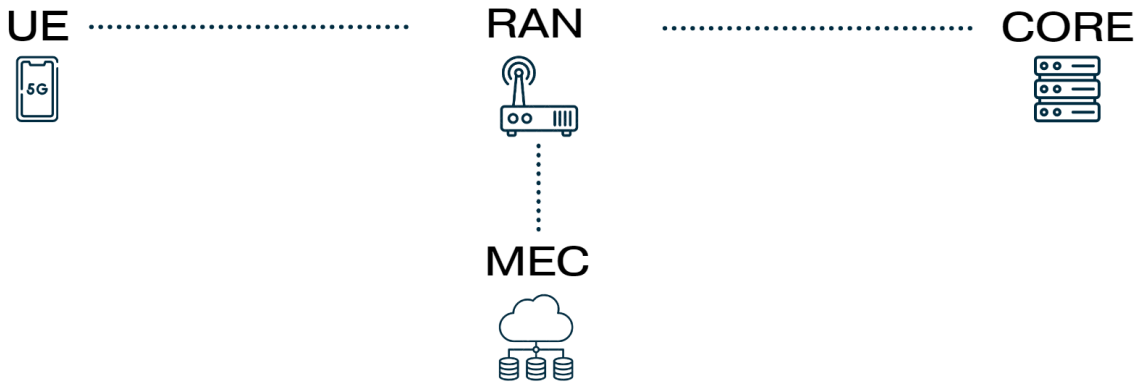

Figure 11: The general scheme for 5G testbed deploying

After analyzing different open-source projects [36] for deploying a 5G network, the following software was chosen:

- Network core: SD-CORE (ONF)
- RAN: openairinterface5G (OAI)
- SMO (Service management and orchestration): Aether ROC (ONF)

Currently, a possible open-source project for MEC deployment was not considered.

In the context of open-source, hardware was considered. It was decided to deploy a network core on the virtual machine on the server with Ubuntu OS, 16GB RAM, 512GB ROM, 16 CPU, and Intel Core processor. Because SDR needed a USB connection with RAN, a PC was chosen. It is also on Ubuntu OS and Intel Core processor, but 8GB RAM, 256GB ROM, and 6 CPU. As a transceiver, SDR Ettus USRP B210 was selected. On the Figure 12 showed a scheme that includes hardware and software.

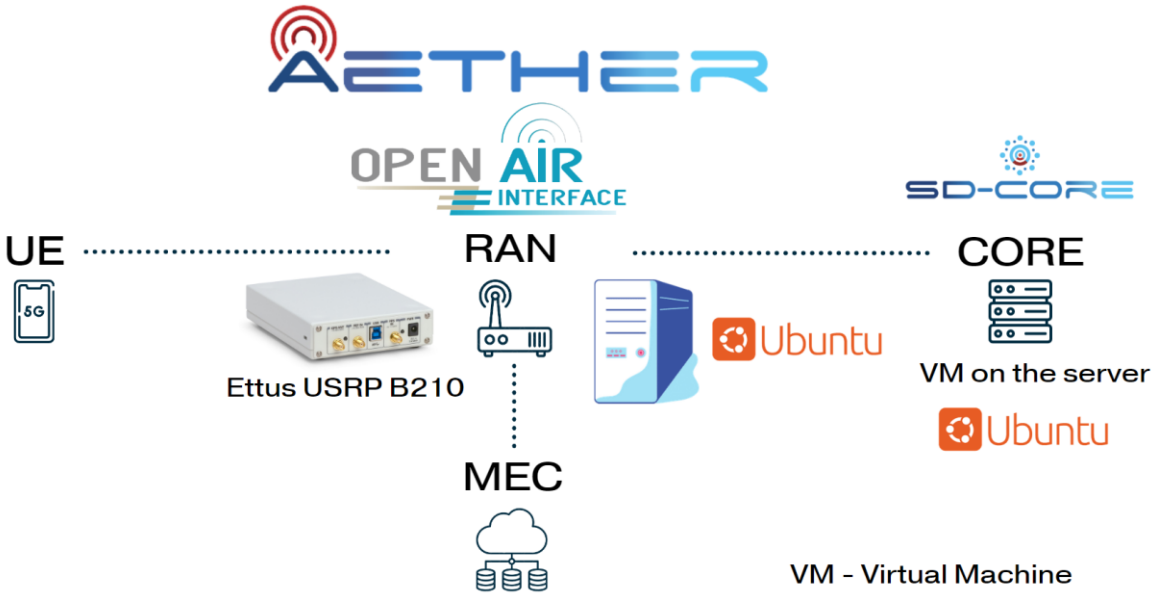

Figure 12: Hardware and software scheme

The full scheme includes communication between all deployed network components, IP addresses, software and hardware (Fig. 13).

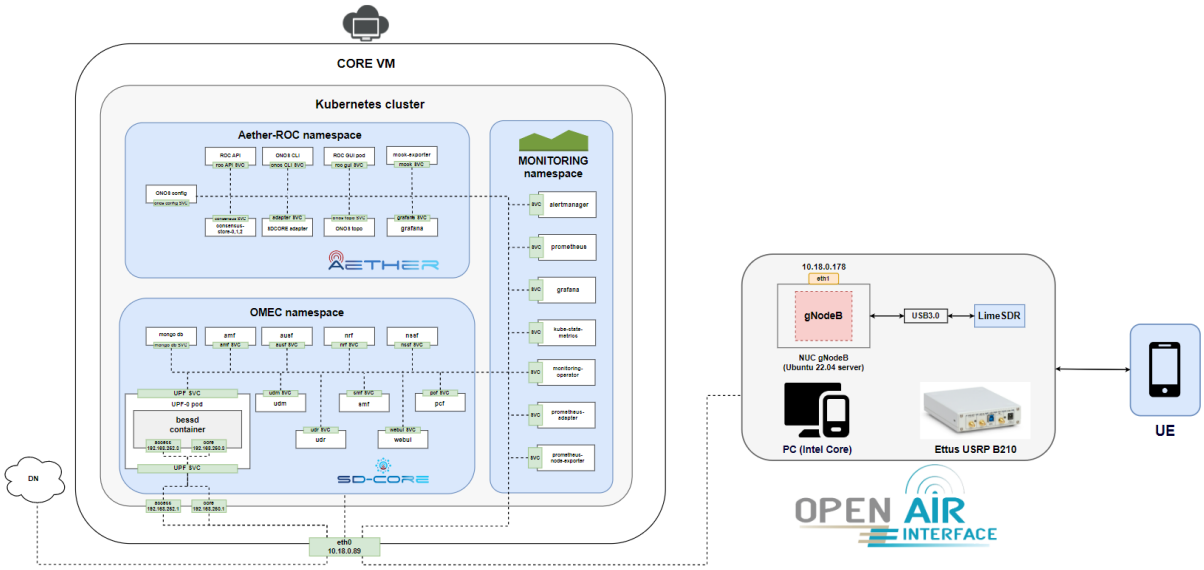

Figure 13: The full deployment scheme

Additionally, a monitoring system Grafana was deployed.

Before using the network for our goals, it was tested with a simulator of UE and gNodeB. For this, the gNBSim [37] project was chosen. In this case, the network core was tested for serviceability. Results of testing showed on Figure 14. During test, five profiles were tested:

- UE registration procedure;
- PDU session establishment (Registration + UE initiated PDU Session Establishment + User Data packets);
- UE deregistration procedure (Registration + UE initiated PDU Session Establishment + User Data packets + Deregister);
- AN release (Registration + UE initiated PDU Session Establishment + User Data packets + AN Release);
- UE trigger service request (Registration + UE initiated PDU Session Establishment + User Data packets + AN Release + UE Initiated Service Request).

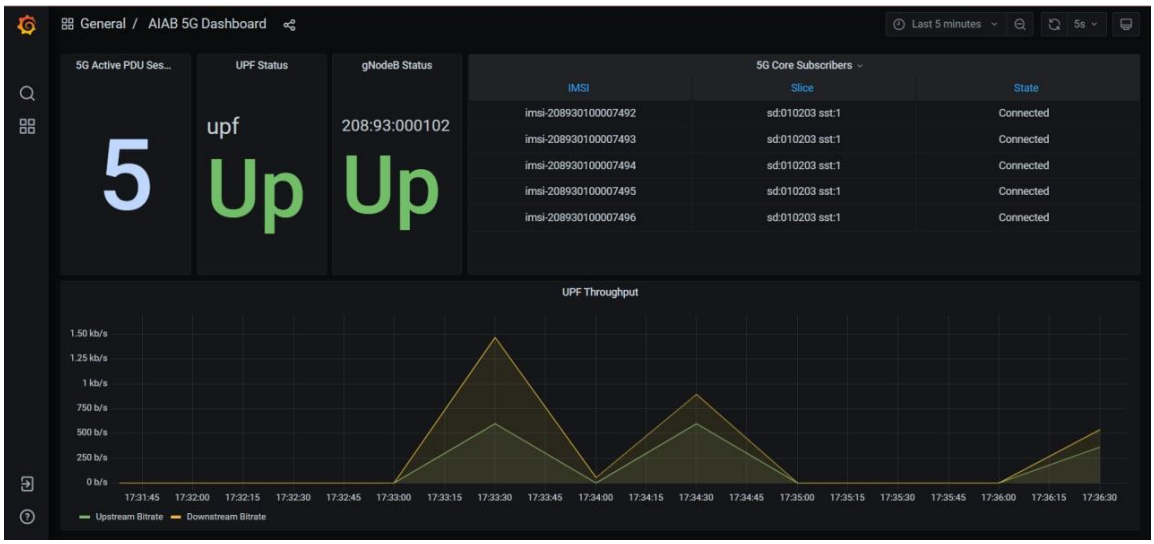

Figure 14: Successful 5G network test

LTE. As we do not currently have the required SDR, similarly, an LTE network scheme was developed and selected software and hardware (Fig.15).

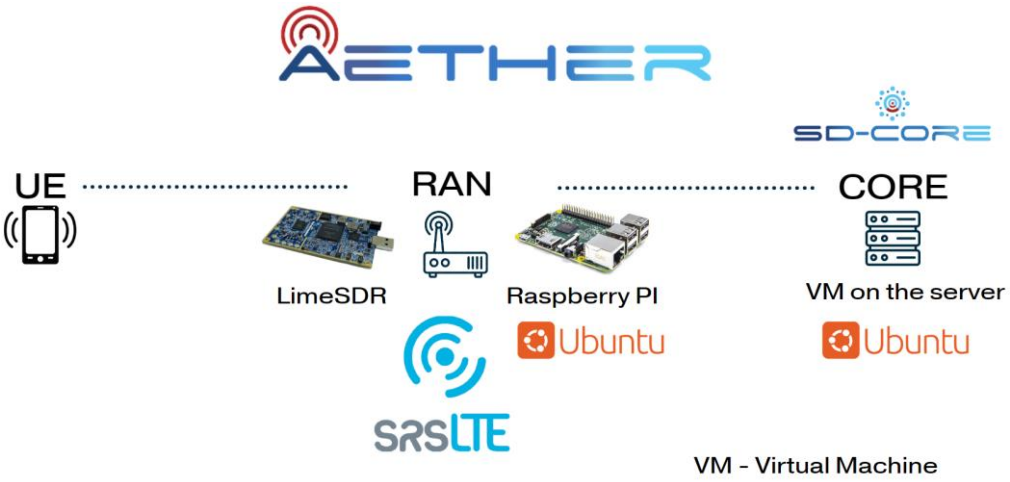

Figure 15: LTE deployment scheme

Thus, for network core uses SD-CORE (EPC), and for RAN uses srsLTE (eNodeB). Regarding hardware, VM for EPC, Raspberry Pi for RAN, were used, and as a transceiver uses LimeSDR.

The similar deployment scheme that includes communication between all deployed network components, IP addresses, software and hardware on the Figure 16.

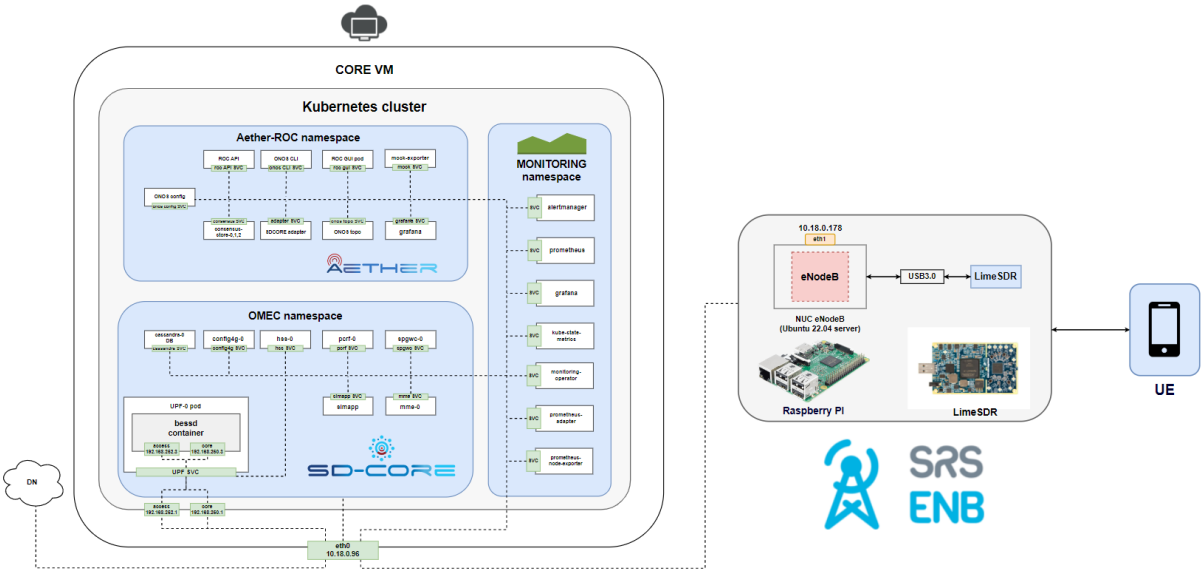

Figure 16: The full deployment scheme

Figure 17: The view of deployment scheme

After configuration of RAN and EPC, a smartphone was prepared for connection (SIM-card programming and APN configuration). Figure 18 showed a successful UE attaching.

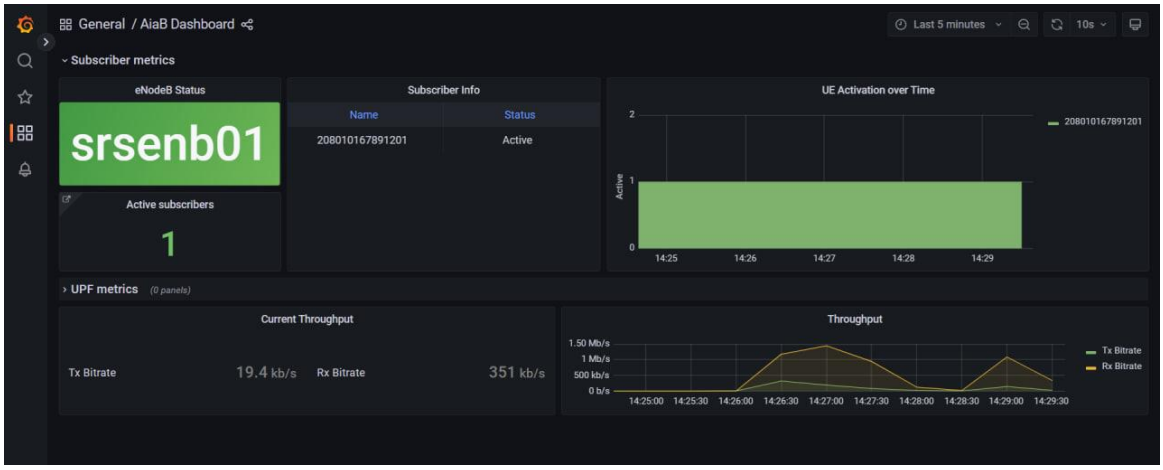

Figure 18: UE attach

On the monitoring we can see the following information about UE: IMSI, active time, throughput by uplink and downlink.

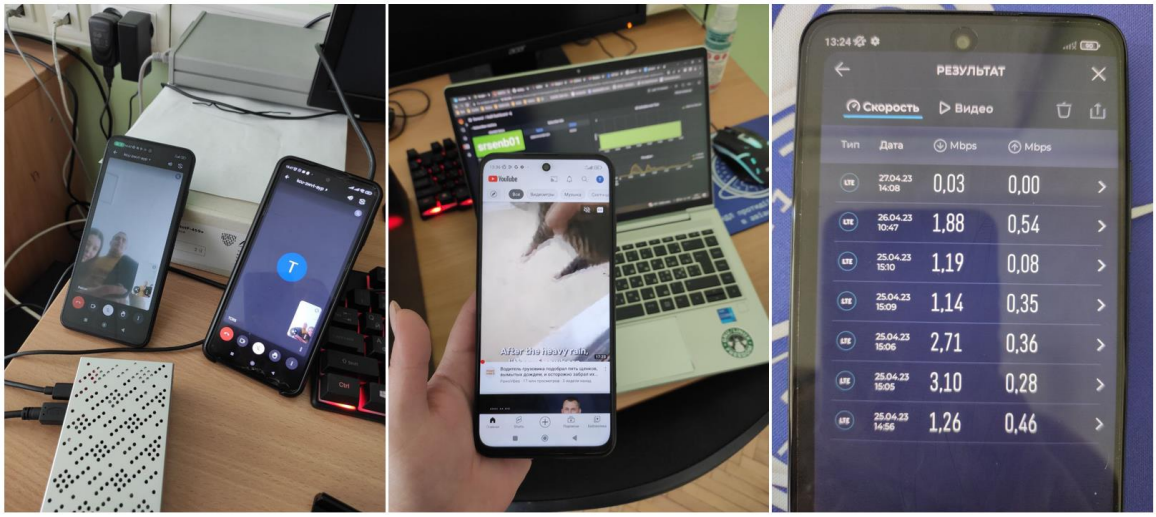

Figure 19: UE attach

7. Conclusions

Currently, commercial 5G networks are being extensively deployed in numerous countries worldwide, while research and development efforts for advancing cellular networks towards 6G are ongoing. In the present era, with the rapid emergence of new cyber threats and global risks, combined with the widespread connection of diverse devices through cellular communication networks, ensuring the requisite level of cybersecurity in these networks has become an urgent priority.

Therefore, in this and previous studies, significant emphasis has been placed on the development of an Intrusion Detection System (IDS) for 5G networks based on artificial intelligence. This work aims to develop the system concept, analyze existing datasets for training purposes, and devise a suitable test architecture for evaluating the AI-based IDS specifically designed for 5G networks.

To construct the test network, various options utilizing open-source solutions were meticulously examined, with a preference given to OpenAirInterface. This choice ensures seamless and practical integration of the developed IDS into the architecture. Additionally, generating the required types of attacks on the network and conducting corresponding traffic analysis will be relatively

straightforward.

Future research endeavors will focus on assessing the performance of the developed system within the established testbed infrastructure.

8. Acknowledgements

This work was supported in part by the European Commission under the 5GASP: 5G Application & Services experimentation and certification Platform (H2020 – ICT-2020, grant agreement ID: 101016448). The views expressed in this contribution are those of the author and do not necessarily represent the project.

9. References

[1] The importance of 5G technology. URL: https://securecommunications.airbus.com/en/meet-the-experts/the-importance-of-5g-technology.

[2] W. Saad, M. Bennis, M. Chen, A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems, IEEE Netw. 34.3 (2020) 134–142. doi:10.1109/mnet.001.1900287.

[3] S. A. Abdel Hakeem, H. H. Hussein, H. Kim, Security Requirements and Challenges of 6G Technologies and Applications, Sensors 22.5 (2022) 1969. doi:10.3390/s22051969.

[4] Telecom Storefront Solutions. URL: https://www.ncr.com/telecom-technology/telecom-storefront-solutions

[5] A. Ibrahim, V. Ford, Observations, Evaluations, and Recommendations for DETERLab from an Educational Perspective, J. Cybersecur. Educ. Res. Pract. № 1 (2021).

[6] C. Jian, H. Shi, B. Krieg-Brückner, SimSpace: A Tool to Interpret Route Instructions with Qualitative Spatial Knowledge, y: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems, Stanford, 2009, p. 47–48.

[7] R. Weiss, F. Turbak, J. Mache, M. E. Locasto, Cybersecurity Education and Assessment in EDURange, IEEE Secur. & Priv. 15.3 (2017) 90–95. doi:10.1109/msp.2017.54.

[8] M. Smyrlis, I. Somarakis, G. Spanoudakis, G. Hatzivasilis, S. Ioannidis, CYRA: A Model-Driven CYber Range Assurance Platform, Appl. Sci. 11.11 (2021) 5165. doi:10.3390/app11115165.

[9] J. Vykopal, R. Oslejsek, P. Celeda, M. Vizvary, D. Tovarnak, KYPO Cyber Range: Design and Use Cases, y: 12th International Conference on Software Technologies, SCITEPRESS - Science and Technology Publications, 2017. doi:10.5220/0006428203100321.

[10] C. Pham, D. Tang, K.-i. Chinen, R. Beuran, CyRIS, y: SoICT '16: Seventh International Symposium on Information and Communication Technology, ACM, New York, NY, USA, 2016. doi:10.1145/3011077.3011087.

[11] M. M. Yamin, B. Katt, V. Gkioulos, Cyber ranges and security testbeds: Scenarios, functions, tools and architecture, Comput. & Secur. 88 (2020) 101636. doi:10.1016/j.cose.2019.101636.

[12] N. Chouliaras, G. Kittes, I. Kantzavelou, L. Maglaras, G. Pantziou, M. A. Ferrag, Cyber Ranges and TestBeds for Education, Training, and Research, Appl. Sci. 11.4 (2021) 1809. doi:10.3390/app11041809.

[13] A. Imanbayev, S. Tynymbayev, R. Odarchenko, S. Gnatyuk, R. Berdibayev, A. Baikenov, N. Kaniyeva, Research of Machine Learning Algorithms for the Development of Intrusion Detection Systems in 5G Mobile Networks and Beyond, Sensors 22.24 (2022) 9957. doi:10.3390/s22249957.

[14] S. Rawat, A. Srinivasan, V. Ravi, U. Ghosh, Intrusion detection systems using classical machine learning techniques vs integrated unsupervised feature learning and deep neural network, Internet Technol. Lett. (2020). doi:10.1002/itl2.232.

[15] W. Haider, J. Hu, J. Slay, B. P. Turnbull, Y. Xie, Generating realistic intrusion detection system dataset based on fuzzy qualitative modeling, J. Netw. Comput. Appl. 87 (2017) 185–192. doi:10.1016/j.jnca.2017.03.018.

[16] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N.

Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, та ін., The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3.1 (2016). doi:10.1038/sdata.2016.18.
[17] A. Shiravi, H. Shiravi, M. Tavallaee, A. A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput. & Secur. 31.3 (2012) 357–374. doi:10.1016/j.cose.2011.12.012.
[18] G. Maciá-Fernández, J. Camacho, R. Magán-Carrión, P. García-Teodoro, R. Therón, UGR‘16: A new dataset for the evaluation of cyclostationarity-based network IDSs, Comput. & Secur. 73 (2018) 411–424. doi:10.1016/j.cose.2017.11.004.
[19] Datasets | Research | Canadian Institute for Cybersecurity | UNB. URL: https://www.unb.ca/cic/datasets/index.html.
[20] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, A. Hotho, A survey of network-based intrusion detection data sets, Comput. & Secur. 86 (2019) 147–167. doi:10.1016/j.cose.2019.06.005.
[21] A. R. Abdellah, O. A. K. Mahmood, A. Paramonov, A. Koucheryavy, IoT traffic prediction using multi-step ahead prediction with neural network, y: 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), IEEE, 2019. doi:10.1109/icumt48472.2019.8970675.
[22] M. Pawlicki, M. Choraś, R. Kozik, Defending network intrusion detection systems against adversarial evasion attacks, Future Gener. Comput. Syst. 110 (2020) 148–154. doi:10.1016/j.future.2020.04.013.
[23] C. Zhang, P. Patras, H. Haddadi, Deep Learning in Mobile and Wireless Networking: A Survey, IEEE Commun. Surv. & Tutor. 21.3 (2019) 2224–2287. doi:10.1109/comst.2019.2904897.
[24] M. Yao, M. Sohul, V. Marojevic, J. H. Reed, Artificial Intelligence Defined 5G Radio Access Networks, IEEE Commun. Mag. 57.3 (2019) 14–20. doi:10.1109/mcom.2019.1800629.
[25] G. Karatas, O. Demir, O. K. Sahingoz, Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset, IEEE Access 8 (2020) 32150–32162. doi:10.1109/access.2020.2973219.
[26] H. Yao, D. Fu, P. Zhang, M. Li, Y. Liu, MSML: A Novel Multilevel Semi-Supervised Machine Learning Framework for Intrusion Detection System, IEEE Internet Things J. 6.2 (2019) 1949–1959. doi:10.1109/jiot.2018.2873125.
[27] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70.1-3 (2006) 489–501. doi:10.1016/j.neucom.2005.12.126.
[28] G. Li, P. Niu, X. Duan, X. Zhang, Fast learning network: a novel artificial neural network with a fast learning speed, Neural Comput. Appl. 24.7-8 (2013) 1683–1695. doi:10.1007/s00521-013-1398-7.
[29] M. H. Ali, B. A. D. Al Mohammed, A. Ismail, M. F. Zolkipli, A New Intrusion Detection System Based on Fast Learning Network and Particle Swarm Optimization, IEEE Access 6 (2018) 20255–20261. doi:10.1109/access.2018.2820092.
[30] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, K. Han, Enhanced Network Anomaly Detection Based on Deep Neural Networks, IEEE Access 6 (2018) 48231–48246. doi:10.1109/access.2018.2863036.
[31] N. Shone, T. N. Ngoc, V. D. Phai, Q. Shi, A Deep Learning Approach to Network Intrusion Detection, IEEE Trans. Emerg. Top. Comput. Intell. 2.1 (2018) 41–50. doi:10.1109/tetci.2017.2772792.
[32] K. Alrawashdeh, C. Purdy, Toward an Online Anomaly Intrusion Detection System Based on Deep Learning, y: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2016. doi:10.1109/icmla.2016.0040.
[33] B. Yan, G. Han, Effective Feature Extraction via Stacked Sparse Autoencoder to Improve Intrusion Detection System, IEEE Access 6 (2018) 41238–41248. doi:10.1109/access.2018.2858277.
[34] M. Al-Qatf, Y. Lasheng, M. Al-Habib, K. Al-Sabahi, Deep Learning Approach Combining Sparse Autoencoder With SVM for Network Intrusion Detection, IEEE Access 6 (2018) 52843–52856. doi:10.1109/access.2018.2869577.

[35] Lam, Jordan, Robert, Abbas, Machine learning based anomaly detection for 5g networks. arXiv preprint arXiv:2003.03474 (2020).
[36] GitHub - calee0219/awesome-5g: Awesome lists about 5G projects. URL: https://github.com/calee0219/awesome-5g.
[37] GitHub - omec-project/gnbsim: gNB simulator. URL: https://github.com/omec-project/gnbsim.