creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Enhancing AI-powered Data Security and Anomaly Detection in 5G Mobile Edge Computing

Jiwon Ock

Department of Future Convergence
Technology Engineering
The Graduate School of
Sungshin Women's University

Enhancing AI-powered Data Security and Anomaly Detection in 5G Mobile Edge Computing

A Master's Thesis
Submitted to the
Graduate School of Sungshin Women's University
in partial fulfillment of the requirements
for the degree of
Master of Future Convergence Technology Engineering

Jiwon Ock

Nov, 2023

This is to certify that we have examined the
Master's Thesis of
Jiwon Ock
Submitted to Department of Future Convergence
Technology Engineering

Approved as to style and content:

Thesis Advisor ..... Seongmin Kim

Committee Chairman ..... Il-gu Lee

Committee Member ..... Seongmin Kim

Committee Member ..... Yeon-sup Lim

The Graduate School of Sungshin Women's University

ABSTRACT

As 5G mobile edge computing gains more attention, ensuring data security and detecting anomalies are becoming increasingly challenging. However, incorporating AI-based solutions can be a promising solution to these challenges. By utilizing machine learning algorithms and predictive models, organizations can significantly enhance their data security and anomaly detection capabilities. By integrating AI-based solutions into 5G mobile edge computing, organizations can monitor and analyze network traffic patterns in real-time. This real-time analysis helps organizations take proactive actions by quickly detecting potential threats and anomalies. Furthermore, AI-based solutions can help detect and prevent data breaches by analyzing user behavior patterns and blocking unauthorized access. By leveraging AI-based solutions, data security and anomaly detection capabilities can be greatly improved. Therefore, organizations should consider utilizing these solutions to protect their systems and data while providing uninterrupted service to customers. This paper explores the NWDAF network function that can be utilized for AI in a 5G mobile edge computing environment. We present appropriate data protection architectures using Intel SGX and simulate them to perform anomaly detection.

Contents

Table Contents

Figure Contents

Ⅰ. Introduction

Mobile Edge Computing (MEC) is a technology designed to process operations in close proximity to terminals that collect data from 5G networks. Within MECs, there's a proposed Edge AI technology that conducts artificial intelligence learning and reasoning directly from data collection points. This innovation offers advantages such as reduced communication and processing delays, along with efficient handover [1]. As a result, it is gaining attention as a key technology for 5G applications like IoT, smart grids, and autonomous driving, positioning itself as a cornerstone for emerging services. These days, ongoing research is directed at establishing an intelligent edge network platform to provide a next-generation distributed intelligent infrastructure environment [2]. To assess, enhance, and progress the performance and security of Edge AI technology, a solid understanding of the 5G network infrastructure is imperative. 5G networks exhibit flexibility in offering various service models, achieved through the softwareization of network components based on technologies such as Software-Defined Networking (SDN) and Network Function Virtualization (NFV) [3]. Network slicing, capable of dividing physical networks into multiple virtual networks to ensure isolation for different applications, further enhances this flexibility. In the context of 5G core networks and wireless networks, there's a shift towards Standalone (SA) methods, differing from the

existing mixed method (NSA). This transition introduces new network functions as outlined by the 3rd Generation Partnership Project (3GPP) Rel-16 [4]. Notably, the Network Data Analysis Function (NWDAF) automates 5G networks, the Network Repository Function (NRF) facilitates interworking between network functions, and the Network Slicing Selection Function (NSSF) provides information for selecting the optimal network slice. These network functions play a pivotal role in 5G network-based Edge AI, offering intelligent security solutions such as real-time network automatic detection and authentication [5].

Addressing security concerns in edge AI models within 5G networks is crucial, especially when dealing with adversarial attacks. Notably, data poisoning attacks are prevalent in distributed infrastructure environments like MEC. In scenarios where MEC collaborates with centralized clouds in federated learning, the model's performance is directly impacted. Despite extensive studies [6, 7] on data poisoning attacks in LTE networks, research in 5G Standalone (SA) networks remains insufficient. Previous studies often focused on poisoning attacks on Radio Access Networks (RAN) [6], neglecting network functions in core networks like NWDAF and NRF. Considering the distinct network processes between LTE and 5G SA modes, poisoning attack detection technology developed for 4G may not be effective in 5G environments. Specifically, data poisoning attacks target AI-based authentication and authorization services controlling access within a 5G network. The frequent authentication

in 5G networks, due to small cell densification, makes efficient authentication mechanisms vulnerable to poisoning during network communication [8]. In 5G SA mode, NF can become cloud-native, programmable, and edge-friendly. The NWDAF, a key player in 5G core network capabilities, facilitates Edge AI for optimizing and enhancing 5G. However, the efficacy of NWDAF heavily relies on the quality of a diverse 5G training dataset. Obtaining data on abnormal conditions is a challenge, leading to imbalances in datasets compared to normal conditions. Additionally, 5G network data is typically privately owned by mobile network operators.

To tackle the shortage and imbalance of 5G data, various data augmentation methods have been proposed. These methods play a crucial role in improving machine learning models' performance on 5G networks. While commercial cloud services offer data scaling technologies, relying on federated learning-based cloud services introduces security risks, as they can't guarantee the integrity and confidentiality of the data scaling process. A federated learning algorithm has been suggested to overcome security and server overload issues associated with data movement in centralized AI learning methods. However, it comes with its challenges, including the potential manipulation of data scaling processes by malicious attackers within cloud services. This manipulation could result in biased or maliciously manipulated magnified data. Traditional methods of removing or filtering data are not applicable in federated learning environments where central servers can't access client

data. Non-central servers are also vulnerable to data leakage, uncommon in central server settings. Therefore, a protective mechanism is essential in a federated learning environment to ensure secure augmentation and analytics of data, even when client-held data includes adversarial input.

NWDAF is integral in analyzing network data collected from various 5G NFs, offering several benefits to existing 5G capabilities such as anomaly detection and load balancing. Anomaly detection, in particular, is crucial for maintaining network reliability in 5G SA mode, making NWDAF a key component for prompt anomaly detection and response. For instance, the 3GPP TS 23.501 [9] standard highlights UE anomaly behavior/detection as a specific use case for NWDAF, employing formula-based or AI-ML analysis. NWDAF, through UE mobility analysis, can identify abnormal UE behavior by adjusting service area limitations for Policy Control Functions (PCFs). Analyzing UE mobility allows for the adjustment of service area restrictions to identify anomalies, enhancing policy control. Despite its significant role, NWDAF poses a risk of data leakage as it consolidates data from multiple NFs. To address this risk, federated learning with NWDAF is proposed to enhance data privacy, particularly when dealing with sensitive information in 5G networks. However, the integration of NWDAF and federated learning presents challenges, including the potential leakage of data during the aggregation process.

Given these considerations, research is essential to safeguard the

secure utilization of data based on artificial intelligence in 5G MEC environments. This study aims to enhance anomaly detection capabilities while addressing privacy concerns associated with the central role of NWDAF in consolidating data from diverse network functions. This paper will help address network data security issues and access rights in 5G MEC. Furthermore, it aims to enable collaborative federated learning across multiple NWDAFs within the 5G ecosystems, employing secure remote attestation using SGX primitives. In summary, this paper makes the following contributions:

1) This paper explores a NWDAF threat model with edge AI technology combined in 5G SA mode. It confirms the possibility of a data contamination attack on NWDAF in the proposed deployment scenario. In addition, we propose a feature selection method to increase the performance of data poisoning attack detection on NWDAF, comparing performance through three classifiers, and demonstrating the necessity of feature selection.

2) This paper shares our preliminary findings on the application of CTGAN for generating synthetic data pertaining to 5G NWDAF. And we evaluate the quality of synthetic data and the usefulness of detecting anomalies with high accuracy to solve the problem of lack of 5G data sets.

3) This paper proposes an architectural approach that harnesses the capabilities of Intel SGX to bolster data privacy and security in the context of federated learning within the 5G NWDAF. Our paper

provides an in-depth examination of the NWDAF federated learning architecture, elucidating its components and highlighting the crucial role played by SGX. Additionally, we showcase the practicality and effectiveness of our approach through the implementation of anomaly detection.

These are the contents of our study. Section 2 provides the essential background knowledge necessary for a comprehensive understanding of our research. In Section 3, we present our research focused on enhancing the performance of data contamination attack detection in the NWDAF for big data analysis capabilities within the context of 5G mobile edge computing. Moving on to section 4, we delve into studies on synthetic data generation designed to improve anomaly detection within the architecture of the 5G NWDAF. In section 5, our attention shifts to a study that introduces Intel SGX as a means to augment privacy through federated learning of NWDAF on 5G MEC. Section 6 is dedicated to the introduction and comparison of studies closely related to our research. In section 7, we will discuss the content covered in this paper and the research direction, and conclude in section 8.

Ⅱ. Background

In section 2.1, our attention is directed towards network functions designed for 5G networks. Section 2.2 delves into the Network Data Analysis Function (NWDAF), a focal point in our research. The subsequent section, section 2.3, provides an overview of poisoning attacks from the perspective of 5G networks. Finally, section 2.4 delves into the functionality of Intel SGX, a crucial aspect explored in our study.

1. 5G Network Function

According to the Reference Architecture within the 5G standard defined by 3GPP, NWDAF and NRF have been added as new network features as key components of the 5G service-based architecture. NWDAF is a new standard feature within the core network system for the automation and intelligence of 5G networks. Artificial intelligence algorithms are used to derive analyses and provide these analyses to other network functions. Specifically, 3GPP Rel-17 defines an interface related to Model Training Logical Function (MTLF) within the NWDAF to provide the AI model learning and learned model to NF users [10]. Currently, the standard only stipulates data collection interfaces and procedures for AI/ML and

does not stipulate detailed algorithms or specific application conditions.

NRF is a network functional service framework component that supports inter-working between them through service status monitoring and interlocking information management of changing 5G core network functions. It maintains a profile of available network feature instances and supported services in a 5G core network, and allows other instances to subscribe to and receive notifications for a specific type of new instance's NRF registration. It also supports a service discovery feature that receives NF discovery requests from each instance and provides information on available instances that meet specific criteria [11].

2. NWDAF on 5G Network

According to the system-level structure defined in the 5G standards by 3GPP, NWDAF and network repository function (NRF) have been added as key components of the 5G service-based architecture along with the traditional core functions, such as access management function (AMF), user plane function (UPF) and session management function (SMF) [9]. Figure 1 illustrates the 5G service-based architecture, highlighting the role of NWDAF and NRF.

The 5G network experiences the generation of network traffic, driven by data from its users.

The NRF plays a pivotal role in facilitating the interworking of various 5G NFs and stores data that passes through multiple NFs such as AMF, SMF, and unified data management (UDM) [9]. Subsequently, the NWDAF receives the data stored in the NRF and initiates a pre-processing stage. Finally, the NWDAF leverages local artificial intelligence models for learning and data analysis.

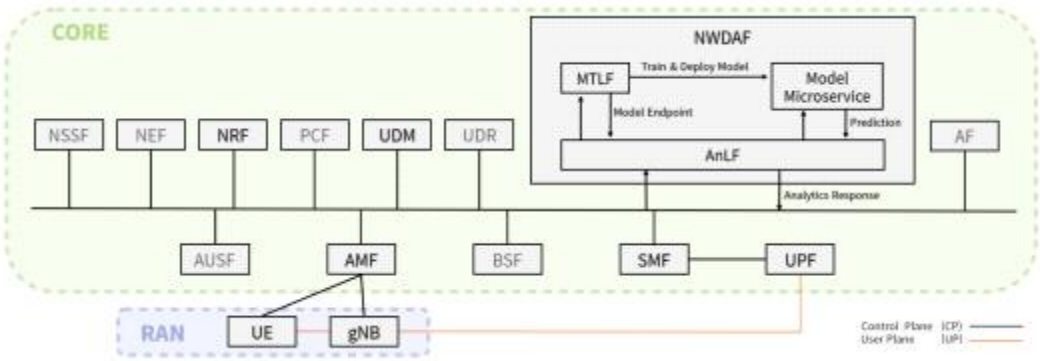


FIGURE 1. NWDAF in 5G core

NWDAF aims to help optimize and improve 5G core network functions [10]. It collects network data from network functions of 5G cores (e.g. NRF, SMF, UDM) and utilizes AI and ML to explore and analyze problems. NWDAF operates in a way that processes and analyzes network data collected from the data collection interface through AI and ML models to other network functions through the data exposure interface. Each network function may directly request specific analysis information from the NWDAF for the analysis result. The NWDAF consists of an analysis logic function (AnLF) and

a model training logic function (MTLF) [10]. AnLF performs data analytics, such as network statistics and predictive information, and serves network analytics information to request 5G NFs. MTLF is dedicated to machine model training and provides a training model provisioning service through a subscription-based procedure.

3. Data Poisoning Attacks in Mobile Networks

A data poisoning attack involves malicious actors tampering with machine learning algorithms. This tampering can include providing incorrect training data or adding noise to the data, which disrupts the learning process and leads to unfavorable outcomes. To execute such an attack, the attacker typically selects a subset of training data to manipulate. They begin by running this subset through a surrogate model, essentially a model that imitates the target machine learning model. Then, they alter the labels of the data in this subset. If the results from deep learning on this modified data significantly differ from the expected outcomes in the surrogate model, the attacker sends this incorrectly labeled data to the target machine learning model. This process constitutes a data poisoning attack. In some prior research [12], there's a specific focus on spectral data poisoning attacks in cognitive wireless communication networks. These attacks allow the adversary to transmit poisoned spectral data during brief idle periods in the communication channel. The poisoned data can then be exploited when AI and

machine learning models are retrained. The outcome is that the modified model becomes vulnerable to data poisoning attacks. The attacker manipulates both data and control plane communication using deep neural networks, ultimately putting the system at risk of data poisoning attacks. This interference disrupts efficient spectrum sharing between existing users and 5G networks [13].

4. Intel SGX

Software Guard Extensions (SGX), developed by Intel, is a secure processor architecture to ensure trust on platforms with sensitive and valuable information [14]. SGX has an isolated protection area called an *Enclave*, which prevents access to code and data except for programs running inside the enclave.

Before creating an enclave, a developer should identify the assets that need to be protected, the data structures that contain the assets, and the code sets that work with them [15]. This step is essential because if you initialize the enclave and load the code into memory, untrusted components cannot read or change the code. Once identified, place them in a separate trusted library. After defining the enclave, a reliable component of the Intel SGX program, it crosses the enclave boundary using *Ecall* and *Ocall* as shown in Figure 2. Ecall serves as an interface function to the enclave from outside, while Ocall serves to the outside application from within the enclave.

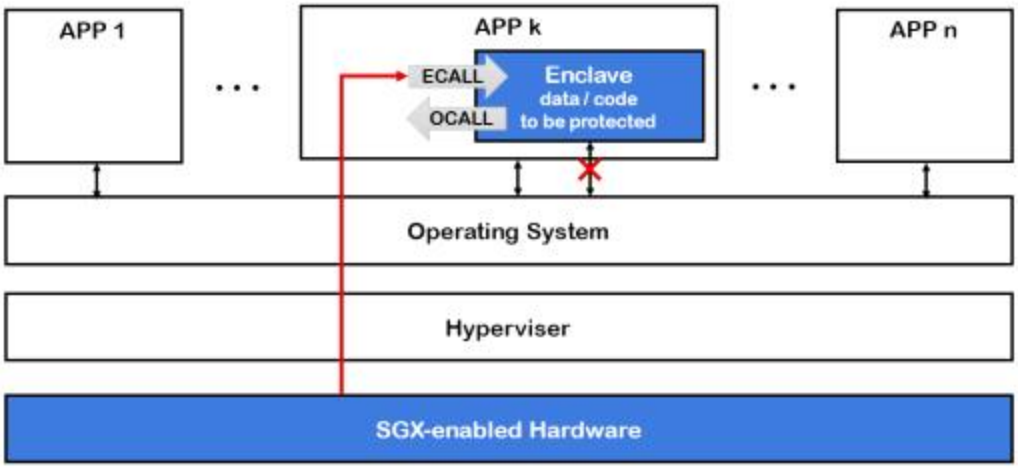


FIGURE 2. The design of Intel SGX (The blue represents the trusted zone.)

Once the code and data to be safeguarded are partitioned within an enclave, any untrusted components, including the operating system and hypervisor, are inherently barred from accessing or manipulating them. This protection is enforced once an enclave region is measured and initialized.

The enclave region is mapped to a specialized memory area encrypted by hardware-specific keys embedded in the CPU. Note that the memory encryption key stored within the CPU is both inaccessible and randomly updated owner epoch at each power cycle. Consequently, enclaves can only be accessed from trusted entry points specified by the enclave developer who signs the enclave, irrespective of the CPU mode or current privilege level. This design effectively resists external software tampering and is

characterized by its robust security features, including memory encryption, remote attestation, and sealed storage. Remote Attestation verifies the integrity of the enclave by checking whether the application is actually running within the enclave of the isolated platform [15]. In addition, SGX provides sealing/unsealing primitives that facilitate the encryption of enclave memory when it is persisted to disk. These mechanisms ensure that enclave data remains protected even when it is stored externally.

Ⅲ Analyzing Data in a 5G MEC Environments: Detecting Data Poisoning Attacks and Abnormal Use Cases

As mobile edge computing (MEC) is gaining attention as a core technology of 5G networks, edge AI technology of 5G network environment based on mobile user data is recently being used in various fields. However, as in traditional AI security, there is a possibility of adversarial interference of standard 5G network functions within the core network responsible for edge AI core functions. In addition, research on data poisoning attacks that can occur in the MEC environment of standalone mode defined in 5G standards by 3GPP is currently insufficient compared to existing LTE networks. In this study, we explore the threat model for the MEC environment using NWDAF, a network function that is responsible for the core function of edge AI in 5G, and propose a feature selection method to improve the performance of detecting data poisoning attacks for Leaf NWDAF as some proof of concept. Through the proposed methodology, we achieved a maximum detection rate of 94.9% for Slowloris attack-based data poisoning attacks in NWDAF.

In section 3.1, we explain why Slowloris attacks can be viewed as poison attacks on 5G networks. Section 3.2 presents an NWDAF threat model including feature selection based on poisoned data in

the 5G MEC environment Slowloris attack scenario, and in section 3.3 we analyze 5G Slowloris attack datasets and conduct classification experiments based on feature selection. In addition, section 3.4 suggests that a federated learning architecture using SGX may be valid for secure data augmentation for image datasets as well as 5G datasets.

1. Slowloris attack

A Slow HTTP Header DoS (Denial of Service) attack is a malicious attempt to disrupt a web server's normal operation by manipulating the information in an HTTP header. This manipulation is done in a way that causes the webserver to wait for complete header information [16]. During this waiting period, there's a limit to the available resources that the web server can allocate to maintain connections. If this limit is exceeded, the attacker essentially blocks or denies access to the server for legitimate users. In the HTTP protocol, the end of the header is marked by a specific sequence, which is represented as '/r/n'. The attacker prolongs this header by continuously adding meaningless variables without sending this specific sequence to mark the end of the header. This process of extending the header without a proper conclusion exhausts the server's available resources to maintain connections, effectively crippling the server and preventing normal access. This is how the Slow HTTP Header DoS attack works.

2. NWDAF Threat Model and Deployment Scenario in a 5G MEC Environment

In this section, we've designed an NWDAF architecture and considered scenarios within a 5G network environment where data poisoning attacks can take place, based on previous studies. We also describe how the NWDAF architecture is deployed using artificial intelligence in a 5G MEC environment. As an illustration of a data poisoning attack, we've chosen the Slowloris attacks as examples. These attacks serve as instances of poisoning attacks that introduce noise into the data. In this case, the web server mistakenly interprets this as ongoing traffic due to alterations in the header information.

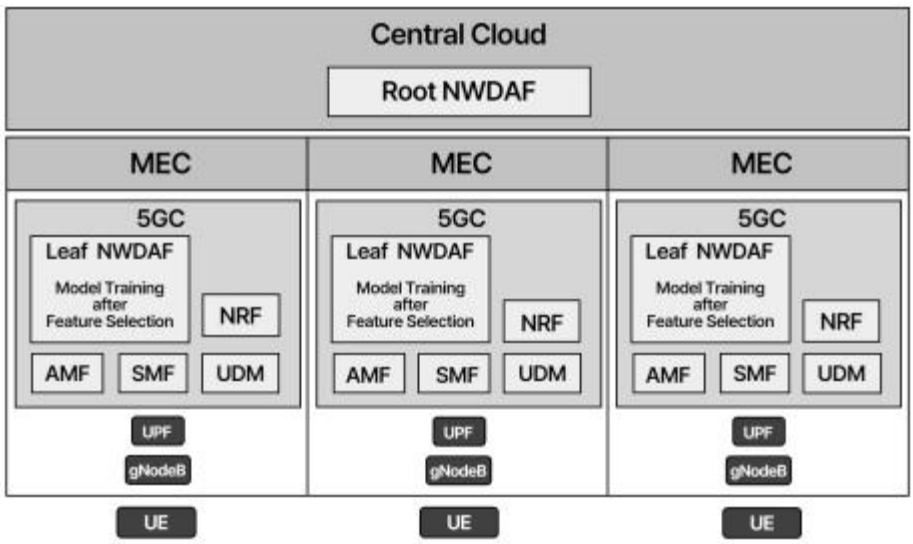


FIGURE 3. A NWDAF Architecture for distributed and federated learning in 5G MEC Environment

The network function deployment scenarios in MEC environments are based on a previously proposed architecture in studies [17, 18]. Figure 3 shows the distributed NWDAF architecture within a 5G MEC environment. This architecture encompasses various network functions, including gNodeB, UE (User Equipment), NRF (Network Repository Function) in the core network, as well as fundamental 5G network components like AMF (Access Management Function), and SMF (Session Management Function). Additionally, there's the NWDAF for machine learning. In this setup, there are two key components: the Root NWDAF, located in the centralized cloud, and the Leaf NWDAF, which exists in each MEC. This dual-layered NWDAF structure is designed for decentralized learning. This architectural choice enhances security by mitigating the risk of a single point of failure, and it enables efficient management of network resource utilization [17].

In this architecture, we consider a threat model where attack data is introduced from UEs and data that has been poisoned by the NRF (Network Repository Function) is present. In this context, 'poisoned data' refers to data that is intentionally altered to manipulate labels or feature values to decrease the accuracy of artificial intelligence learning models. The attackers' goal is to make their malicious attack traffic indistinguishable from normal network traffic. If such malicious traffic is incorrectly recognized as normal and is used for learning, it's considered to be poisoned data. For instance, if malicious traffic, like that generated by a Slowloris attack, which

alters HTTP header information, is mistakenly seen as normal and used for learning, it falls into the category of poisoned data. When this type of poisoned data is employed within the NWDAF, it can reduce the accuracy of the NWDAF's learning model and lead to errors. To address this, in the proposed scenario, labels are assigned based on detection results, classifying traffic as either normal or malicious. If poisoned data is identified, it is correctly labeled as malicious traffic to mitigate its impact on the learning process.

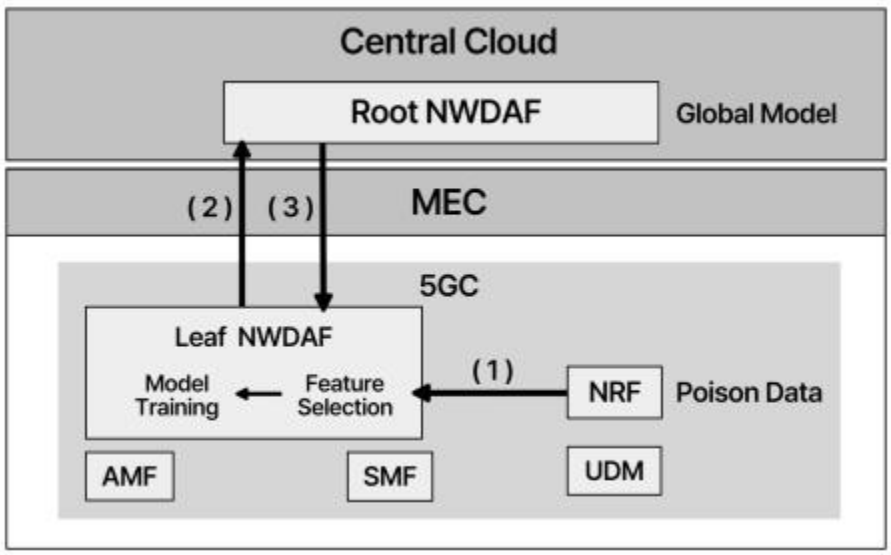


FIGURE 4. Operation process of NWDAF architecture including feature selection in 5G MEC environment

The specific operational process of the NWDAF architecture, considering both normal and poisoned data, is illustrated in Figure

4. In this setup, gNodeB receives data from terminals, where one terminal is connected per MEC. This data then passes through a UPF (User Plane Function) to reach AMF and SMF. This section focuses on a scenario where feature selection is conducted within the Leaf NWDAF to enhance model performance within each MEC. The overall operational process can be broadly divided into three phases. For the sake of clarity, we'll explain the execution flow within a single MEC environment.

Here's a more readable explanation of the operational process within the MEC-based 5G NWDAF architecture:

(1) Initially, network traffic containing poisoned data is generated by malicious users on 5G networks. Both normal and poisoned data, introduced into the terminal through various NFs like AMF, SMF, and UDM, are stored in the NRF, which supports intercommunication between different 5G NFs. The Leaf NWDAF then acquires this mix of normal and poisoned data from the NRF and subjects it to preprocessing and feature selection. Following this, local artificial intelligence models within the Leaf NWDAF undergo learning. However, it's important to note that poisoned data can lead to misclassifications due to malicious interference, resulting in the generation of learning information (weights) for the dataset in each Leaf NWDAF that exhibits poor learning performance.

(2) The Leaf NWDAF in each MEC conveys the weight information derived from the learning results to the Root NWDAF. These weights are aggregated to create a global model in a centralized cloud

where the Root NWDAF is deployed. A noteworthy aspect here is that akin to federated learning, this process enables the construction of a global model without sharing the actual data across each MEC.

(3) The Root NWDAF, functioning as an aggregation server, generates a global model that a data poisoning attack has influenced and shares it with the Leaf NWDAFs. The Leaf NWDAFs use this globally affected model, which includes adversarial disturbances, to update their local models. Consequently, each MEC becomes impacted by the data poisoning attack.

Our objective is to assess the impact of data poisoning attacks and the effectiveness of feature selection in detecting and addressing them within this MEC-based 5G NWDAF architecture. We will evaluate these aspects through experiments and analysis.

3. Enhance data poisoning attack detection with feature selection

In this section, we delve into feature selection techniques aimed at enhancing the detection of machine learning-based data poisoning attacks in the 5G MEC deployment scenarios we've described. It's important to note that obtaining real 5G network data, which belongs to telecommunications operators, is quite challenging due to privacy concerns, and there's a scarcity of openly available datasets. For our purposes, we assume that Slowloris attack data in

5G-NIDD, a source that provides intrusion detection data in a dual 5G network, serves as the poisoned data responsible for misclassifying malicious traffic as normal, as previously discussed in the deployment scenarios. The Slowloris attack is a prime example of poisoning, impacting not just the labels of the data but also the actual feature values themselves.

We will provide a more in-depth exploration of this in section 3.3.1, with a particular focus on the primary features involved. Our aim is to evaluate the performance of data poisoning attack detection within the Leaf NWDAF, which corresponds to the regional model in the entire federated learning process. This evaluation serves as a proof of concept for the entire architecture. In this proposal, we employ feature selection as a means to address underfitting. We achieve this by eliminating features that tend to degrade prediction performance. This, in turn, enhances the detection performance of data poisoning attacks. In our evaluation, we aim to assess the effectiveness of the proposed feature selection methodology by comparing the classification performance when utilizing decision tree, linear regression, and SVM (Support Vector Machine) classifiers.

3.1 5G-NIDD Datasets

The 5G-NIDD [17] dataset provides comprehensive network intrusion detection data generated over a 5G wireless network. Specifically, the dataset is a dataset that contains and labels data

built and extracted from a 5G testbed for users developing and testing AI/ML-related solutions. The 5G testbed is linked to the University of Oulu's network in Finland. The data within the dataset is extracted from two base stations equipped with both attacker nodes and 5G users [19]. These attacker nodes are responsible for targeting servers within 5GTN MEC environments. The attack scenarios encompass activities such as denial of service attacks and port scanning. In this section 3, we use Slowloris attack data, one of the denial of service attacks, as a learning dataset.

TABLE 1. 5G-NIDD dataset overview (Slowloris Attack)

Number of features	113 (SrcID, Flgs, RunTime, Sum, Rate, Label, etc)
Total instances	31,015
values of 'Label' feature	Benign, Malicious
# of 'Benign' instances	24,808
# of 'Malicious' instances	6,207

Table 1 shows the details regarding the Slowloris attack dataset. This dataset contains information extracted from network traffic that led to resource depletion. This depletion occurred due to a Slowloris attack on a specific web server, and it took place over approximately 10 minutes during a half-hour session. In the Label column, the values of Benign and Malicious can be found, and the term "Malicious" designates data that is regarded as poisoned or

contaminated, as it has the potential to cause misclassification in learning outcomes.

3.2 Perform feature selection

Feature selection is the selection of important features without modifying the features constituting the model [20]. Features with weak correlations with labels have the potential to lower the performance of the model in the learning process. Moreover, if the model is trained with an excessive number of features, it may encounter an issue known as overfitting. Consequently, to enhance the detection of data poisoning attacks, feature selection was conducted to create an improved learning model. This was achieved by isolating key features from the multitude available in the Slowloris attack dataset. First of all, pre-processing was performed to reduce from 113 to 30 by removing features with no pattern of distribution, too many missing values (Null), overlapping, or impossible-to-identify patterns, as if the value of the features were all zero.

Feature selection methods are broadly categorized into Filter, Wrapper, and Embedded techniques, and for our purposes, the Embedded method was chosen. The Embedded method combines the strengths of both Filter and Wrapper methods [20]. The Filter method measures the relevance of individual features, while the Wrapper method gauges the utility of feature subsets. The Embedded method integrates the benefits of both, enabling direct learning of

each feature using built-in metrics and optimizing the learning process while evaluating the usefulness of feature subsets. With this approach, we aimed to select features that significantly contributed to the model's accuracy. Within this selection process, key features were identified through a decision tree-based algorithm. Decision tree-based algorithms simplify classification and prediction tasks by constructing classification trees based on the frequency of data points within each category [21].

To analyze feature importance, we employed a decision tree model algorithm [21] with the DecisionTree Classifier, using the preprocessed dataset containing 30 features. In the context of a Slowloris attack, one of these features, called "Rate," plays a pivotal role. This is because, during a Slowloris attack, the modification of the final character in the packet header impacts the Rate feature within the data used for learning. The Rate characteristic represents the response rate corresponding to echo requests. In the case of Slowloris attacks, this rate decreases significantly compared to normal traffic. Slowloris attacks work by continuously sending HTTP headers to specific web servers, preventing the establishment of complete connections. The web server, in turn, waits without responding to echo requests until the full header information is received. Consequently, the Rate feature extracted from the attacked and modified data becomes zero, a significant departure from normal data. Slowloris attacks, while considered denial-of-service attacks in terms of availability, can be seen as data poisoning

attacks from the perspective of NWDAF. They distort the learning results by manipulating the feature values within the learning data. Hence, the Rate feature is critical in determining the label for Slowloris attacks, as it helps distinguish malicious traffic from benign traffic. After splitting the dataset into a 70% training set and a 30% testing set, a decision tree model was created and validated. The accuracy of the test dataset was 93.7%. Subsequently, we used the Python library *feature_importance_* to assess the importance of each feature based on the Rate feature. This analysis revealed that the Sum feature was about 43%, SrcRate was about 8%, and TotPkts of 7%, indicating that these four features are important characteristics.

TABLE 2. Information on the main features of the Slowloris dataset

Main features	Information
Label	(Integer) Slowloris attack or not
Rate	(float) Response rate per echo request
SrcRate	(float) Source to Destination packets per second
Sum	(float) Total duration of aggregated record
TotPkts	(Integer) Total transaction packet count

Table 2 describes the features of high importance among the 30 features pretreated in a total of 113 Slowloris attack data set features. The data was collected using the Argus tool [22], which is

designed for monitoring network flows. The table includes information regarding *Label* feature, which indicates whether an attack has occurred or not. Additionally, it highlights the high-importance features, namely *Rate, Sum, TotPkts*, and *SrcRate* features, respectively.

FIGURE 5. Accuracy of test dataset and training dataset environment

We conducted an experiment by initializing memory for verification and applying the same algorithm to all features, except for the four high-importance features: Rate, Sum, TotPkts, and SrcRate. The results showed a decrease in the accuracy of the test dataset by approximately 2%, dropping from 93.7% to 91.7%. This decline demonstrates the importance of the four excluded features mentioned above. The results of evaluating the accuracy of the test

dataset, while excluding these crucial features, are presented in Figure 12-(a). In conclusion, our approach involves selecting and learning the four key features, which include Sum, TotPkts, SrcRate, and Rate (a feature critical for distinguishing Labels). Subsequently, we will assess the performance of Slowloris attack detection, which is one of the data poisoning attacks in the realm of 5G.

3.3 Analysis of classification results

To assess the effectiveness of feature selection, we employed three supervised learning-based machine learning classifiers: Decision Tree, Linear Regression, and SVM. These classifiers were used to determine whether a given dataset could be classified based on a 'Label' feature. In the case of the Slowloris attack dataset, we focused on 29 features, excluding 'Label'. Additionally, we included four data features—Rate, Sum, TotPkts, and SrcRate—identified as highly important through the feature selection process for the Slowloris attack dataset. Throughout this analysis, the learning rate of the dataset was set at 0.7 and the random_state was configured to 42.

In the case of the machine learning toolchain to build an experimental environment, Google Colab was used. Experiments were conducted using GeForce RTX 3090 GPU and installing nvidia-driver-525 and CUDA version 12.0 on hosts with Ubuntu 22.04 version installed.

TABLE 3. Comparison of classification performance according to feature selection

	Type (# of feature)	Decision Tree	Logistic Regression	SVM
Accuracy	ALL features (29)	0.917	0.789	0.801
	proposed approach (4)	0.938	0.833	0.949
Precision	ALL features (29)	0.920	0.725	0.636
	proposed approach (4)	0.942	0.736	0.966
Recall	ALL features (29)	0.916	0.080	0.798
	proposed approach (4)	0.938	0.251	0.773
F1	ALL features (29)	0.916	0.150	0.708
	proposed approach (4)	0.937	0.375	0.859

Table 3 is the classification metrics, including Accuracy, Precision, Recall, and F1 score, for the test dataset, which accounts for 30% of the entire dataset. These results are rounded to four decimal places. Notably, the classification performance demonstrates improvement when feature selection is employed to enhance the detection of data poisoning attacks, compared to conducting

learning with all pre-processed features in all three classifiers. It is noteworthy that not only is the decision tree classifier performance used in the feature importance analysis but also the performance is improved in Linear Regression and SVM models. This indicates the efficacy of feature selection in enhancing classification performance.

In the case of the Linear Regression classifier, it's worth noting that Recall and F1 score exhibit lower results compared to other metrics. This difference arises from the distinct approach taken in generating boundaries to distinguish different classes. In the context of Linear Regression, the values of Recall and F1 score can vary based on the threshold set for the linear boundary determination. Thus, the classification performance depends on how this threshold is specified. On the other hand, Decision Tree and SVM exhibit superior classification performance because they can accommodate nonlinear boundaries and automatically determine optimized boundaries for classifying different classes.

4. Additional: SGX-enabled edge cloud architecture for secure data augmentation

Data augmentation plays a vital role in addressing data shortages and improving the performance of machine learning models. As the demand for data augmentation increases, recent commodity cloud services offer a variety of scaling technologies such as rotation, cropping, flipping, and color adjustment, providing a rich set of

data for a wide variety of applications. A case in point is Stratio's augmented data fabric platform [23] and Oracle's fusion analytics warehouse [24]. To address the security and server overload issues caused by data movement in centralized AI learning methods, a federated learning algorithm has been proposed. This algorithm enables the generation of high-performance models without the need for data sharing.

However, relying on federated learning-based cloud services poses security risks as it cannot guarantee the integrity and confidentiality of data augmentation processes [25]. Malicious actors within cloud services can manipulate or corrupt these processes, leading to biased or maliciously manipulated augmentation data. Such actions can have significant implications for machine learning tasks. In the context of federated learning, where a central server doesn't have access to client data, traditional methods of removing or filtering data are not feasible. Furthermore, non-central-server clients are susceptible to data poisoning attacks, which are not typically encountered in a central-server setup. Therefore, in a federated learning environment, a protection mechanism is required to ensure the secure augmentation of data, even when the data held by clients may contain adversarial inputs.

We propose SGX-enabled edge cloud architecture that ensures reliable and secure data augmentation, utilizing the security features offered by Intel SGX. SGX provides a secure environment that safeguards sensitive and valuable information, protecting augmented

code and resulting images from unauthorized access or tampering. By analyzing the security risks of adversarial attacks against the augmentation processes in the untrusted cloud, we explore the design space of a secure augmentation by leveraging the state-of-the-art confidential computing mechanism, Intel SGX.

4.1 Security Risks in cloud-based data augmentation

Cloud service-based data augmentation poses security risks, including unauthorized access to augmented data stored in the cloud, potential data breaches, and privacy violations [25]. System failures or cyberattacks targeting the cloud service can typically result in data loss, service disruptions, and downtime. Likewise, a malicious insider may intercept data transmission between the client and the cloud service or tamper with the data. Concerns arise regarding the honest-but-curious cloud service provider's access to augmented data and the potential misuse of sensitive information. Dependency on a third-party cloud service exposes organizations to vulnerabilities in the provider's security measures [26]. In particular, this paper deals with adversarial attacks against data augmentation among these security risks.

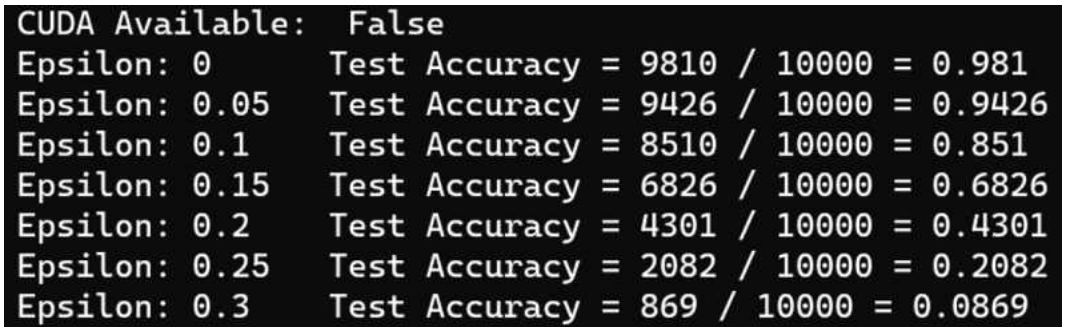


FIGURE 6. FGSM attack based on epsilon value

Adversarial attack e.g. FGSM: FGSM (Fast Gradient Sign Method) is an adversarial attack method commonly used in augmented image generation in cloud services [27]. An attacker can calculate the gradient of the loss function for the input image to determine the direction in which the image pixels are agitated to maximize classification errors. By injecting these attacked images into the training dataset, attackers can manipulate the learning process of cloud services, potentially resulting in poor model performance or biased results. We demonstrate the FGSM attack to prove its effectiveness by changing the distortion multiplier epsilon from 0 to 0.3 in Ubuntu 20.04 on Intel i9-10900K CPU. Figure 6 shows that the accuracy decreases as the degree of distortion increases according to the epsilon value.

4.2 SGX-enabled Edge Cloud architecture for reliable augmentation

To develop an SGX-enabled reliable data augmentation procedure, we assume a hierarchical edge cloud architecture that employs

federated learning, a typical AI model, to tackle the challenges of data centralization to a central cloud [28]. Poison data, which adversely affects the AI model's learning results, flows from the client's input and exists in the local model. Normal and poisoning data are received from the client's input and transmitted, along with the local model, to an SGX-enabled cloud for augmentation. The weights learned from the local model are obtained from the central source and averaged. The shared model is then sent back to the local model on each edge, updating it. The updated local model on each edge provides results to the client, ensuring improved privacy and distributed learning capabilities.

As our threat model, we consider the one-shot kill poison attack as an adversarial attack, a clean-label poison attack. The attack performance was confirmed by dividing it into general and intelligent attackers according to the attacker's attack strategy. They propose a data augmentation defense mechanism at the time of Inference and demonstrate the effectiveness of adversarial attack defense by reducing the attack success rate by up to 65% compared to conventional methods. To mitigate this, we build a defense mechanism inspired by the previous study [29]. It proposed a technique to degrade attack success rates by applying variations to learning and inference data without a separate detection and removal process for poison data.

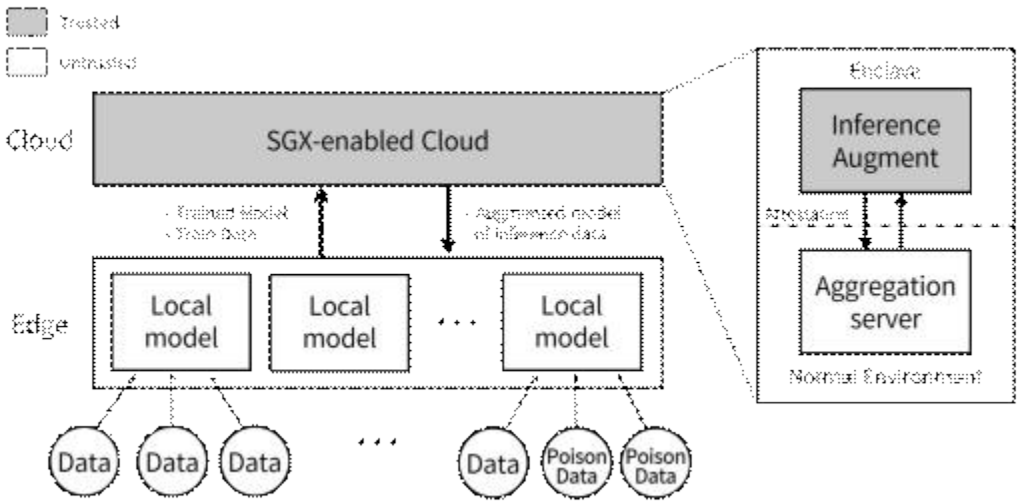


FIGURE 7. SGX architecture with secure augmentation for adversarial attack defense

The proposed architecture has two assumptions. First, the source code for Augmentation itself is trust. Second, the edge cloud itself is not malicious; from a local point of view, it is unreliable only for input data. Figure 7 schematically shows an architecture that augments using SGX. The following describes the scenario specification regarding the central cloud, SGX Enclave, and edge cloud.

Central Cloud: The central cloud provides code for augmentation services. Here, the types of augmentation include rotation, crop, flip, RGB, etc.; if clients want to make augmentation more complex, they can deploy additional types. Note that the host in the central cloud offers hardware support relevant to SGX functionality.

SGX Enclave: The training model and data received from the edge cloud may be untrusted. However, the inference augmentation code executed within the SGX Enclave, along with the resulting augmentation data, can be trusted. This allows for the delivery of a trusted model and data to the edge cloud. The augmentation providers utilize unique provisioning passwords to establish secure admissions that identify SGX hardware [14]. The software validation process commences upon Enclave initialization, and the augmented code is loaded into the Enclave using specific commands. To ensure confidentiality, the augmentation is performed as a black box within the Enclave, preventing attackers from knowing the specific type of augmentation applied. If the encryption hash of the built Enclave matches the expected value, it confirms to the data provider that a reliable augmentation code has been executed on SGX. In other words, from the client's perspective, the code intended for distribution in the central cloud for aggregation can verify whether it is applied within the Enclave.

Edge Cloud: Edge clouds can ensure reliable operations through aggregation and augmentation processes that operate on Enclave within the central cloud. This helps to trust Enclave and establish confidence in the augmentation process, even if other entities attempt to insert malicious elements. Compared to the Central Cloud, the local models have limited SGX-related resources. Federated learning is mainly used to improve the performance of the local model. At this time, data protection is required when

modelers need to provide data when handling local models, but there is a limitation that training within SGX has a large cost. According to existing papers [28], a solution was proposed that only certain layers use SGX. Only the first few layers of the artificial intelligence model can be deployed to the EPC to overcome the performance part using SGX, and the rest of the layers can be sent outside the EPC.

According to the proposed architecture, SGX Enclave requires only performing data augmentation code, so there is less overhead. A simple experiment was conducted to verify that the data augmentation code worked well in SGX. This experiment was performed in a server consisting of an Intel 3.70GHz Core™ i9-10900K CPU equipped with SGX capability and installing sgx_linux_x64_driver_2.11 version on Ubuntu 20.04. When performing data rotation, shifting, and zooming codes on a trial basis, we can check the data operation well in SGX.

Ⅳ. Exploring Synthetic Data Generation for Anomaly Detection in the 5G NWDAF Architecture

We explore the methodology to obtain proper and high-quality 5G data by leveraging conditional tabular generative adversarial network (CTGAN) [30], the state-of-the-art synthetic data generator based on deep learning. To address the data scarcity issue, a recent study utilizes a typical GAN model for generating 5G network data [31]. However, the generator produces a limited variety of samples for the tabular data whose columns contain various data types, including discrete and categorical data, as it is basically designed to utilize continuous variables. We believe that the state-of-the-art GAN model, ensuring high-quality synthetic data containing discrete and continuous variables, helps overcome such limitations. We conduct the assessment of the effectiveness when leveraging CTGAN on the NWDAF dataset with various metrics and provide a preliminary result that demonstrates its viability in anomaly detection.

Section 4.1 explains why synthetic data is needed in a 5G MEC environment, and Section 4.2 explains the 5G NWDAF dataset and generates synthetic data. In section 4.3, we evaluate synthetic data as a case study and analyze anomaly detection performance.

1. Limitations of Data Collection in 5G MEC Environments: Why Synthetic Data Is Needed

To accommodate high-speed and low-latency 5G mobile services, the deployment mode of 5G has been changed from the traditional non-standalone(NSA) mode to the standalone(SA) mode [9], which fully connects to 5G RAN and core network rather than relying on the legacy 4G LTE control plane. Such innovation enables network functions(NFs) to be cloud-native, programmable, and edge-friendly. Among them, the network data analytics function(NWDAF) standardized by 3GPP plays a key role in edge AI in 5G to optimize and enhance the 5G core network functions. It analyzes network data collected from other 5G NFs and produces some benefits on existing 5G functionalities, such as anomaly detection [9, 32] and load balancing [33].

In particular, anomaly detection is one of the crucial requirements in 5G SA mode to maintain network stability and reliability, and NWDAF would be a pivotal component in detecting abnormal situations with prompt responses. For example, 3GPP TS 23.501 [9] standard specifies the UE Abnormal behavior/anomaly detection as a use case of formula-based/AI-ML analysis based on NWDAF. By analyzing UE mobility with NWDAF, service area restrictions can be adjusted to the policy control function(PCF) to identify abnormal behavior [10]. The efficiency of NWDAF to ensure such properties highly depend on the quality of the extensive 5G training dataset.

However, the core challenge to procuring the datasets is the scarcity of data for abnormal states, leading to data imbalance compared to benign conditions and furthermore, 5G network data is typically privately owned by mobile network operators.

2. Synthetic 5G data generation

Due to the limited privilege of accessing the 5G dataset, we regard the public NWDAF dataset as a baseline for generating training data. Currently, only a select group of entities (e.g., mobile network operators) occupy the raw 5G network data. To resolve this issue, a recent study [34] releases an open-source synthetic dataset based on the fields defined in the 3GPP 5G standards. The goal was to generate a labeled dataset for 5G cellular networks, enabling them to be used for predicting network loads and classifying various types of anomaly situations that occur in 5G networks.

The dataset consists of six features: *t, cell_id, cat_id, pe_id, load*, and *has_anomaly*, described as follows.

· **t**: the amount of data transmitted during a specific time period (in bytes).

· **cell_id**: the Remote Radio Unit (RRU) cell number, with a total of 5 RRUs.

· **cat_id**: the ID based on the subscriber category name (Platinum, Gold, and Silver).

· **pe_id**: the type of connected personal device (IoT devices,

vehicles, mobiles, smartwatches, and tablets).

· **load**: the amount of data in bytes over time (calculated from LSTM and RNN using linear regression).

· **has_anomaly** (label): a feature that detects anomalies in network traffic using XGBoost and logistic regression models. Outputs a binary value, with 0 for normal and 1 for abnormal traffic.

The table contains six numerical features with various data types: 1) *t, cell_id, cat_id, pe_id*, and *has_anomaly* (categorial) are discrete and 2) *load* has continuous value.

However, a typical GAN is not suitable for such discrete variables because they rely on back-propagation for training, which only works for continuous values. CTGAN was proposed to address these issues by learning the frequency of each category to closely resemble that of real data [30]. It can handle various data types and is particularly effective at generating synthetic data with diverse feature types, such as NWDAF. The result shows that CTGAN surpasses the vanilla GAN, showing up to a 51.8% increase in F1 score, as well as improvements in L_{syn} by approximately -10 to -5 and L_{test} by approximately -60 to -3. It is worth noting that CTGAN is a specialized model for tabular data generation, where it relies on a conditional generator and discriminator. As tabular data consists of multiple feature columns with different characteristics, increasing the complexity of the data distribution, such design choice of CTGAN enables it to successfully handle complex data distributions. Consequently, we prospect that CTGAN effectively generates

high-quality synthetic data for 5G NWDAF.

3. Case Study: Preliminary Results

3.1 Environment setup

We use Google Colab with a GeForce RTX 3090 GPU and CUDA version 11.6 on a Ubuntu 22.04 machine. The nwdaf_data dataset had a total of 1,296,000 entries with 1,004,400 anomalous and 291,600 benign data. We randomly select 291,600 anomalous data and used them alongside all the benign instances for training. To avoid storage issues, we employed random sampling to generate 10,000 synthetic data that were similar to the original population.

3.2 Evaluation of synthetic data

To evaluate how similar the synthetic dataset for NWDAF is to the real dataset and how data can be generated dynamically, we estimate metrics related to the synthesizing quality and measure the anomaly detection performance. We first estimate the metrics provided by SDV (Synthetic Data Vault) [35]. The Report library provided by SDV evaluates how well the synthetic data includes mathematical attributes in the real data. Since the NWDAF dataset is a single table consisting of columns of continuous and discrete numbers, the attributes between the synthetic and real data are measured by correlation similarity using the Single Table API. Each

result is expressed as a percentage, and the closer it is to 100%, the best.

Column Shapes compare the overall distribution of column attributes. The higher the score, the more similar the distribution of the real and synthetic data is. Since all attributes of the NWDAF dataset are numbers, it is measured using the Kolmogorov-Smirnov statistic [36]. To compute this statistic, they convert a numerical distribution. Column Pair Trends show the relationship between two features within the synthetic dataset. For a pair of columns, this test computes a correlation using The Pearson [37] and Spearman [38] rank correlation coefficients. In both indicators, after obtaining the measured value for each column of the data set, the average for the entire column may be obtained to obtain the measured value of the final composite data set. The Overall Quality Score represents the average of the measured Column Shapes and Column Pair Trends values and represents the similarity between the real and synthetic datasets. The evaluation results for the NWDAF synthetic data show that Column Shapes are 96.21%, Column Pair Trends are 98.47% and Overall Quality is 97.34%.

In addition, We use SDMetrics to validate whether the synthetic data adhered to the minimum/maximum error range of the real data. Boundary Adherence returns the percentage of columns of synthetic data that comply with the boundaries of real data. The boundary adherence value of all features in the nwdaf_data dataset is 1.0, indicating that the synthetic data closely matched the real

data. The synthetic data features are within the range of the real data, as illustrated in Figure 8, demonstrating a close resemblance to the real data.

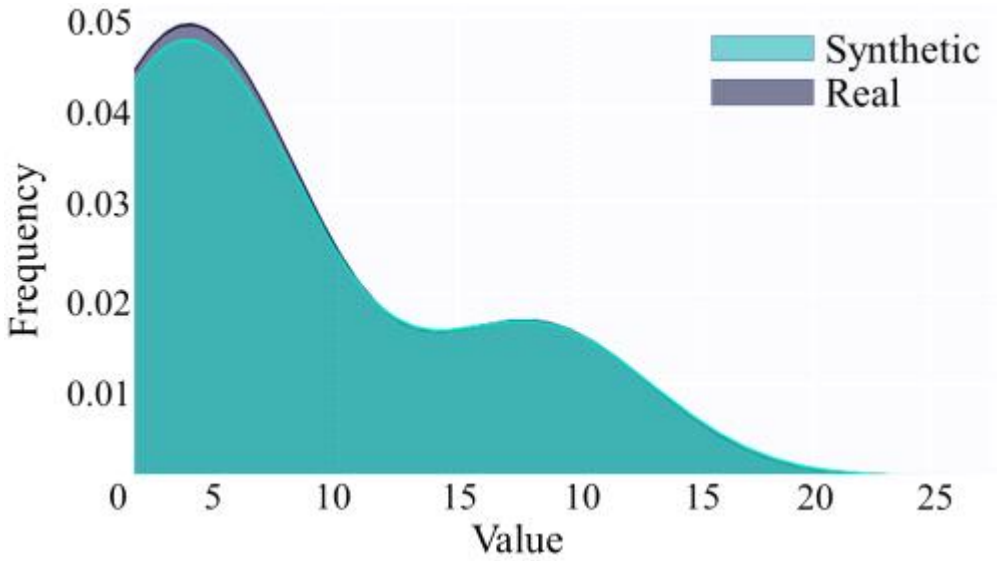


FIGURE 8. Real vs Synthetic Data for column 'load'

3.3 Anomaly detection performance

To evaluate the performance of anomaly detection based on synthetic data, we use the decision tree classifier. For the nwdaf_data dataset, the real data and synthetic data were compared and analyzed separately for five features `t, cell_id, cat_id, pe_id, load' excluding `has_anomaly'. In this case, the dataset's training rate was set to 0.7, and the random_state was set to 42.

TABLE 4. Comparison result based on Decision Tree

	Real Data	Synthetic Data
Accuracy	0.748	0.663
Precision	0.849	0.748
Recall	0.821	0.759
F1	0.835	0.771

Table 4 shows the Accuracy, Precision, Recall, and F1 score for the `has_anomaly' classification on the test dataset. We believe that utilizing the synthetic data for anomaly detection is viable if we further optimize the data quality.

Ⅴ. Enhancing Privacy through Federated Learning of NWDAF using SGX: Anomaly Detection in 5G MEC

The rapid evolution of the 5G network environment increases the demand for machine learning solutions that place importance on safety and privacy. In this context, 5G network-based edge AI plays a pivotal role in providing accurate, low-latency services. The 3GPP has improved quality of service (QoS) and network efficiency by introducing Network Data Analysis (NWDAF) to analyze and process information from various network components. However, the central role of NWDAF in integrating data from multiple sources increases the risk of information leakage, and the adoption of AI-driven network capabilities in cloud infrastructure poses a potential data exposure risk. This paper introduces an architectural approach that leverages the capabilities of Intel Software Guard Extensions (SGX) to enhance data privacy and security within the context of 5G NWDAF federated learning. The proposed framework not only addresses concerns related to data security and data access rights in Cloud Service Providers (CSPs) but also promotes collaborative federated learning across multiple NWDAFs capabilities within the 5G ecosystem. We present a comprehensive analysis of the NWDAF federated learning architecture and its components, highlighting the

pivotal role played by SGX. To validate our approach, we conducted anomaly detection experiments using the NWDAF dataset with Decision Tree, XGBoost, and Logistic Regression models. The proof-of-concept evaluation shows that the proposed framework introduces moderate overhead while achieving enhanced security in the NWDAF federated learning architecture.

In section 5.1, we present a federated learning architecture for NWDAF using SGX. Section 5.2 conducts anomaly detection within the proposed scenario to assess performance with a focus on SGX.

1. Enhancing Data Privacy in NWDAF Federtaed Learning with Intel SGX

The design goal of our proposed system is to mitigate the threat of data leakage in the NWDAF FL scenario and ensure that 5G network customers can securely receive network data analysis results from CSPs when AI-based NFs operate in the cloud. To achieve this, the NWDAF FL server must guarantee secure aggregation while preserving data privacy. Simultaneously, the FL client must operate in a manner that shields individual data points or parameters from identification by unauthorized entities, even in cases where they have been input and generated. However, conventional NWDAF FL approaches have shown vulnerabilities, where sensitive information can be derived from updated parameters tied to local data in the FL Client. These parameters play the vital

role of the aggregator, and the potential for a membership inference attack exists if they are ever exposed or leaked [39, 40]. Consequently, it is imperative to implement a robust security architecture within NWDAF federated learning to safeguard sensitive information and effectively mitigate these risks. As a promising solution in such a scenario, we explore the design that leverages SGX technology in the context of NWDAF FL. If the proposed architecture relies on Intel SGX, it may still be vulnerable to attacks such as DoS and side-channel attacks. However, for the purpose of this paper, it is assumed that SGX itself has no flaws or vulnerabilities.

This section describes the data flow and architecture of NWDAF federated learning, focusing on SGX. Section 5.1.1 discusses the process of federated learning among multiple NWDAFs within a 5G core network. Section 5.1.2 proposes an architecture that leverages SGX technology to enhance data privacy and security in the context of NWDAF FL.

1.1 Data Flow from the NWDAF FL Perspective

Federated learning among multiple NWDAFs is a machine learning technique applied within the core network [41]. It enables the training of machine learning models across various distributed NWDAF clients using their local datasets without the need for sharing these datasets. This approach effectively addresses critical concerns like data privacy, data overloads, and data access rights

[17]. It stands in contrast to traditional centralized machine learning techniques where all NWDAF client datasets are centralized on a single server. Intel SGX technology plays a crucial role in securing these datasets, enhancing trust in the NWDAF server from the customer's perspective. Details of SGX protection are elaborated in Section 5.1.2. In federated learning scenarios supported by multiple NWDAFs, one NWDAF acts as the FL server, while others serve as FL clients.

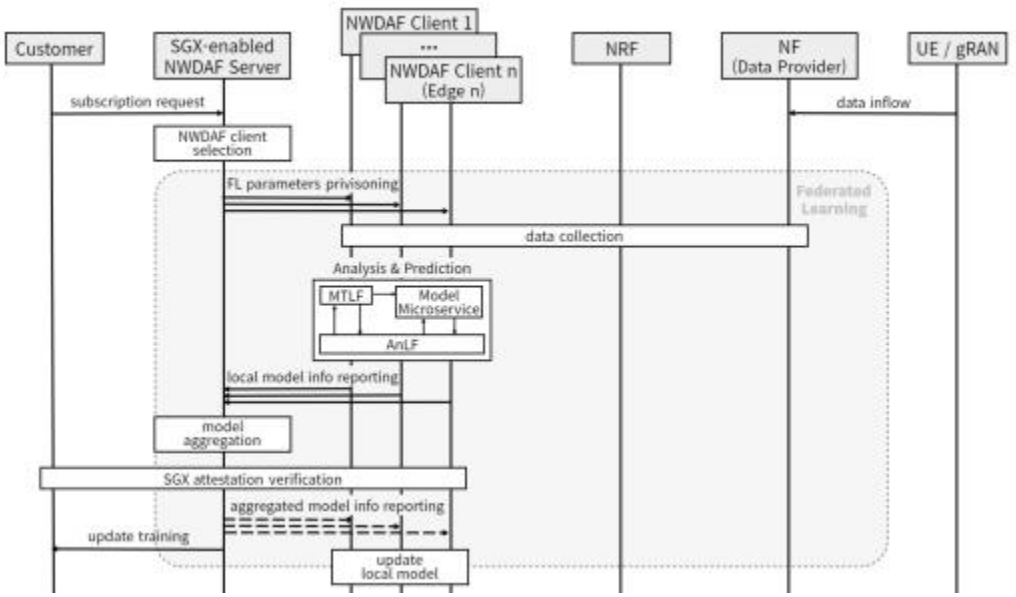


FIGURE 9. Data flow about NWDAF FL

The process of NWDAF federated learning using SGX is depicted in Figure 9, following the data flow. When a customer requests a subscription to the server, the NWDAF FL server with SGX searches

for NWDAF clients willing to participate in the federated learning process, selecting them based on the required number. Each chosen NWDAF client is tasked with local model learning and model information provision. During this phase, the NWDAF client gathers data through NFs and the NRF, which have received data from UE and gNB.

NWDAF clients then train local models using data collected through AnLF, MTLF, and Model Microservice components, respectively, generating model weights. During this process, sensitive data that cannot be shared with other users (such as personal information or data without data access rights) is securely sealed using the SGX sealing function. The learned local ML model information and sealed data are transmitted to the NWDAF FL server. The NWDAF FL server aggregates the local model information of NWDAF clients to create a global ML model. Model aggregation is executed within SGX, and the Intel attestation service verifies the aggregation's integrity to ensure its quality for customers. The aggregated global model is then shared with NWDAF clients, who use it to update their local models and repeat the learning process if necessary.

1.2 NWDAF FL Architecure with SGX

In the context of the described data flow, we offer a comprehensive explanation of the servers and clients employed in NWDAF federated learning, with a particular emphasis on SGX. The

proposed architecture is based on two key assumptions. Firstly, we assume that the source code for the aggregation process itself is trustworthy. Secondly, we consider the edge cloud itself as non-malicious; it's viewed as unreliable solely in terms of input data. And a single cloud service provider has the flexibility to assemble multiple edges as required and exclusively transfer sealed data among these edges. Data sharing becomes possible only when the edge of the cloud service provider shares the same seal key, known as the Common Seal Key (CSK). In practical application, the Intel attestation service efficiently manages this shared seal key on a one-to-one basis.

Figure 10 provides a schematic representation of the federated learning architecture employing SGX. Further details on the scenario specifications of the NWDAF FL server, SGX Enclave, and NWDAF FL client are explained below.

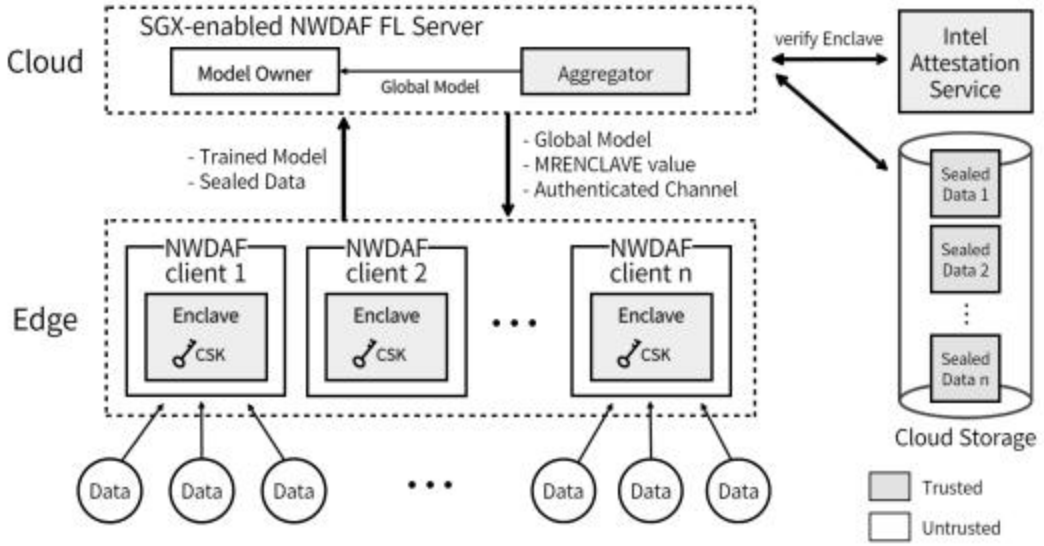


FIGURE 10. High-level overview of NWDAF FL arhcitecture with SGX

Secure Aggregation in NWDAF FL Server. The NWDAF FL server, corresponding to the Central Cloud, is well-equipped with SGX-related resources. It offers analysis codes for data processing within the NWDAF FL server, covering various functions such as anomaly detection, load balancing, and traffic analysis tailored to the specific NWDAF role. Additionally, different types of analysis codes can be deployed as needed.

Before sending the model back to the NWDAF client, the updated models from each client need to be combined. For the sake of robust and secure operations, the model weights must be aggregated within SGX [28]. This unique feature of SGX ensures that only SGX

can identify the source of each weight update. While it's possible to identify which part of the model has changed, the specific NWDAF client responsible for the changes remains private.

Furthermore, because aggregation takes place within the SGX enclave, cloud providers are unable to access the data transmitted by individual clients [28]. Even if the data were accessible, it would remain unreadable without a CSK, utilizing SGX's sealing function. Hence, SGX functionality empowers the NWDAF FL server to securely aggregate data while ensuring the confidentiality and privacy of the transmitted information.

NWDAF Client. The NWDAF FL client, corresponding to the Edge Cloud, can ensure reliable operations through aggregation and analytics processes that operate on Enclave within the central cloud. And each NWDAF client shares the same CSK. This CSK is used to seal data that originated in another NF and entered the NWDAF client before being sent to the NWDAF FL server. Each client sends the weight of the model and the sealed data used in the MTLF to the NWDAF FL server via the AnLF. This helps to trust the enclave and establish confidence in the analysis and anomaly detection processes, even when other entities attempt to maliciously access data or processes.

Comparatively, the Edge Cloud possesses limited SGX-related resources in its local models. Federated learning is employed to enhance the performance of the NWDAF client associated with these local models. During this process, data protection is crucial when

modelers need to share data for handling these local models. It's worth noting that training within SGX can be resource-intensive. To address this, a solution proposed in prior research [28] suggests that only specific layers of the model utilize SGX. This approach mitigates the performance impact by confining only the initial layers of the AI model within the SGX Enclave Page Cache (EPC) while sending the remaining layers outside of the EPC.

Utilizing SGX. SGX plays two pivotal roles within our proposed architecture. Firstly, at the edge stage, each NWDAF client securely stores a Common Seal Key (CSK) within its enclave. This CSK is used to seal the client's data, ensuring its integrity and confidentiality, before safely transmitting the sealed data to the FL server for storage in the Cloud. Secondly, SGX serves as a secure platform for executing the aggregation code within its enclave. SGX acts as a guardian, shielding sensitive data as it enters the NWDAF and as it undergoes analysis through the Anlf of the NWDAF.

SGX hardware incorporates inherent secrets used for secure attestation [15]. These secrets are an integral part of the manufacturing process, with each hardware possessing a unique value. In the context of SGX, NWDAF leverages these built-in secrets to establish secure attestation, thereby identifying the specific SGX hardware in use. The software verification process commences with enclave launch, and the NWDAF Anlf code is loaded into this enclave using specific commands. During this phase, the sensitive data undergoes analysis within a sealed 'black

box.' Meanwhile, the secure hardware log meticulously tracks the Enclave deployment method, known as 'measurement.' This measurement employs cryptographic hashes to verify their contents, comparing them against the expected cryptographic hashes of the built Enclave. If the hashes match, we can confidently demonstrate to the data provider that trustworthy analysis code has been executed within SGX. In essence, from the client's perspective, the Intel attestation service can verify whether the code, deployed by the central cloud for network data analysis, has indeed undergone proper analysis.

2. Experiment: non-SGX vs SGX

We evaluate the proposed architecture through a use case inspired by NWDAF data. In Section 5.2.1, we provide an overview of the NWDAF dataset, which includes the 'has_anomaly' feature. Section 5.2.2 analyzes the anomaly detection evaluation on the NWDAF. Finally, Section 5.2.3 investigates the assessment of SGX overhead associated with anomaly detection runtime on SGX and non-SGX. In this case, for convenience of explanation, the execution flow is described based on one MEC environment. Figure 11 shows the operation process in which abnormal data can be introduced and used in the NWDAF Federated Learning scenario.

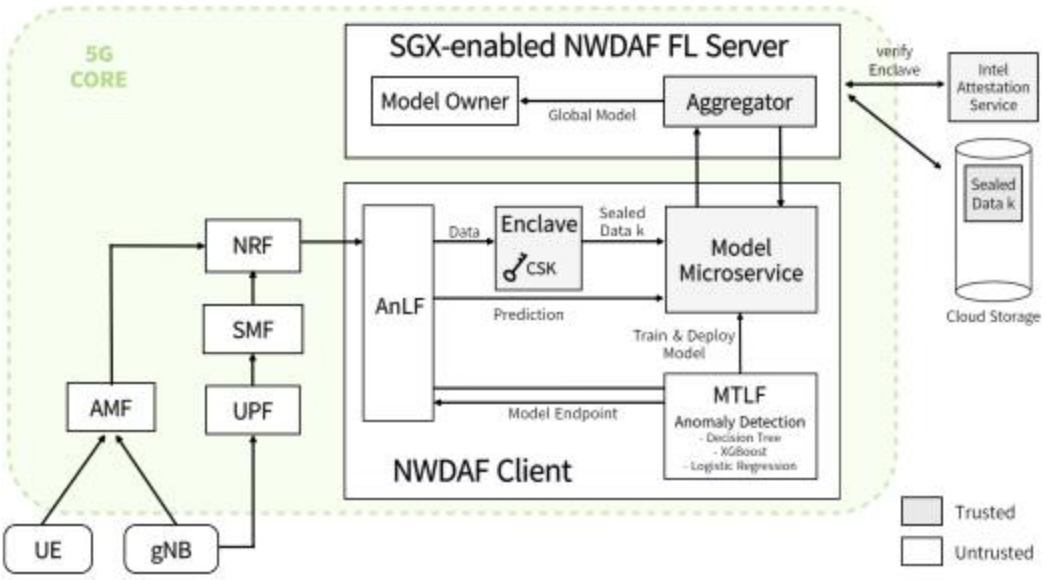


FIGURE 11. Anomaly detection in NWDAF federated learning scenario with SGX

2.1 NWDAF synthetic dataset description

5G network data based on actual traffic is basically the data of mobile network operators, which is difficult to obtain due to privacy concerns and the limited availability of open-source datasets. To solve the collection problem, We used CTGAN [30] to generate synthetic data similar to the characteristics of the original dataset. These synthetic data[1)] were then utilized for the evaluation in this study.

1) In section 4, we evaluated the quality of synthetic data in the NWDAF context and confirmed that they showed performance close to the real data.

2.2 Anomaly Detection Performance

We employed three supervised learning-based machine learning classifiers: Decision Tree, XGBoost, and Logistic Regression, to assess the detection of NWDAF anomalies. The classification was based on the `has_anomaly' class in the NWDAF dataset. For all experimental models, we standardized the learning rate to 0.7 and set the random_state to 15. We evaluated the classifiers using metrics such as Accuracy, Precision, Recall, and F1 score, rounding the results to four decimal places. This experiment was performed in a server consisting of an Intel 3.70GHz Core™ i9-10900K CPU equipped with SGX capability and installing sgx_linux_x64_driver_2.11 version on Ubuntu 20.04.

When the NWDAF dataset resides within SGX and the code to access it is shared outside of SGX, a Permission Denied message is triggered. SGX establishes a secure environment to safeguard sensitive and authorized data, ensuring the protection of the NWDAF dataset within the enclave from unauthorized access or testing. However, in this table of contents, both the code and the dataset were uploaded to the SGX enclave to demonstrate that performance remains guaranteed even when SGX is utilized. Table 5, 6 shows the anomaly detection performance in scenarios without SGX and scenarios with SGX, respectively.

non-SGX. In scenarios without SGX, the Accuracy of anomaly detection models shows performance of at least 68.5% when using the Decision Tree model and up to 76.0% when using XGBoost.

Precision reaches at least 73.5% with the Logical Regression model and 78.8% with the Decision Tree model. For recall, the Decision Tree model achieves a minimum of 78.0%, whereas the Logical Regression model excels with a recall of 99.0%. Finally, the F1 score ranges from at least 78.4% with the Decision Tree model to as high as 85.0% with XGBoost. This content is shown in Table 5.

TABLE 5. Anomaly detection performance in non-SGX

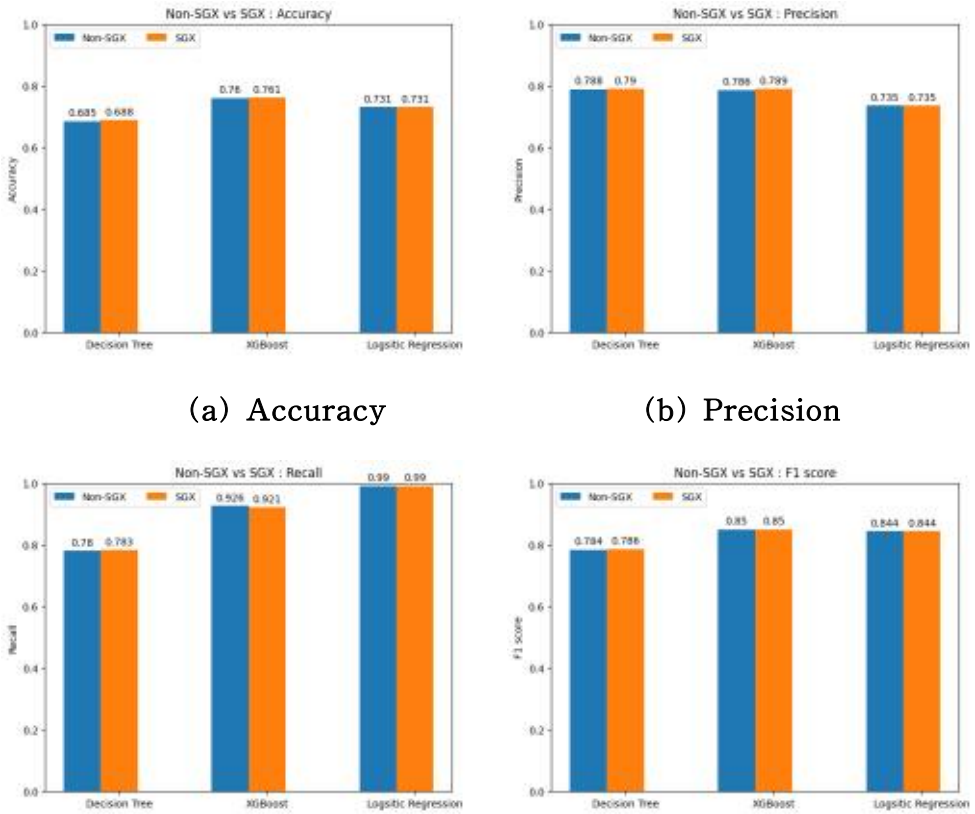
	Decision Tree	XGBoost	Logisitc Regression
Accuracy	0.685	0.760	0.731
Precision	0.788	0.786	0.735
Recall	0.780	0.926	0.990
F1	0.784	0.850	0.844

SGX. In scenarios with SGX, the Accuracy of anomaly detection models shows performance of at least 68.8% when using Decision Tree models and up to 76.1% when using XGBoost. Precision is a minimum of 73.5% with the Logical Regression model and rises to 79.0% with the Decision Tree model. The Decision Tree model attains a recall of at least 78.3%, and the Logical Regression model excels with a recall of 99.0%. In terms of the F1 score, the Decision Tree model delivers a minimum of 78.6%, with XGBoost achieving as high as 85.0%. This content is shown in Table 6.

TABLE 6. Anomaly detection performance in SGX

	Decision Tree	XGBoost	Logisitc Regression
Accuracy	0.688	0.761	0.731
Precision	0.790	0.789	0.735
Recall	0.783	0.921	0.990
F1	0.786	0.850	0.844

The comparison of Tables 5 and Table 6 is visualized in Figure 12. The results show that there is minimal performance overhead when SGX is utilized for each matrix Figure 12-(a), Figure 12-(b), Figure 12-(c), and Figure 12-(d). This is similar to the results obtained when SGX is not used at all.



(a) Accuracy (b) Precision

(c) Recall (d) F1 score

FIGURE 12. Anomaly detection performance: non-SGX vs SGX

2.3 SGX Overhead

Gramine LibOS. Gramine [42], formerly known as Graphene, is a lightweight library operating system designed for Linux multi-process applications. It is implemented in C and offers support for Intel SGX. To ensure SGX compatibility, the library OS can be tailored to

provide a glibc-like interface. The key feature of Gramine is its ability to run existing applications within SGX enclaves without requiring any modifications to the applications themselves. Instead, you need to provide an accompanying manifest file detailing security settings and configuration. Gramine then uses this manifest file to verify authenticity, ensure integrity, and load applications along with their dependencies. Gramine currently offers two backends: gramine-direct running on the host Linux OS and gramine-sgx running inside Intel SGX enclave [43].

Runtime Comparison. Gramine provides a time() system call that can be timed to run. We show comparison results according to SGX usage for inference latency of each model considering execution time using Gramine. Figure 13 compares the execution times of each model with SGX and without SGX.

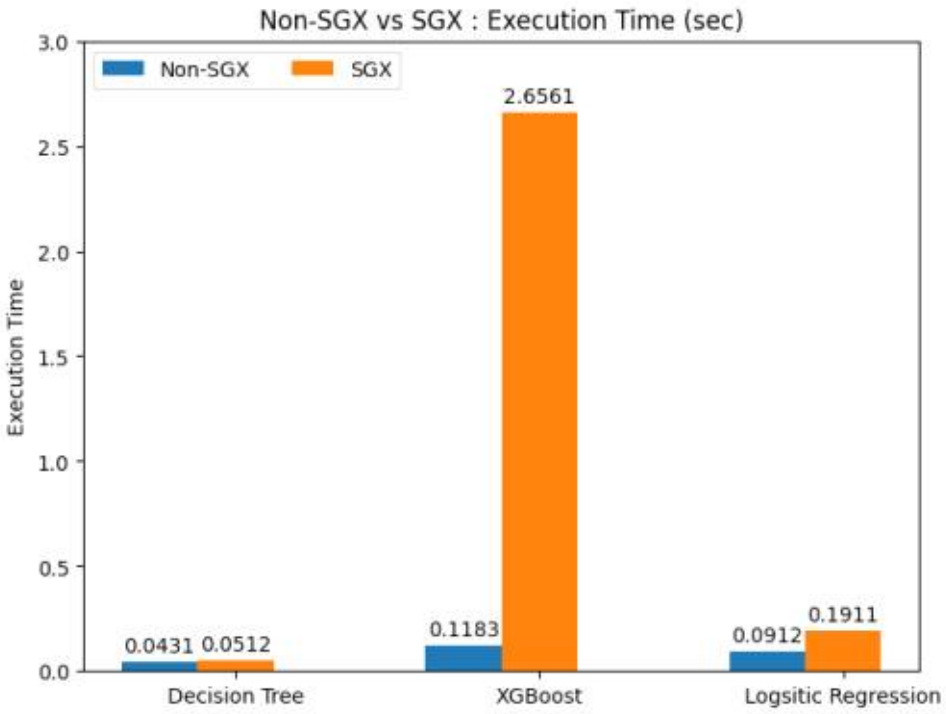


FIGURE 13. Inference execution time: non-SGX vs SGX

Inference latency refers to the duration taken to dispatch an information request to the classifier and receive a response. In particular, the use of SGX itself has little overhead for Decision Tree and Logical Regression models. However, a slight variance is observed when emulating the raw syscall command, as seen in the case of XGBoost. Employing raw syscalls in Gramine causes an overhead that necessitates raw emulation, consequently diminishing performance. To mitigate this overhead, improvements can be achieved by patching applications and employing the Gramine Syscall API.

Ⅵ. Related Work

This section introduces meaningful studies related to our research.

Chafika Benzaid et al. [12] explored how artificial intelligence can affect 5G network security. It emphasizes that there is no all-in-one solution to solve all artificial intelligence threats in 5G networks, so it should be partially divided to apply defense techniques. Accordingly, it represents an ITU-T integrated architecture for artificial intelligence and machine learning in 5G networks and specifies attack surfaces for data contamination attacks, avoidance attacks, and API-based attacks in this architecture. It also presents defense techniques that can be used on each surface. However, the proposed attack and defense techniques deal only with data injection, manipulation, and logic corruption attacks from a general point of view, making it difficult to regard them as specific to 5G networks, and have limitations as discourse techniques. In addition, the proposed architecture also did not fully consider processes such as how 5G network functions collect, preprocess, and transmit/receive data.

Yalin E et al. [13] identified new attack surfaces and corresponding attacks of adversarial machine learning for wireless communication in 5G systems and discussed the major vulnerabilities of 5G systems for adversarial machine learning.

Specifically, they considered a scenario that interrupts 5G communication by learning a 5G deep learning pattern that shares a spectrum between 5G and user terminals and jamming data and control signals, and a spoofing attack-based adversarial attack scenario for an attacker to pass through a deep learning-based physical layer authentication system in 5G gNodeB. At this time, the attacker sends a similar signal to access the 5G network slicing application to gain permission and carry out a hostile attack. That is, an attacker can wirelessly learn synthetic data to generate a spoofing signal and transmit it to penetrate 5G signal authentication. In the case of this paper, it has contributions in that it creates intentional errors in 5G systems to allow users to learn inaccurate models and proves the possibility of hostile attacks that negatively affect performance. However, in the case of this paper, no consideration was given to the edge cloud, a key paradigm of 5G networks. Because numerous terminal devices generate vast amounts of data on 5G networks, distributed learning takes place in MECs, and federated learning performed in conjunction with central cloud servers is drawing attention, identification of attack areas should also be considered [16]. Specifically, in a federated learning scenario considering edge clouds, a potential data contamination attack that can use malicious data as input from a local model can occur. In addition, there is a difference from this study that considers data contamination attacks in the core network in that it targets wireless access networks.

Zhou et al. [44] aimed to address the challenge of bridging the gap between different NWDAF datasets, which often contain unique data characteristics. To tackle this issue, they introduced a novel approach called Partial Homomorphic Encryption (PHE)-based federated learning. Traditional Differential Privacy (DP)-based federated learning faces a problem where the global model's robustness deteriorates when the amount of noise introduced exceeds a certain threshold. To overcome this limitation, they employed PHE-based federated learning, which enhances the model's resilience. In their simulation, they utilized one federated learning server and engaged with 100 clients, training on the MNIST dataset. For the experiment, they generated HE (Homomorphic Encryption) key pairs in advance for all clients. They then selected 10 clients at random to encrypt each weight value of the multi-perceptron model during 10 mini-batches and 10 epochs of training. The results demonstrated that their HE-based approach, designed to ensure the secure utilization of federated learning in NWDAF architecture, outperformed the DP-based federated learning method by 1.5 times. Furthermore, it exhibited similar performance to the federated learning architecture without HE when evaluated for the accumulation of Multi-Layer Perceptron (MLP) models. However, it is important to note that they use datasets irrelevant to NWDAF for evaluation, making it challenging to conclusively prove its applicability to NWDAF-specific scenarios. In comparison, our study is effective for evaluation in that we used NWDAF datasets collected

from the 5G network.

P. Rajabzadeh and A. Outtagarts [45] introduced a centralized federated learning approach within a distributed NWDAF setting. This method incorporated Local Differential Privacy (LDP) and a Feedback Mechanism, demonstrating its effectiveness compared to the traditional centralized NWDAF approach. In this study, they implemented four NWDAF instances as virtual machines and developed 9 NFs subscribed to NWDAF. These experiments were conducted using the NetSoft2020-Tutorial4-Demo2-Exp1 framework from LABORA [46]. When they compared Centralized Machine Learning with Centralized Federated Learning in the proposed distributed NWDAF environment, utilizing a lightweight LSTM prediction model, the results showed that the proposed method significantly improved prediction accuracy for network function CPU usage. Specifically, for the prediction of network function CPU usage, the proposed method achieved up to a 4.7% higher accuracy, and the prediction time was reduced by 16.6 seconds. In summary, the study indicated that the distributed NWDAF architecture outperformed the traditional centralized NWDAF architecture. However, it's essential to note that there is a limitation in this research. The simulation of the NWDAF environment used Docker containers, and it's important to protect the input data for the NWDAF to prevent potential security attacks on the distributed NWDAF architecture. In contrast, our study aims to minimize the risk of data leakage by safeguarding the input data to NWDAF using

SGX during the federated learning process.

Jorquera et al. [47] introduced a novel concept of TEE-as-a-service, aiming to seamlessly integrate hardware-based solutions with 5G technology. This innovation addresses previous challenges associated with virtualized 5G networks relying on third-party infrastructure, which often suffered from trust issues. They proposed leveraging Intel SGX for securing 5G multi-domain networks. The focus was on safeguarding 5G data and applications not only during runtime within a security zone but also during transmission and shutdown phases. The proposed solution aimed to establish a secure environment for 5G core network components, such as the Virtualized Infrastructure Manager (VIM) and Virtualized Network Function (VNF), especially when deployed in third-party infrastructures. However, it's worth noting that the paper lacks a proof of concept for the proposal, and therefore, the suitability of 5G components for memory-limited SGX was not demonstrated. On a positive note, our study provided a proof of concept for the feasibility of utilizing SGX in 5G data analysis and anomaly detection, specifically within the 5G NWDAF.

Zheng Yan et al. [48] have introduced Intel SGX as a comprehensive privacy solution integrated into 5G positioning and location-based services (LBS) within the 5G positioning service chain. They identified a vulnerability in which sensitive information about the position owner could be inferred from User Equipment (UE) location privacy and LBSP (Location-Based Service Provider)

data used in the 5G positioning service chain. They address this concern by employing Intel SGX to ensure the confidentiality of UE locations in edge form and LBSP data assets. By utilizing the security features provided by the Enclave during the UE positioning process, the UE's location remains undisclosed to the Fusion Center (FC) during location calculation and is not leaked to the LBSP during the provision of LBS. Moreover, secure secret sharing between UEs in edge scenarios supports multiple location-based service providers without requiring frequent key exchanges. Notably, the Enclave is created only when necessary and promptly discarded after the service is completed. This study marks a significant contribution as the first method to provide mutual privacy for both UE and LBSP. The proposed framework is appropriately demonstrated through a comparison and analysis of the execution time of Intel SGX's Enclave. Additionally, our study suggests that Intel SGX can extend its utility beyond securing location information, emphasizing its potential applicability in 5G data analysis and anomaly detection.

Ⅶ. Discussion

In this section, we discuss the limitations of this paper as compared to its contributions and outline the future research direction related to the contents of section 4 and 5.

We believe that this work is highly beneficial for data analysis and anomaly detection cases of NWDAF in a 5G MEC. The proposed scenario suggests that federated learning can reduce the cost of transmitting massive amounts of raw data, which in turn reduces the relative performance overhead of the data analysis process. Additionally, applying Intel SGX can mitigate security risks such as data leakage that may occur during the federated learning process.

However, a limitation of this work is that it is challenging to claim it as a perfect result for 5G MEC as it used the NWDAF dataset obtained from simulated studies instead of an established 5G MEC environment. Therefore, to obtain the best experimental results, it is necessary to create a 5G MEC environment that can be combined with Intel SGX, such as 5G simulator, as a follow-up study. Furthermore, it needs to be adjusted to suit specific applications. According to the proposed architecture, the NWDAF dataset is located within the enclave of SGX, and the data analysis results are also located within the enclave. Data protection is ensured as the data in the enclave is not accessible from the outside due to the

characteristics of Intel SGX. However, several evaluations need to be performed in terms of data protection. We can conduct additional experiments by referring to existing studies [49, 50].

Regarding Section 4, there are methods to safely generate synthetic data without combining Intel SGX. When predicting the success rate of the Membership Inference Attack (MIA) using the SDV library for the synthetic data generated in Section 4, a dangerous result of about 88.1% was obtained. To mitigate these attacks, we emphasize the need to introduce a trusted execution environment such as Intel SGX. Other studies have shown the possibility of generating synthetic data while preserving privacy when applying differential privacy [51]. Therefore, various studies are needed to achieve privacy preservation and data protection such as differential privacy and Intel SGX.

In Section 5, among the federated learning, Decision Tree, Logical Regression, and XGBoost methods used for the anomaly detection experiment, the ensemble algorithm is excellent when combined with federated learning. Instead of a decision tree, it is possible to increase performance by replacing it such as a Gradient Boosting Decision Tree and an Autoencoder [52, 53, 54]. As a follow-up study, we need to explore which model can perform well in federated learning among the three classifiers used in this paper, Decision Tree, Logic Regression, and XGBoost, and which model is suitable for the federated learning scenario in some cases to improve performance.

Ⅷ. Conclusion

As the evolution of 5G networks continues and transitions towards standalone execution modes, the imperative need for secure and privacy-conscious solutions in the realm of data analytics and AI-driven network functions becomes increasingly evident. At this time, 5G NFs can be strategically deployed in base stations, which serve as a rich source of data that can provide valuable network-wide insights into the role of NWDAFs. Federated learning was introduced to effectively process a large amount of data entering the NWDAF. However, when data collected from NF for training federated learning models are sent to a central server such as NWDAF, there are potential security risks, especially with regard to end-user mobility patterns and other sensitive information. In addition, the introduction of cloud infrastructure and AI-based network capabilities introduces new security risks associated with data exposure, especially through membership inference attacks.

To mitigate these risks effectively, this paper proposes the incorporation of TEE, notably Intel SGX, within NWDAF federated learning scenarios. The proposed deployment scenario, featuring TEE, serves as a demonstrative model for enhancing data privacy through robust data access controls. The overarching objective of 5G NWDAF federated learning with SGX is to prioritize data

protection, ensure analysis code integrity, and enable secure model weight aggregation. To substantiate this approach, an anomaly detection model such as Decision Tree, XGBoost, and Logistic Regression was successfully executed on the NWDAF dataset using SGX, validating its effectiveness. We believe our approach contributes to developing secure and privacy-conscious solutions for 5G network data analytics, thereby ensuring data privacy, integrity, and the realization of the full potential of network analytics in 5G environments.

References

[1] Albert Chun Chen Liu, Oscar Ming Kin Law, Jeremiah Liao, Jeffrey Y.C. Chen, Andy Jia En Hsieh, and Cheng Hung Hsieh. “Traffic safety system edge ai computing,” In 2021 IEEE/ACM Symposium on Edge Computing (SEC), pages 01-02, 2021.

[2] S.W.Hong, C.S.Lee, S.C.Kim, S.Moon, J.C.Shim, S.B.Hong, and H. Y.Ryu. “Technologies of intelligent edge computing and networking,” In Electronics and Telecommunications Trends, volume 34, pages 23-35, 2019.

[3] Electronics and Telecommunications Research Institute. “Development of network data analytics function (nwdaf) and intelligence technology standards for 5g network automation,” Technical Report ETRI TRKO202100009073, 2021.

[4] 3GPP. “Release 16 description; summary of rel-16 work items,” Technical Report 3GPP TR 21.916, 2020.

[5] Sidi-Mohammed Senouci, Hichem Sedjelmaci, Jiajia Liu, Mubashir Husain Rehmani, and Elias Bou-Harb. “Ai-driven cybersecurity threats to future networks,” IEEE Vehicular Technology Magazine, 15: 5-6, 09 2020.

[6] Syed Hussain, Omar Chowdhury, Shagufta Mehnaz, and Elisa Bertino. “Lteinspector: A systematic approach for adversarial testing of 4g lte,” 2018.

[7] Chuan Yu, Shuhui Chen, Fei Wang, and Ziling Wei. “Improving 4g /5g air interface security: A survey of existing attacks on differen t lte layers,” Comput. Netw., 201(C), dec 2021.

[8] He Fang, Xianbin Wang, and Stefano Tomasin. “Machine learning for intelligent authentication in 5g and beyond wireless networks,” IEEE Wireless Communications, 26(5):55-61, 2019.

[9] 3GPP. “System architecture for the 5g system (5gs),” Technical Re port 3GPP TS 23.501, 2018.

[10] 3GPP. “Architecture enhancements for 5g system (5gs) to support network data analytics services.” Technical Report 3GPP TS 23.28 8, 2018.

[11] 3GPP. “5g; 5g system; network function repository services; stage 3.” Technical Report 3GPP TS 29.510, 2019.

[12] Chafika Benza¨ıd and Tarik Taleb. “Ai for beyond 5g networks: A cyber-security defense or offense enabler?” IEEE Network, 34(6):14 0-147, 2020.

[13] Yalin E. Sagduyu, Tugba Erpek, and Yi Shi. “Adversarial machin e learning for 5g communications security,” arXiv:2101.02656, 202 1.

[14] Matthew Hoekstra, Reshma Lal, Pradeep Pappachan, Vinay Phega de, and Juan Del Cuvillo. “Using innovative instructions to create trustworthy software solutions.” HASP@ ISCA, 11(10.1145):2487726-2488370, 2013.

[15] Victor Costan and Srinivas Devadas. “Intel sgx explained.” Crypto

logy ePrint Archive, 2016.

[16] Dimitrios Moustis and Panayiotis Kotzanikolaou. “Evaluating security controls against http-based ddos attacks.” In IISA 2013, pages 1-6, 2013.

[17] Youbin Jeon, Hyeonjae Jeong, Sangwon Seo, Taeyun Kim, Haneul Ko, and Sangheon Pack. “A distributed nwdaf architecture for federated learning in 5g.” In 2022 IEEE International Conference on Consumer Electronics (ICCE), pages 1-2, 2022.

[18] Seungyeol Lee and Myung-Ki Shin. “Federated learning over private 5g networks: Demo.” In Proceedings of the Twenty-Third International Symposium on Theory, Algoꟍrithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, MobiHoc ’22, page 295-296, New York, NY, USA, Association for Computing Machinery, 2022.

[19] Esa Piri, Pekka Ruuska, Teemu Kanstren, Jukka Makela, Jari Korva, Atso Hekkala, A. Pouttu, Olli Liinamaa, Matti Latva-aho, Kari Vierimaa, and Harri Valasma. “5gtn: A test network for 5g application development and testing.” pages 313-318, 06 2016.

[20] Girish Chandrashekar and Ferat Sahin. “A survey on feature selection methods.” Comꟍputers & Electrical Engineering, 40(1):16-28, 40th-year commemorative issue, 2014.

[21] Nooritawati Md Tahir, Aini Hussain, Salina Abdul Samad, Khairul Anuar Ishak, and Rosmawati Abdul Halim. “Feature selection for classification using decision tree.” In 2006 4th Student Conference on Research and Development, pages 99-102, 2006.

[22] [Online] Argus. Argus: System + network monitoring. https://jaw0.github.io/argus5docs/docs/, (accessed on March 2023).

[23] [Online] Stratio. https://www.stratio.com, (accessed on June 2023).

[24] [Online] Oracle. Oracle's fusion analytics warehouse. https://docs.oracle.com/en/cloud/saas/analytics/22r2/fawag, (accessed on June 2023).

[25] Mohamed Almorsy, John Grundy, and Ingo Muller. "An analysis of the cloud computing security problem," 2016.

[26] Saeid Abolfazli, Zohreh Sanaei, Ejaz Ahmed, Abdullah Gani, and Rajkumar Buyya. "Cloud-based augmentation for mobile devices: Motivation, taxonomies, and open challenges." IEEE Communications Surveys & Tutorials, 16(1):337-368, 2014.

[27] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples," 2015.

[28] Eugene Kuznetsov, Yitao Chen, and Ming Zhao. "Securefl: Privacy preserving federated learning with sgx and trustzone." In 2021 IEEE/ACM Symposium on Edge Computing (SEC), pages 55-67, 2021.

[29] So-Eun Jeon, Ji-Won Ock, Min-Jeong Kim, Sa-Ra Hong, Sae-Rom Park, and Il-Gu Lee. "Efficient Poisoning Attack Defense Techniques Based on Data Augmentation." Jouranl of Information and Security, 22:25-32, 2022.

[30] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan V

eeramachaneni. “Modeling tabular data using conditional gan.” Advances in neural information processing systems, 32, 2019.

[31] Dae gyeom Kim, Myeong jin Ko, Sung woo Moon, Sung hyun Kim, Kyung-Yul Cheon, Seungkeun Park, Yunbae Kim, Hyungoo Yoon, and Yong-Hoon Choi. “Generation of 5g traffic using lstm-dccnn gan.” In Proceedings of Symposium of the Korean Institute of communications and Information Sciences (KICS) Summer Conference, pages 1245-1246, 2021.

[32] MK Shin, SH Lee, and JH Yi. “Trends of 5g network automation and intelligence technologies standardization.” Electronics and Telecommunications Trends, 34(2):92-100, 2019.

[33] 3GPP. “Study on 5g transport network enhancements for the support of urllc and mmtc.” Technical Report 3GPP TR.23.791, 2018.

[34] Salih Sevgican, Meric¸ Turan, Kerim Gokarslan, H. Birkan Yilmaz, and Tuna Tugcu. “Intelligent network data analytics function in 5g cellular networks using machine learning.” Journal of Communications and Networks, 22(3):269-280, 2020.

[35] [Online] SDV. Evaluating single table data-sdv 0.10.0 documentation. https://docs.sdv.dev/sdv/single-table-data/evaluation, (accessed on March 2023).

[36] [Online] Kolmogorov-Smirnov test. In Wikipedia. https://en.wikipedia.org/wiki/Kolmogorov_Smirnov_test, (Accessed on December 2023).

[37] [Online] Pearson correlation coefficient. In Wikipedia. https://en.

wikipedia.org/wiki/Pearson_correlation_coefficient, (Accessed on December 2023).

[38] [Online] Spearsman correlation coefficient. In Wikipedia. https://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient, (Accessed on December 2023).

[39] Yi Liu, Jialiang Peng, Jiawen Kang, Abdullah M. Iliyasu, Dusit Niyato, and Ahmed A. Abd El-Latif. “A secure federated learning framework for 5g networks.” IEEE Wireless Communications, 27(4):24-31, 2020.

[40] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. “Membership inference attacks against machine learning models.” In 2017 IEEE Symposium on Security and Privacy (SP), pages 3-18, 2017.

[41] Solmaz Niknam, Harpreet S. Dhillon, and Jeffrey H. Reed. “Federated learning for wireless communications: Motivation, opportunities, and challenges.” IEEE Communications Magazine, 58(6):46-51, 2020.

[42] Chia-Che Tsai, Donald E Porter, and Mona Vij. “Graphene-sgx: A practical library os for unmodified applications on sgx.” In 2017 USENIX Annual Technical Conference (USENIX ATC 17), pages 645-658, 2017.

[43] [Online] Gramine. Gramine documentation, https://gramine.readthedocs.io/en/stable/, (accessed on October 2023).

[44] Changshi Zhou and Nirwan Ansari. “Securing federated learning

enabled nwdaf architecture with partial homomorphic encryptio n." IEEE Networking Letters, 2023.

[45] Parsa Rajabzadeh and Abdelkader Outtagarts. "Federated learning for distributed nwdaf architecture." In 2023 26th Conference on In novation in Clouds, Internet and Networks and Workshops (ICIN), pages 24-26, 2023.

[46] [Online] C. Macedo and K. Cardoso. Netsoft2020-tutorial4-demo2-exp1 (labora), 2020. https://github.com/LABORA-INF-UFG/NetSoft2020-Tutorial4-Demo2-Exp1, (accessed on October 2023).

[47] Jose Maria Jorquera Valero, Pedro Miguel Sanchez Sanchez, Alexi os Lekidis, Pedro Martins, Pedro Diogo, Manuel Gil Perez, Alberto Huertas Celdran, and Gregorio Martinez Perez. "Trusted Execution Environment-Enabled Platform for 5G Security and Privacy Enhan cement," pages 203-223. Springer International Publishing, Cham, 2022.

[48] Zheng Yan, Xinren Qian, Shushu Liu, and Robert Deng. "Privacy protection in 5g positioning and location-based services based on sgx." ACM Trans. Sen. Netw., 18(3), Aug 2022.

[49] Junwei Luo, Xuechao Yang, and Xun Yi. "SGX-based Users Matc hing with Privacy Protection," In Proceedings of the Australasian Computer Science Week Multiconference (ACSW '20). Association f or Computing Machinery, USA, Article 4, pages 1-9. 2020.

[50] Lin, Dan, and Anna Squicciarini. "Data protection models for ser vice provisioning in the cloud." Proceedings of the 15th ACM sym posium on Access control models and technologies. pages 183-19

2. 2010.

[51] Jiwon Ock, Taewhi Lee, and Seongmin Kim, "Privacy-preserving Approximate Query Processing with Differentially Private Generative Models," 2023 IEEE International Conference on Big Data (Big Data), Sorrento, Italy, pages 6242-6244, 2023.

[52] Q. Li, Z. Wen, and B. He, “Practical Federated Gradient Boosting Decision Trees”, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 04, pages 4642-4649, Apr. 2020.

[53] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha and G. Srivastava, "Federated-Learning-Based Anomaly Detection for IoT Security Attacks," in IEEE Internet of Things Journal, 9 (4), pages 2545-2554, Feb, 2022.

[54] Weinger, B., Kim, J., Sim, A., Nakashima, M., Moustafa, N., and Wu, K. J. Enhancing IoT anomaly detection performance for federated learning. Digital Communications and Networks, 8(3), pages 314-323, 2020.

논문 개요

5G 모바일 에지 컴퓨팅에서 AI 기반 데이터 보안 및 이상 탐지 향상에 관한 연구

옥지원
미래융합기술공학과
성신여자대학교 대학원

5G 모바일 에지 컴퓨팅이 주목받는 가운데, 데이터 보안 및 이상 탐지는 여전히 핵심 과제로 남아 있다. AI 기반 솔루션의 도입은 이러한 과제에 대한 유망한 해결책으로 나타나고 있다. 기계 학습 알고리즘과 예측 모델을 활용함으로써 조직은 데이터 보안과 이상 탐지 기능을 획기적으로 강화할 수 있다. 5G 모바일 에지 컴퓨팅에 AI 기반 솔루션을 통합함으로써, 조직은 실시간으로 네트워크 트래픽 패턴을 모니터링하고 분석할 수 있다. 이러한 실시간 분석은 잠재적인 위협과 이상 징후를 신속히 감지하여 조직이 사전 조치를 취할 수 있도록 지원한다. 게다가, 인공지능 기반 솔루션은 데이터 침해를 감지하고 예방하는 데 도움이 될 수 있다. 사용자 행동 패턴을 식별하고 분석함으로써, 인공지능 기반 솔루션은 의심스러운 활동을 탐지하고 승인되지 않은 접근을 차단할 수 있다. 5G 모바일 에지 컴퓨팅에 AI 기반 솔루션을 통합함으로써, 데이터 보안과 이상 탐지 기능이 크게 향상될 수 있다. 조직은 이러한 솔루션을 활용하여 시스템과 데이터를 보호하면서도 고객에게 중단 없는 서비스를 제공하는 것을 고려해야 한다. 본 논문은 5G 모바일 에지 컴퓨팅 환경에서 인공지능에 활용될 수 있는 NWDAF 네트워크 기능에 대해 탐구한다. Intel SGX를 사용하여 적절한 데이터 보호 아키텍처를 제시하고, 시뮬레이션을 통한 5G상 NWDAF 이상 탐지를 수행한다.

ACKNOWLEDGEMENTS

본 논문을 지도해주신 김성민 교수님께 감사드립니다.