B5GCASP: Decentralized Federated Anomalous Signaling Protection Architecture Using Functionally Layered Network

Cong Li, Xingxing Liao, Zilong Wang, *Member, IEEE*, Xinsheng Ji, and Guoqiang Mao, *Fellow, IEEE*

***Abstract*—As the brain of B5G networks, the core networks enable more ubiquitous intelligent connectivity over previous generations of mobile networks, thanks to the decentralized user plane close to edges. As mitigation against abnormal signaling attacks on edge core networks, signal protection mechanisms for the user planes at the N4 interface are widely investigated. However, the prior art fails to adequately address the distribution characteristics of abnormal signaling at this interface, where single-point defences are insufficient for the complexities of beyond 5G (B5G) distributed architecture. This article proposes a decentralized federated anomaly signaling protection framework, called B5GCASP, based on a functionally layered anomalous signaling detection model (FLAD). Mainly, B5GCASP analyses abnormal signaling distribution under the packet forwarding control protocol (PFCP) at the N4 interface, distinguishing significant and nonsignificant anomalies. Coupled with a decentralized, federated detection mechanism, B5GCASP creates a comprehensive point-and-area detection architecture. Extensive experiments on the 5GC PFCP dataset show that B5GCASP achieves higher accuracy and faster detection of abnormal signaling compared to single-point defending baselines, which offer robust anomaly signaling protection for the B5G core network.**

***Index Terms*—Abnormal detection, beyond 5G (B5G), packet forwarding control protocol (PFCP), point-and-area detection, user plane.**

I. Introduction

With the evolution of beyond 5G (B5G) networks, the era of ubiquitous connectivity is approaching. B5G aims to create an integrated network that spans space, air, oceans, and land, providing seamless global information coverage [1]. Efforts from various countries and research institutions are accelerating the development of B5G networks, focusing on the core network.

Received 12 December 2024; revised 31 December 2024; accepted 9 January 2025. Date of publication 13 February 2025; date of current version 23 May 2025. This work was supported by the National Key Research and Development Program of China under Grant 2022YFB2902204. *(Corresponding author: Xinsheng Ji.)*

Cong Li and Zilong Wang are with the State Key Laboratory of Integrated Service Networks, School of Cyber Engineering, Xidian University, Xi'an 710126, China (e-mail: licong55@126.com; zlwang@xidian.edu.cn).

Xingxing Liao is with Purple Mountain Laboratories, Nanjing 210000, China (e-mail: liaoxingxing@pmlabs.com.cn).

Xinsheng Ji is with Tsinghua University, Beijing 100000, China (e-mail: jixs@pmlabs.com.cn).

Guoqiang Mao is with the Research Institute of Smart Transportation, Xidian University, Xi'an 710126, China (e-mail: g.mao@ieee.org).

Digital Object Identifier 10.1109/JIOT.2025.3531047

According to the third generation partnership project (3GPP), key issues in B5G core networks include data transmission security, access control, authorization, and privacy protection [2]. One primary concern is that the high volume of signaling traffic in IoT environments may hinder detecting anomalous signaling attacks. Additionally, as B5G networks incorporate technologies like network slicing, virtualization, and software defined networking (SDN) [3], they become more flexible and complex, introducing new attack vectors. Adequate signaling protection is, therefore, crucial to safeguard critical applications, especially in scenarios requiring real-time data integrity, such as in smart cities and remote healthcare.

In multi-intelligent scenarios with massive user access, attackers can easily conceal their activities by launching abnormal signaling attacks toward the control plane from the user plane, as illustrated in Fig. 1. The user plane function (UPF), responsible for processing and forwarding user data to the control plane via the packet forwarding control protocol (PFCP) over the N4 interface, becomes a critical target for attackers. The N4 interface, therefore, represents a key vulnerability. Robust detection and monitoring anomalous signaling at the N4 interface are crucial to maintaining B5G core network security and integrity. However, the complexity of the B5G core network environment, which encompasses technologies, such as network slicing, virtualization, and SDN, along with diverse network architectures and functionalities, leads to intricate interactions among network functions (NFs). This complexity creates numerous opportunities for anomalous signaling attacks and complicates their detection. The core network must manage dense, high-speed, and concurrent user plane data flows, rendering anomaly detection at the N4 interface inherently challenging. Moreover, the stealthy nature and the sophisticated characteristics of signaling attacks further exacerbate this difficulty.

It is worth noting that research on detecting anomalous signaling at the user plane N4 interface needs to be improved, primarily due to the scarcity of datasets under the PFCP and the challenges associated with identifying specific attack features. Panagiotis proposed an intrusion detection system (IDS) for the 5G core (5GC) that utilizes artificial intelligence to detect potential network attacks targeting the N4 interface under PFCP [4]. This work is notable for using a real PFCP dataset, significantly advancing 5GC anomaly detection. However, it does not adequately address the specific

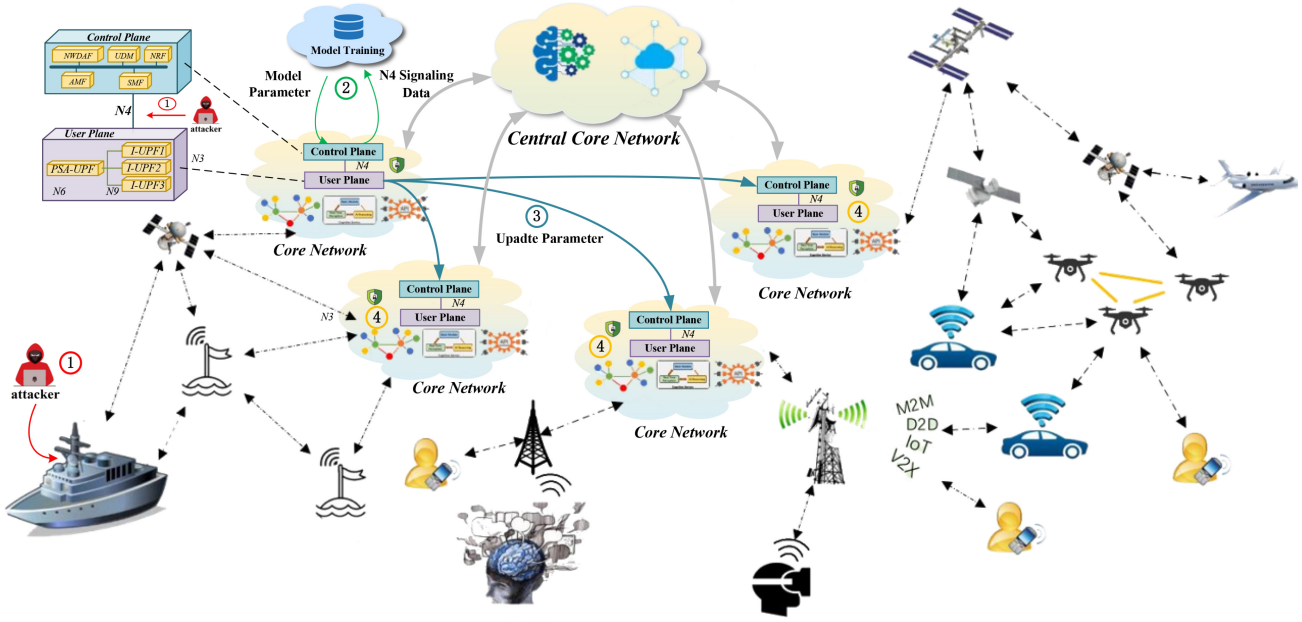

Fig. 1. Decentralized federated learning anomaly signaling protection architecture under the B5GC consists mainly of the user and control planes. In this architecture, ① a single core network at the N4 interface is attacked by an attacker launching anomalous signaling attacks. ② the attacked core network can capture and learn some new anomaly signaling features. ③ the updated weights containing the new anomaly features are sent to other connected core networks. Moreover, ④ when the attacker targets other core networks again, the core networks that received the new model parameters can effectively resist the attack promptly.

attack features associated with PFCP anomalous signaling, which limits its detection accuracy. Furthermore, by the 3GPP standards, the B5G core network (B5GC) is anticipated to evolve toward a more distributed architecture. This evolution will increase the deployment of distributed core network nodes, decentralizing critical NFs closer to the edge. Such an architecture reduces latency, enhances performance, and broadens the potential attack surface.

Moreover, the distributed nature of edge core networks increases the complexity of detecting and mitigating abnormal signaling attacks that can originate from compromised edge nodes and spread across interconnected network segments [5]. Meanwhile, various radio map (RM) technologies are introduced to address the functional degradation issues of different types of nodes at the user plane [6]. In a multiedge core network environment, where core functions are distributed across geographically diverse locations, the interactions between network segments become increasingly complex. Attackers can exploit these interactions to initiate attacks that propagate from one compromised edge core node to others, potentially affecting multiple network parts. Current defence systems are often built on isolated, point-based detection approaches. Even though each network segment may implement local defence mechanisms to detect abnormal signaling, these mechanisms can only sometimes coordinate across networks and share threat intelligence, leading to slower detection and limited effectiveness when dealing with attacks that span multiple nodes or networks [7]. A point-and-area integrated defence architecture is needed to ensure that individual detection models can be retrained or adapted to new threats at each network point, which can address the full scope of modern threats. In RingSFL [8], an adaptive model splitting mechanism is utilized to efficiently distribute the training process between the global model and multiple clients in parallel. This refined distributed architecture effectively reduces the latency associated with traditional approaches, resulting in significant performance improvements. Inspired by RingSFL, this article proposes a decentralized federated anomaly signaling protection architecture for the B5GC to address these challenges. This architecture is designed to complement existing single-point protections by enabling collaborative, cross-network detection and defence, thus offering a more robust and scalable solution to anomalous signaling attacks in B5G networks.

In our proposed architecture, we introduce the functional layered anomaly defense (FLAD) network as a point defence mechanism specifically tailored to the characteristics of anomalous signaling attacks within the PFCP framework. The decentralized defence architecture comprises multiple FLAD instances, each capable of updating attack parameters in real time, providing an integrated point-and-area defence mechanism. This design enhances anomaly detection efficiency and improves the ability to sense anomalies. The main contributions of this article are as follows.

1) We propose a hierarchical anomaly detection algorithm to distinguish between significant and nonsignificant anomalies in PFCP signaling attacks. This approach is novel in categorizing these attacks into distinct feature types. Extensive real-world testing validates the algorithm's effectiveness, improving detection accuracy and resilience against signaling attacks.
2) Our approach addresses the challenge of knowledge exchange among distributed core networks for anomaly detection in signaling protection. Traditional single-point

defence mechanisms often struggle with geographically dispersed, homogeneous attacks. By enabling knowledge sharing across networks, our method enhances the system's ability to detect anomalies more effectively.

3) This article presents a novel decentralized, federated anomaly detection architecture for collaborative B5G cloud-network environments. It combines point-and-area detection to ensure robust protection while preserving single-point defence capabilities. The architecture adapts to the distributed nature of B5G networks, enabling efficient, scalable anomaly detection while maintaining data privacy.

The remainder of this article is organized as follows. Section II discusses the threat modeling. Section III reviews the related work. Section IV presents the proposed architecture. Section V illustrates the evaluation results. Section VI concludes this article.

II. Threat Modeling

As outlined in the introduction, abnormal signaling attacks have attracted considerable attention due to their potential to disrupt communication networks and compromise system security. Anomalous signaling attacks, in their diverse forms and methods, significantly impact the operation of core networks. Attackers exploit these techniques to exhaust critical resources of the 5G core (5GC) through high-volume signaling requests or by sending malicious signals to confuse network protocols and traffic, obstructing normal request handling, leading to service disruptions or increased latency. Anomalous signaling attacks are often characterized by a high frequency of occurrence owing to their low operational complexity.

Within our threat model, we specifically focus on the N4 interface of the core network, primarily involving the PFCP and TCP protocol. Below is a summary of four PFCP-related network attacks identified in our investigation as follows.

1) *PFCP Session Establishment Flooding DoS Attack:* This attack aims to exhaust UPF resources by overwhelming it with legitimate session establishment requests and detection signal requests, hindering the 5G core network ability to successfully establish new protocol data unit (PDU) sessions. This exploit affects both the N4 and N6 interfaces. Connectivity can be restored by restarting the UE or by entering another gNbs coverage range. These actions will associate a new SEID with the UE PDU session, effectively neutralizing the attack.
2) *PFCP Session Deletion Flooding DoS Attack:* This attack aims to disconnect an UE from the DN. By sending numerous session deletion request messages, this attack depletes resources of the target UPF, impacting its normal operations. This exploit affects both the N4 and N6 interfaces. Connectivity can be restored by restarting the UE or by entering another gNbs coverage range. These actions will associate a new SEID with the UE's PDU session, effectively neutralizing the attack.
3) *PFCP Session Modification DoS Attack (DROP):* This attack invalidates packet processing rules for specific sessions, thereby disconnecting user equipment (UE) from the data network (DN). During rule updates, UPF deletes forwarding action rules (FAR) entries associated with tunnel endpoint identifiers (TEID) and base station IP addresses, resulting in subscriber GTP tunnels for downlink data transmission being severed, thus preventing DN access. However, sending data to the UPF can restore the GTPU tunnel. It is important to note that this exploit specifically targets client-DN PDU sessions, without disconnecting the UE from the 5G RAN or 5GC. Its impact is limited to the DN, and the attack occurs via the N4 interface, affecting the N6 interface.
4) *PFCP Session Modification DoS Attack (DUPL):* Exploiting the DUPL flag in the apply action field, this attack forces UPF to replicate session rules, creating multiple paths for a single source data. This instability in the N6 interface can lead to DN traffic replication and potentially be utilized for distributed Denial of Service (DDoS) attacks against the entire DN, consuming UPF resources.

These identified attacks underscore the critical need for robust anomaly detection mechanisms within B5GC to ensure resilience against malicious exploitation of PFCP vulnerabilities. In response to this research demand, some literature has proposed research directions and highlighted the importance of the research. However, effective anomaly signaling defense mechanisms specifically targeting the user plane N4 interface of B5GC remain lacking.

III. Related Works

After defining the threat model, it is essential to review and analyze the existing literature and related work pertinent to the threat model. In this section, we review related work on abnormal signaling attacks and detection, outlining relevant research on abnormal signaling attacks and the application of artificial intelligence and machine learning in anomaly signaling defence. This review provides insights into the current focus of existing research on abnormal signaling detection and lays the groundwork for future research.

Abnormal Signaling Attack: B5G adopts an evolutionary approach, integrating cloudification and service-oriented in a single stride. Service orientation leverages flexible application programming interfaces (APIs) and protocol interfaces, enhancing the openness, flexibility, and scalability of 5G networks [9]. However, it also inherits corresponding security threats. The B5GC employs protocols, such as PFCP, TCP, and HTTP2, widely used across the Internet, yet susceptible to security vulnerabilities, such as four types of attack signaling targeting the PFCP protocol [10] and the SIP IPSec disable attack [11] occurring under the SIP protocol exploit vulnerabilities in these protocols, allowing attackers to disrupt everyday network communications by forging legitimate signaling. Typical signaling attack methods can generally be categorized into three forms: primarily, attackers may disguise or conceal IP addresses, DNS names, and MAC addresses to send abnormal signaling to the network, deceiving B5GC NFs into carrying out malicious operations, such as falsifying location data or manipulating authentication information [12].

Second, attackers could exploit the most vulnerable radio areas in the B5GC network by deploying fake base stations to intercept communications and steal sensitive information exchanged between parties [13]. Finally, and most critically, attackers may attempt to incapacitate base stations or core network systems by flooding them with a high volume of anomalous or invalid signaling, causing the network to generate excessive signaling and rendering services unavailable through Denial of Service (DoS) attacks that exhaust critical resources on core network systems, such as the UPF [14].

Anomalous signaling attacks pose significant threats to B5GC, potentially causing irreparable harm. These vulnerabilities necessitate robust security measures to safeguard against such attacks, ensuring the integrity and reliability of B5G network operations amidst its transformative capabilities.

Abnormal Signaling Detection: Anomalous signaling detection is a technique aimed at monitoring and analyzing signaling traffic within communication networks to identify and detect abnormal or malicious signaling behaviors. The process typically involves data collection, feature extraction, anomaly detection, reporting, and response.

Most existing abnormal signaling detection methods in networks primarily rely on traditional machine learning techniques, utilizing mature mathematical theories and interpretability to infer and classify unknown signaling, such as support vector machines (SVMs) [15], decision trees (DTs) [16], and random forests (RFs) [17], all of which have notable shortcomings. SVM requires selecting appropriate kernel functions for different data samples and may not perform well with large-scale data. DT are prone to overfitting and have poor generalization performance. Although RF mitigate overfitting by constructing multiple DT, they also increase computational complexity. Another widely used method for abnormal signaling detection is learning and detecting based on the distinct distribution characteristics of different abnormal signaling categories, such as autoencoders [18] and generative adversarial networks (GANs) [19]. These methods rely on the assumptions that there are significant differences between abnormal signaling patterns and that abnormal samples are more difficult to reconstruct than normal ones. Other popular methods include constructing long short-term memory networks (LSTMs) [20] and convolutional neural networks (CNNs) [21]. However, these models often face challenges, such as insufficient feature extraction and inadequate data processing.

Related work needs to be done to effectively research the distribution characteristics of abnormal signaling in B5GC, and the B5GC PFCP dataset has not been appropriately utilized. A suitable anomalous signaling detection model must be tailored specifically for B5G networks. Addressing this research gap is critical for developing effective anomaly detection systems that protect B5G networks from emerging and sophisticated security threats. The inherent complexity and diverse applications associated with B5G networks present significant challenges that existing anomaly detection methodologies may need to address adequately. Consequently, there is a pressing need for a specialized model that can effectively address these unique challenges. This article aims to systematically investigate these challenges and propose potential solutions, thereby advancing the field of network security and contributing to the robustness of B5G network infrastructures.

IV. Methodology

A. Preliminary

In the context of anomaly detection tasks in the core network of mobile communication networks, identifying anomalies typically involves various types of features, primarily, including essential attributes, protocol attributes, and attack behaviors. Primary attributes mainly refer to fundamental information, such as destination address, source address, and session duration for each sample within the mobile communication network. These features exhibit minimal differences and high similarity during the recognition process. Protocol attributes describe specific behaviors of each protocol within the network and detailed data exchange models. Usually, datasets are collected for normal and abnormal data concerning specific protocols. Attack behaviors manifest as anomalous characteristics when attacks are launched against particular protocols. Most anomaly categories exhibit high variability in protocol attributes and attack behaviors, making them relatively easier to detect. Conversely, other anomaly categories exhibit high similarity in features, posing more incredible difficulty for detection. Simultaneously, while collecting abnormal information in mobile communication networks, datasets become highly dimensional due to accumulating numerous essential, protocol, and attack behavior attributes. This high dimensionality impedes the intuitive analysis of the distribution of these anomalies.

To address the challenges of imbalanced detection difficulty and high dimensionality in anomaly detection within mobile communication networks, the data is first subjected to dimensionality reduction analysis in this article. Principal components analysis (PCA) [22] is chosen as a commonly used method for multidimensional data analysis. The process of dimensionality reduction with PCA is as follows.

1) Assume the sample data x, the standardization process for sample data involves the following formula:
$$z = \frac{(x - u_x)}{\sigma_x} \tag{1}$$
where u_x denotes the mean of each column in the dataset, σ_x represents the standard deviation of each column's features, and z signifies the standardized data.
2) Using the formula
$$C = \frac{1}{n} z^T z \tag{2}$$
the covariance matrix C of data z is obtained. Performing eigenvalue decomposition on the covariance matrix C yields eigenvalues and eigenvectors. Select the top k rows to form the matrix P.
3) Using the formula $Y = P \times z$, obtain the k dimensional principal components Y. In the process of dimensionality reduction, the cumulative contribution rate of the principal components is calculated, and the top k principal

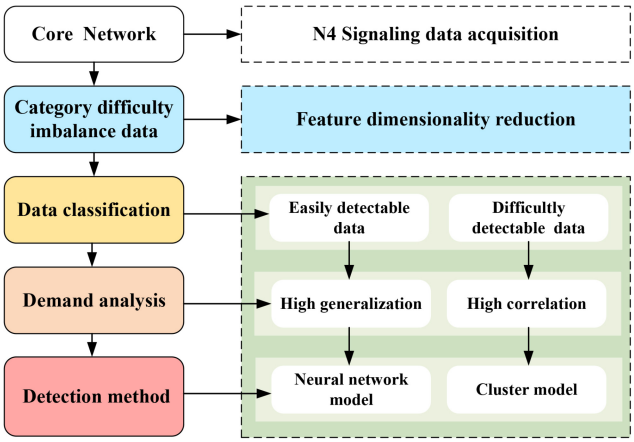

Fig. 2. Analysis and design diagram. In the core network, there are easily identifiable and difficult-to-identify data samples. The technical approach of this article is to handle these two types of anomalies with different detection difficulties separately.

components that reach the set threshold are selected as the basis for classification. The cumulative contribution rate of the top k principal components is expressed as

$$h = \frac{\sum 5_{i=1}^{k} \lambda_k}{\sum_{j=1}^{d} \lambda_j} \tag{3}$$

where λ_k as the eigenvalue of the kth principal component and d as the dimensionality of the original data.

After processing anomaly handling in network signaling attacks through cumulative contribution rate acquisition and PCA dimensionality reduction, it was observed that there are $N_1 + N_2$ types of anomalies inherent in the attacks. Among these, N_2 types of anomalies are often misclassified into the same category during classification due to the aggregation of different category features and blurred boundary delineations. This article defines these N_2 types of anomalies as nonsignificant. Conversely, N_1 other types of anomalies exhibit significant differences and clear boundary delineations compared to the nonsignificant anomalies grouped. Both traditional methods and neural networks show promising results in detecting these $N_1 + 1$ types of anomalies. Therefore, this article categorizes them as significant anomalies.

B. Functional Layered Anomaly Defense Network

Based on the analysis presented in preliminary, we identified two distinct types of anomalous signaling in the B5G core network: significant anomalies and nonsignificant anomalies. This differentiation underscores the need for a detection model capable of addressing the varying characteristics of anomalous signaling. In the case of significant anomalies, the model must be proficient in extracting and distinguishing critical features with high accuracy. Conversely, for nonsignificant anomalies, a model with heightened sensitivity to sample proximity is required to detect subtle deviations. Therefore, we propose a hierarchical detection model that can adapt to the distinct distribution patterns of these two types of anomalies, ensuring robust and accurate detection across the spectrum of anomalous signaling. Given the above analysis, we propose a detection mechanism tailored for both significant and nonsignificant anomaly features, as shown in Fig. 2. This mechanism is highly relevant for anomaly detection of N4 interface signaling in the B5G core network and provides a reliable research framework for future studies on B5G anomalous signaling detection.

Fig. 2 illustrates that different detection methods are required for anomalies with distinct feature distributions to enhance overall detection performance. For significant anomalies, deep feature extraction is necessary to uncover intricate patterns and neural networks can be employed as detection models, given the real-time requirements of anomaly detection, a lightweight neural network should be chosen. For nonsignificant anomalies, the features tend to be similar, and the decision boundaries are often ambiguous. Traditional machine learning algorithms are better suited to capture the data characteristics and achieve satisfactory classification results for such high-similarity, low-dimensional data. Appropriate machine learning algorithms can be selected for this purpose.

Specifically, there are $N_1 + 1$ types of significant anomaly samples and N_2 types of nonsignificant anomaly samples in the dataset. Given a set of training objects consisting of $m + p$ samples from both types of anomalies $X = \{x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{m+p}\}$ with $x_i \in \mathbb{R}^{1\times M}$, in which $\chi = \{x_1, x_2, \ldots, x_m\}$ is significant anomaly sample and $\mu = \{x_{m+1}, x_{m+2}, \ldots, x_{m+p}\}$ is nonsignificant anomaly sample. For a given raw input data x, after feature extraction and transformation by the upper-stage model $f_{\text{upper}}(\cdot)$, a new feature set is generated. This new feature set can be transformed into predicted probabilities for each category through a softmax function

$$\left(P_u{}^1, \ldots, P_u{}^{N_1}, P_u{}^{N_1+1}\right) = \text{Softmax}\left(f_{\text{upper}}(x)\right). \tag{4}$$

In the upper-stage model, it can obtain predicted probabilities for each category of significant anomalies and determine whether they are nonsignificant anomalies, when the probability value $P_u{}^{N_1+1} < \forall P_u^i (i = 1, 2 \cdots N_1)$, the loss function at the upper-stage is represented as

$$\text{Loss}_1 = \ell_1\left((f_{\text{upper}}(x)), y_x\right) \tag{5}$$

where ℓ_1 is a multiclass loss function and y_u^i denotes the target class of sample x^i.

When the probability value $P_u{}^{N_1+1} > \forall P_u^i (i = 1, 2 \cdots N_1)$, the raw input data x undergoes processing by the lower-stage model $f_{\text{lower}}(\cdot)$ to obtain probabilities for sample categories within nonsignificant anomalies

$$\left(P_1{}^1, P_1{}^2, \ldots, P_1{}^{N_2}\right) = \text{Softmax}\ \left(f_{\text{lower}}(x)\right). \tag{6}$$

Therefore, the loss function of the lower-stage model is

$$\text{Loss}_2 = \ell_2\left((f_{\text{lower}}(x)), y_x\right). \tag{7}$$

Finally, after aggregating and combining the output probabilities from the upper and lower stages, we obtain the final classification probabilities $N_1 + N_2$ for the categories $P = (P_u{}^1, \ldots, P_1{}^{N_1+N_2})$. It is important to note

$$\begin{cases} \left(P_u{}^1, P_u{}^2, \ldots, P_u{}^{N_1}\right) = 0\ P_u{}^{N_1+1} > \forall P_u^i (i = 1, \ldots, N_1) \\ \left(P_1{}^1, P_1{}^2, \ldots, P_1{}^{N_2}\right) = 0\ P_u{}^{N_1+1} < \forall P_u^i (i = 1, \ldots, N_1) \end{cases}. \tag{8}$$

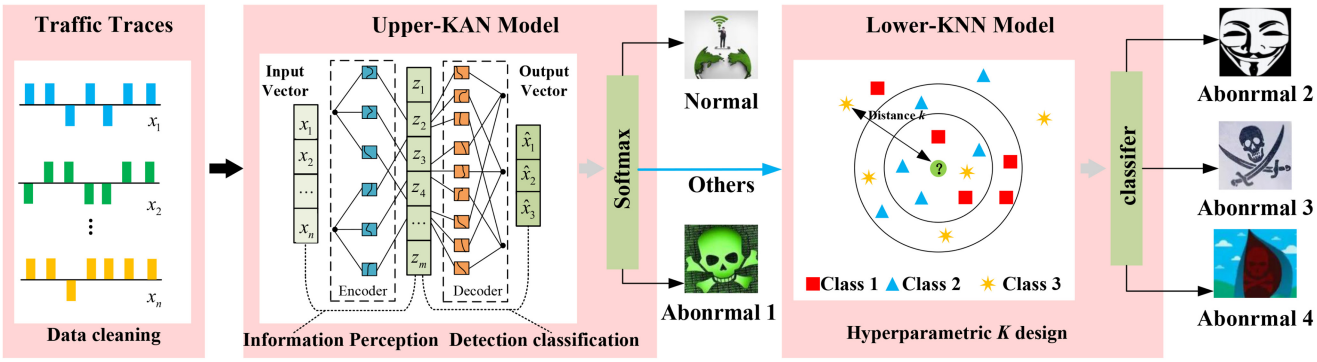

Fig. 3. Network architecture diagram of an FLAD. The data samples are directly separated for some significant anomalies by the upper-KAN model, while the remaining nonsignificant anomalies can be further distinguished by the lower-KNN model.

Based on the loss functions of the upper and lower stages, the overall loss function can be defined as follows:

$$\begin{aligned}\text{Loss}_{\text{total}} &= \text{Loss}_1 + \text{Loss}_2 \\ &= \ell_1\big((f_{\text{upper}}(x)), y_x\big) + \ell_2\big((f_{\text{lower}}(x)), y_x\big). \end{aligned} \tag{9}$$

Based on the detection mechanism presented in Fig. 2, our paper proposes an FLAD, as shown in Fig. 3. By separating anomalies with different detection difficulties and optimizing with relevant loss functions, the approach aims to improve the classification performance of the algorithm. This method effectively combines the detection of significant and nonsignificant anomalies, better handling complex abnormal scenarios in mobile communication network data. In the FLAD model described in this article, the upper model $f_{\text{upper}}(\cdot)$ is constructed using the Kolmogorov–Arnold network (KAN). In contrast, the lower model $f_{\text{lower}}(\cdot)$ adopts the traditional KNN detection method.

1) Significant Anomaly Classification Based on KAN Architecture: In the anomaly signaling data flow at the N4 interface of the B5GC core network, significant features exhibit apparent differences in the high-dimensional information between anomaly classes. The neural network model can efficiently learn and capture the deep features of the signaling data, thereby improving classification and detection accuracy. Given the real-time requirements of 5G anomaly signaling detection, a lightweight network architecture should be designed. Based on the above analysis, we select the KAN proposed by Liu [23] as the upper-layer structure of FLAD, as shown in Fig. 3. The KAN network eliminates the dependence on a linear weight matrix by using learnable functions instead of fixed activation functions. Additionally, KAN significantly reduces the complexity and the number of parameters required for precise modelling by focusing on optimizing these 1-D functions rather than the entire multivariate space.

By composing multiple univariate functions, KAN achieves an effect similar to that of a multilayer perceptron (MLP). While KAN features a fully connected structure similar to an MLP, it differs because KAN employs learnable activation functions at the edges rather than fixed activation functions at the nodes, as in MLPs. A KAN model with L layers can be represented as a nested composition of multiple KAN layers

$$\text{KAN}(\text{x}) = (\Phi_{L-1} \circ \Phi_{\text{L}-2} \circ \cdots \circ \Phi_1 \circ \Phi_0)(\text{x}) \tag{10}$$

where Φ_i denotes the ith layer of the entire KAN network. Each KAN layer has n_{in}-dimensional inputs and n_{out}-dimensional outputs. The function Φ consists of $n_{\text{in}} \times n_{\text{out}}$ learnable activation functions ϕ, and is given by the following formula:

$$\Phi = \{\phi_{q,p}\}, p = 1, 2, \ldots, n_{\text{in}}, q = 1, 2, .., n_{\text{out}}. \tag{11}$$

The computation results of the KAN model from layer k to layer $k+1$ can be represented in matrix form as follows:

$$x_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\Phi_l} x_l \tag{12}$$

where $\phi_{l,i,j}$ denotes the activation function between the ith neuron in the lth layer and the jth neuron in the subsequent layer.

In practice, a single-layer KAN model has limited expressive power due to its few parameters and difficulty in capturing complex patterns. We extended the single-layer KAN model to a multilayered structure, as illustrated in (10). In the upper-stage model, after feature extraction and transformation using the KAN model, the training is optimized using a cross-entropy loss function

$$\text{Loss} = -\sum_{i=1}^{N_1+1} {y_u}^i \log\big({P_u}^i\big) \tag{13}$$

where $N_1 + 1$ represents the number of significant anomaly categories, and y_u is the one-hot encoding of the significant anomaly category label. The position corresponding to the encoded label y is set to 1, with all other positions set to 0. P_u is the predicted probability from the upper stage.

By repeatedly optimizing the objective function, the model can better capture the features of complex data, there by improving classification accuracy and generalization ability.

2) Nonsignificant Anomaly Classification Based on KNN Algorithm: KAN has demonstrated high efficiency in detecting significant anomaly features. However, when handling nonsignificant anomalies, neighborhood relationships between similar anomalous classes result in nonlinear features and high similarity, which negatively impacts the classification performance of the KAN network. To address this issue, we propose a lower-level network model specifically designed to handle anomalies in neighborhood relationships. This model complements the KAN network to enhance the precision of network anomaly signaling detection.

Nonsignificant anomalies typically exhibit sparse and difficult-to-observe features and ambiguous boundaries, which neural network methods may find challenging to address effectively. Traditional machine learning algorithms, such as K-nearest neighbor (KNN) [24], DT, and RF, have shown promising results in nonlinear data with high similarity, making them suitable for handling these types of problems. Since similar anomalous samples exhibit neighborhood relationships, KNN is particularly well-suited for this task. Therefore, we construct a lower-level detection network based on the KNN algorithm to improve anomaly detection performance.

In the lower stage of the FLAD model shown in Fig. 3, the KNN model for anomaly classification involves three main steps. First, the distance between the test sample and each object in the training set is computed. Next, the K nearest training objects to the test sample are identified. Finally, the test sample is classified based on the predominant class among these KNNs. Distances between objects in the training set are generally measured using a standard metric to quantify the similarity between the test and training samples. This distance measurement considers the squared differences between each feature value of the test and training samples. Then, it takes the square root to combine these differences into a single distance value, effectively applying the Euclidean distance formula.

At the lower-level stage, optimizing the KNN algorithm focuses on selecting the value of k. During optimization, the value of k is adjusted to achieve an optimal classification result. For the original data $X \in R^{1\times M}$ with N types of nonsignificant anomalies, after applying (14), the probability values of each type of nonsignificant anomaly can be obtained. Among the most recent k values, the probability value C for category P can be expressed as

$$P = \frac{\sum_{i:x_i\in N_k(X)} D(X, X_i)\cdot \delta(y_i, C)}{\sum_{i:x_i\in N_k(X)} D(X, X_i)} \tag{14}$$

where $\delta(y_i, C)$ is an indicator function such that it takes the value of 1 if y_i equals category C, and 0 otherwise. $D(X, X_i)$ represents the distance metric between the data to be classified and the training sample X_i. $N_k(X)$ denotes the set of the nearest neighbors of all samples.

By using (14), the probability values $({P_1}^1, {P_1}^2, \ldots, {P_1}^{N_2})$ for each type of nonsignificant anomaly can be calculated at a given value of k. The k values are optimized using the cross-entropy loss function

$$\text{Loss} = -\sum_{k=1}^{N_2} {y_1}^k \log\left({P_1}^k\right) \tag{15}$$

where N_2 represents the number of nonsignificant anomaly categories, y_l is the one-hot encoding of the nonsignificant anomaly category label. The position corresponding to the encoded label y is set to 1, with all other positions set to 0. P_l is the predicted probability from the lower stage.

The optimal k value is determined by minimizing the loss function. In the lower-level stage of the two-layer detection algorithm illustrated in Fig. 3, the KNN algorithm continuously optimizes the k value, considering the number of nearest neighbor samples for classification, thereby avoiding potential misclassification that may arise from a global model.

3) Training and Testing Process of Functional Layered Anomaly Defense Algorithm: Algorithm 1 outlines the training and testing procedures of the functional layered anomaly defence algorithm. The training set is divided into datasets for significant and nonsignificant anomalies during the training phase. Each model layer is trained on its respective feature set, employing distributed training to allow each model to focus on the characteristics of each anomaly type, thus enhancing the model's ability to fit the corresponding anomaly type. These trained models are then loaded into the two-layer detection algorithm for application during the testing phase.

The test dataset is first classified by the upper-layer KAN network in the testing phase. This upper-layer model identifies significant anomalies within the test data and treats nonsignificant ones as a single group. Samples identified as significant anomalies are then passed to the lower-layer KNN model, further refining the classification of nonsignificant anomalies. The classification results from both layers are subsequently consolidated to identify different anomaly categories across the entire dataset comprehensively. By handling significant and nonsignificant anomalies in separate layers, the algorithm improves the accuracy of anomaly detection and classification, thereby significantly enhancing the overall performance of anomaly detection.

C. B5G Distributed Core Network Integrated Point-and-Area Decentralized Federated Anomaly Detection Architecture

Conventional single-point core network defence mechanisms often face challenges in promptly updating newly discovered vulnerabilities across geographically dispersed peer core networks. This limitation can lead to adjacent core networks remaining vulnerable to the same anomaly x, potentially resulting in successful attacks. To address this issue, we propose a decentralized, federated anomalous signaling protection architecture (B5GCASP) for core networks, as illustrated in Fig. 1. In this architecture, when a node detects a new anomaly, it updates its model weights to incorporate the characteristics of the newly identified anomaly. These optimized model weights are then shared with other nodes within the network, facilitating an integrated point-to-surface anomaly detection mechanism. This decentralized approach accelerates the enhancement of detection capabilities across multiple core networks and reduces the overall detection time by enabling more rapid dissemination of anomaly related information.

Algorithm 1 FLAD Algorithm

Input: Train data: $X = \chi \cup \mu$; Test data: T; Upper-layer Model $f_{upper}(\cdot)$: $M_u(\theta_u)$; Lower-layer Model $f_{lower}(\cdot)$: $M_l(\theta_l)$; FLAD model: $(M_u(\theta_u)|M_l(\theta_l))$; Number of training iterations: E;

Output: FLAD Model Test Results $R : \left\{R^1, R^2, \cdots, R^N\right\}$

// Model training process
Initialize the parameters θ_u of the upper-layer model M_u and the parameters θ_l of the lower-layer model M_l
for $i = 1, ..., E$ **do**
  X is divided into significant anomalies χ and nonsignificant anomalies μ
  Randomly sample M samples from the dataset χ to train the parameters θ_u of the model M_u: $\theta'_u \leftarrow \theta_u$
  Randomly sample M samples from the dataset μ to train the parameters θ'_l of the model M_l: $\theta'_l \leftarrow \theta_u$
end for
// Model testing process
Load the trained parameters θ'_u and θ'_l into the FLAD model: $(M_U(\theta'_u) \mid M_l(\theta'_l) \leftarrow (M_U(\theta_u) \mid M_l(\theta_l)$
Randomly sample N test data points $\{x_1, x_2, \ldots, x_n\}$ from T
for $i = 1, ..., n$ **do**
  Classification results through the upper layer of the FLAD model $f_{\text{upper}}(\cdot)$: $R^i_u \leftarrow f_{\text{upper}}(x_i)$
  if $\mathrm{R}^i_u \in \chi$ **then**
    Update the results: $R^i \leftarrow \mathrm{R}^i_u$
  else
    Classification results through the lower layer of the FLAD model $f_{\text{lower}}(\cdot)$: $R^i_l \leftarrow f_{\text{lower}}(x_i)$.
    Update the results: $R^i \leftarrow R^i_l$
  end if
end for
return R

Assume that there are initially n independent FLAD models within the core networks, with each core network having the same model structure and weights. Core network i can be represented as M_{θ_i}, where θ_i denotes the weight parameters of model i. If core network i is the first to be attacked by anomalous data x, it will undergo recovery and subsequently train on dataset X, which includes the anomaly x. Let $L(\theta_i)$ represent the loss of the new anomaly detection model during the parameter update process of core network i. Then, the loss $L(\theta_i)$ can be expressed as

$$L_{\theta_i} = -\left(\sum_{m=1}^{N_1+1} y_u^m \log\left(P_u{}^m\right) + \sum_{k=1}^{N_2} y_l^k \log\left(P_l{}^k\right)\right) \tag{16}$$

where $N_1 + 1$ and N_2 represent the number of significant and nonsignificant anomalies, P_u^m and P_1^k are the probability values of the mth significant sample and the kth nonsignificant sample. Respectively, y_u^{m} and y_1^k denote the true labels of the anomalous data. The updated model parameters can be expressed as θ'_i

$$\theta'_i = \theta_i + \nabla\theta_i \tag{17}$$

Algorithm 2 B5G Distributed Core Network Integrated Point-and-Area Decentralized Federated Anomaly Detection Architecture

Input: N core networks FLAD model $M = \{M_{\theta_1}, ...M_{\theta_n}\}$; Attacked core network i model M_{θ_i}; Train datasets $X = \{Train_1, ..., Train_n\}$; The dataset used for training after the recovery of core network $Train^i_{attacked}$; Number of iterations n_epochs;

Output: Each updated core network model after being attacked $\{M_{\theta'_1}, \ldots, M_{\theta'_n}\}$

Before being subjected to anomalous signaling attacks.
Initialize the parameters of each core network's FLAD model $\{\theta_1, ..., \theta_n\}$
for $i = 1, ..., n\_epochs$ **do**
  Send the model weights to neighboring core networks and update its own parameters: $\theta^{t+1} \leftarrow \theta^t$
  Each core network FLAD model calculates the loss using eq. (9):$L \leftarrow eq(9)$
end for
// After launching an anomaly signaling attack on the core network i and obtaining $Train^i_{attacked}$.
for $i = 1, ..., n\_epochs$ **do**
  Using $Train^i_{attacked}$, calculate the loss value using eq. (16) L_{θ_i}
  Update the parameters of core network model i: $\theta_i^{t+1} \leftarrow \theta_i^t + \nabla\theta_i$
end for
Send the trained model parameters a of core network i to other core networks $\theta'_j = \theta'_i (i \neq j)$
return Updated model $\{M_{\theta'_1}, \ldots, M_{\theta'_n}\}$

$$\nabla\theta_i \leftarrow L_{\theta_i}. \tag{18}$$

Based on the principles of decentralized federated learning model updates [25], when the model weights of a core network change, the updated weight information needs to be communicated to other core networks. The remaining $n - 1$ core networks need to update their model weights by incorporating the gradient information from core network i. Specifically, the core network model $M_{\theta'_i}$ needs to share its updated weights θ'_i with other core networks M_{θ_j} $(j \neq i)$. The following formula can represent this process

$$\theta'_j = \theta'_i (i \neq j). \tag{19}$$

Compared to traditional single-point core network defence mechanisms, the decentralized, federated learning architecture proposed in this article establishes an integrated point-to-surface anomaly signaling detection mechanism. This architecture leverages decentralized, federated learning, ensuring the efficient transmission of data weight characteristics between edge core networks without central cloud network involvement. In Algorithm 2 presents B5G distributed core network integrated point-and-area decentralized federated anomaly detection architecture (B5GCASP).

TABLE I
COMPARISON TABLE OF OVERALL ANOMALY CLASSIFICATION

Algorithm	Accuracy	FPR	F1
KNN	0.573	0.1273	0.558
Decision tree	0.647	0.0996	0.644
Random Forest	0.637	0.1017	0.636
SVM	0.588	0.6667	0.167
CNN	0.588	0.6667	0.167
MLP	0.588	0.6667	0.167
KAN	0.588	0.6667	0.167
Transformer	0.618	0.1146	0.571
FLAD	**0.688**	**0.0919**	**0.678**

V. EXPERIMENT

A. Experimental Dataset Description

To validate the effectiveness of the hierarchical detection algorithm for handling two specific types of imbalanced anomalies in the core network, we utilized the PFCP dataset as a crucial test basis. The PFCP dataset was collected by Amponis [10] through attacks on the PFCP protocol under a 5G core network, specifically for intrusion detection purposes. This dataset comprises anomalous and normal signaling for four types of PFCP protocol attacks. The training set consists of 2307 instances, while the test set includes 615 instances. To address the imbalance in class distribution, random sampling was employed during both training and testing phases to ensure balanced representation across each class.

In the experiment, after performing feature dimensionality reduction on the PFCP dataset, the cumulative contribution rate of representing the original dataset remained above 85% when reduced to two dimensions, as depicted in Fig. 4(a). This was determined by computing the ratio of the number of principal components to the cumulative variance contribution rate. In the low-dimensional plane, distinct distributions of the four types of anomalous signals and normal signals could be clearly observed. As shown in Fig. 4(b), Class 1 (NORMAL), Class 2 (Mal_EST), and Class 3 (including Mal_DEL, Mal_MOD1, and Mal_MOD2) exhibit clearly distinguishable boundaries, defining them as significant anomalies in this study. Within Class 3, the boundaries between Mal_DEL, Mal_MOD1, and Mal_MOD2 are blurred, characterized by fewer relevant data features and less straightforward observation, hence categorized as nonsignificant anomalies.

B. Experimental Setup and Baseline Models

This experiment utilized a computer equipped with a NVIDIA GeForce RTX 4060 graphics card and an Intel i7 12700K CPU. The experimental procedures were implemented in Python and involved constructing model algorithms based on the PyTorch framework. The PyTorch and CUDA versions utilize GPU acceleration for training, which are 2.1.2 and 11.8, respectively. The proposed FLAD algorithm employs four-layer KAN network structure, while the lower-stage KNN model utilizes a hyperparameter k set to 25. The learning rate for the FLAD model is configured at 0.01, with the Adam optimizer used for parameter optimization.

FLAD algorithm was compared with various methods. Among traditional machine learning approaches, we selected KNN, DT, RF, and SVM as comparison benchmarks. For neural network models, we chose MLP, CNN, KAN and the recently popular Transformer for comparison. The descriptions of these comparison methods are as follows.

1) KNN [26]: An unsupervised learning algorithm that utilizes distance-based metrics to classify a sample or predict its value by referencing its KNNs in the training set, widely used in classification and regression tasks.
2) DT [4]: A supervised learning algorithm that employs feature-based partitioning to recursively select the optimal features for splitting data into subsets and defining decision boundaries using specific thresholds, resulting in a tree-structured model.
3) RF [27]: An ensemble learning method that combines multiple independent DT, with the final output determined by majority voting for classification or averaging for regression tasks.
4) SVM [28]: A supervised learning algorithm that uses kernel functions to project data into a high-dimensional space, enabling the identification of an optimal linear decision boundary by maximizing the margin between classes for superior classification or regression performance.
5) MLP [29]: A deep learning architecture consisting of multiple fully connected layers, which transforms inputs into outputs through nonlinear activation functions and serves as a foundational model in neural network research.
6) CNN [30]: A deep learning architecture extensively employed in domains, such as image processing and speech recognition. It leverages convolutional layers to hierarchically extract local features from input data, followed by fully connected layers to perform classification tasks, thereby exhibiting substantial efficacy in anomaly detection applications.
7) KAN [23]: A lightweight neural network architecture derived from the Kolmogorov–Arnold representation theorem, capable of expressing high-dimensional functions as finite nested compositions of univariate continuous functions, with decision boundaries efficiently approximated using fully connected layers and B-splines.
8) Transformer [31]: A deep learning model utilizing self-attention mechanisms, characterized by an encoder–decoder architecture, and renowned for its advanced capabilities in global feature extraction and representation, extensively applied in tasks, such as text classification and anomaly detection.

C. Overall Anomaly Classification Comparison

First, we present a comprehensive overview highlighting the advantages of the proposed approach. Table I displays the performance of different models in terms of ACC, FPR, and FI in the comparative overall abnormal classification of the functionally layered network (FLAD). The FLAD model

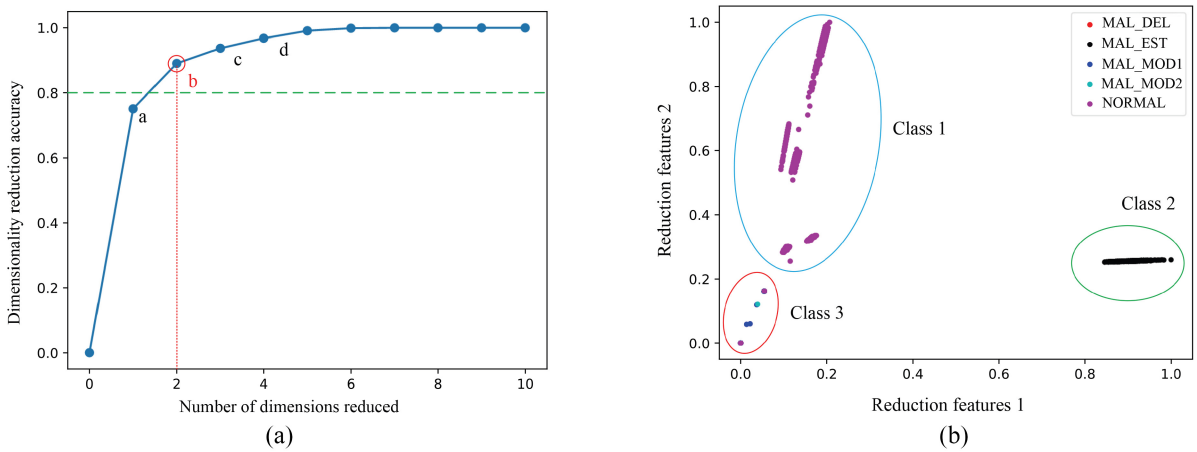

Fig. 4. Dimensionality reduction analysis: a) statistical calculation of the cumulative explained variance as the number of reduced dimensions increases and b) distribution of dataset sample types when the principal components are reduced to 2. (a) Pca dimensionality reduction information loss. (b) Dataset 2-D principal component analysis category relationship.

TABLE II
COMPARISON TABLE OF OVERALL ANOMALY CLASSIFICATION

Algorithm	MAL_DEL	MAL_EST	MAL_MOD1	MAL_MOD2	NORMAL	Accuracy
KNN	0.75	0.90	**0.40**	0.263	0.3375	0.573
Decision tree	0.636	0.92	0.313	0.35	0.934	0.647
Random Forest(RF)	0.536	0.91	0.325	0.386	0.928	0.637
SVM	0.97	0.94	0.00	0.00	0.918	0.586
CNN	0.98	0.93	0.00	0.00	0.932	0.588
MLP	0.95	0.96	0.00	0.00	0.923	0.587
KAN	**0.99**	0.98	0.00	0.00	0.929	0.585
Transformer	0.763	**1.00**	0.3875	0.00	0.937	0.618
FLAD	0.686	**1.00**	0.388	**0.425**	**0.938**	**0.688**

achieves approximately 4% higher accuracy and F1 score compared to the best results. Compared to traditional machine learning methods, such as KNN, DT, and SVM, it improves accuracy by around 4%. It also achieves an increase of approximately 10% in accuracy compared to neural network models, such as CNN, MLP, and KAN.

To highlight the advantages of using the hierarchical detection algorithm in handling two types of anomalies in the dataset, precise statistical analysis was conducted on the results of these models during the experimental process. Table II shows that the FLAD model performs comparably well in detecting significant anomalies MAL_EST and NORMAL, not inferior to models like DT and RF. However, FLAD does not exhibit superior performance in detecting MAL_DEL and MAL_MOD1 anomalies. This is because FLAD misclassifies both MAL_MOD1 and MAL_DEL as MAL_DEL anomalies, a phenomenon also observed in SVM, CNN, and MLP models. However, in terms of overall performance, FLAD demonstrates the most robust performance across different categories of anomalies, showing a balanced detection capability. Overall, it achieves approximately 4% improvement compared to the best-performing DT model. In order to better visualize the detection performance of FLAD across different anomaly categories, we have included Fig. 5, which provides a clear representation of each category's contribution to the total detection samples. This visualization serves to further substantiate the balanced detection capability of FLAD. Moreover, based on the prediction outcomes for each class of detection samples in Table II, confusion matrices for these models were plotted to showcase the experimental model's advantages.

Fig. 6 illustrates the confusion matrices for various models over the 5GC PFCP dataset, where C1–C5 refer to NORMAL, MAL_EST, MAL_DEL, MAL_MOD1, and Mal_MOD2, respectively. Overall, the FLAD model introduced in this research performs better than existing models when processing datasets containing significant and nonsignificant anomaly features. Specifically, the upper layer of the FLAD model employs the KAN network architecture, achieving significant performance enhancements of 10% and 60.05%, respectively, over traditional KNN methods for handling C1 and C2 data categories. The lower layer of the FLAD model leverages the KNN algorithm, resulting in notable improvements of 38.8% and 42.5% in processing C4 and C5 data categories compared to neural network models, such as MLP, CNN, and KAN. It is worth noting that although advanced neural network models, including MLP, CNN, SVM, and Transformer, report higher accuracy rates in the detection of C3 data, this elevated accuracy is primarily attributed to their misclassification of all nonsignificant anomalies as C3, thereby inflating the accuracy of C3 data detection. This phenomenon further elucidates why the FLAD model exhibits inferior performance to specific neural network models in predicting C3 data, as it prioritizes minimizing false positives to ensure the precision of anomaly detection. Furthermore, compared with traditional machine

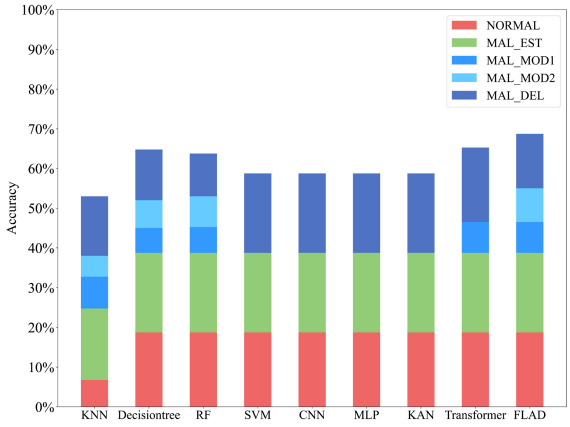

Fig. 5. Contribution ratio of accuracy for each category in the overall anomaly of the model.

learning approaches, such as RF, the FLAD model achieves performance improvements of 15%, 6.55%, and 3.9% in predicting C3–C5 data categories, respectively. Relative to baseline DT models, the FLAD model also demonstrates enhanced predictive capabilities, with performance increases of 5%, 7.7%, and 7.5% across these three categories.

D. Upper Significant Anomaly Classification, Lower Nonsignificant Anomaly Classification and Ablation Analysis

Fig. 7 illustrates the detection performance of significant and nonsignificant anomalies across the different functional layers of the network. Fig. 7(a) depicts the results of significant anomaly detection by various algorithms, highlighting the contribution proportions of significant anomalies in each category across different models. Both traditional machine learning and neural networks were trained and tested separately in this process. The results demonstrate that the KAN model achieves outstanding performance, with an accuracy of 98.7% enabled by multilevel feature extraction that captures subtle data features. Following closely are CNN and MLP, achieving classification accuracies of 98.1% and 98.3%, respectively. The KAN model excels in extracting features at different layers within each network, enabling it to capture anomaly patterns in the data from simple to complex and from local to global perspectives. This analysis indicates that the KAN model exhibits notable performance in classifying significant anomalies, particularly in the higher functional layers of the network.

Fig. 7(b) shows the results of different algorithms in detecting nonsignificant anomalies and illustrates the contribution proportions of nonsignificant anomalies in each category across different models. The KNN model performs well in nonsignificant anomaly classification, achieving an accuracy of 49.6%. It exhibits a distinct advantage in classifying nonsignificant anomalies, particularly suitable for handling nonlinear data and data with high similarity and insignificant features. Following are traditional machine learning models, such as DT and RF, with accuracies of 44.2% and 42.9%, respectively. In contrast, neural network models show insufficient feature significance in nonsignificant anomaly detection, resulting in poor generalization and low accuracy (33.3%). From Fig. 7(b), it is evident that nonsignificant anomalies exist in the detection of core network signaling attacks, posing additional challenges in detection.

As shown in Table III, the FLAD module employs the lightweight network structure KAN as an upper-layer component to learn and capture the features of salient anomalies. This approach yields results comparable to those of the Transformer model regarding overall performance. However, KAN exhibits a notable advantage in the number of covariates, requiring only 2.5% of the covariates needed by the Transformer model. Specifically, after fixing KNN as the lower module, the other modules selected for comparison in this study achieved the best results of 0.688, 0.919, and 0.678 in terms of accuracy, false positive rate (FPR), and F1 score, respectively, when either KAN or Transformer is used. However, the Transformer model has approximately 40 times more parameters than KAN. Compared to alternative candidate modules, such as DT, RF, and SVMs, KAN demonstrates a 3% improvement in accuracy and F1 score. Compared to neural networks like MLP and CNN, KAN shows a 2% improvement in accuracy and F1 score. The results of the comparison experiments on the selected different upper-layer modules indicate that both the KAN network and the Transformer model are effective in perceptual recognition when dealing with salient anomaly features. However, as shown in Table III, the Transformer model is better suited for handling large-scale, high-dimensional, and complex information, albeit with significantly larger model parameters. Given the real-time requirements of 5G anomaly signalling detection, this article ultimately selects the lightweight network structure KAN, which provides sufficient expressive power without increasing model complexity and exhibits superior performance outcomes with lower parameter counts.

Table IV shows the application of the FLAD module using the conventional KNN approach as an underlying component for addressing nonsignificant anomalies with high nonlinearity and similarity. Applying the KNN method as the lower module yields the optimal accuracy, FPR, and F1 score results. Compared to the DT with the highest overall performance, there is a 3% improvement in accuracy.

Additionally, in terms of the number of parameters, KNN is comparable to SVMs, RF, and DT, and it offers the advantage of a lower number of parameters and superior accuracy when utilizing traditional methods. The experimental results demonstrate that traditional machine learning algorithms can address the issue of nonsignificant anomalies arising from near-neighbor relationships between similar anomaly classes characterized by highly nonlinear and similar features. After comparing various modules, this article has validated the appropriateness of selecting KNN as the lower module. KNN demonstrates superior performance in handling similar anomaly samples with a certain degree of near-neighbor relationships compared to traditional methods, such as DT and RF.

E. B5G Distributed Core Network Integrated Point-and-Area Decentralized Federated Anomaly Detection Architecture

Fig. 8 compares the proposed decentralized B5GC anomaly signaling protection (B5GCASP) framework and traditional

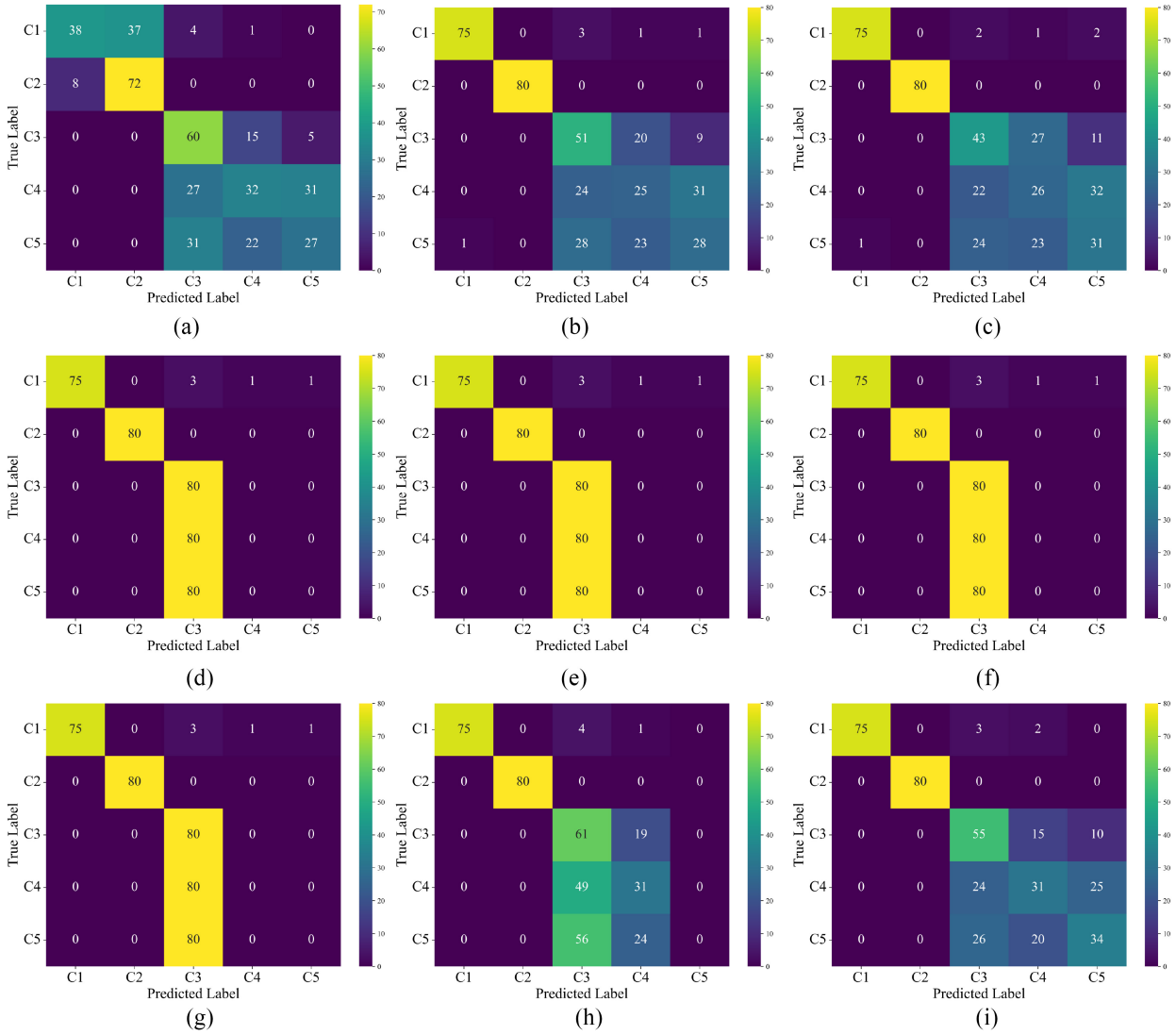

Fig. 6. Confusion matrices of various models on the 5GC PFCP dataset are shown, from which it can be observed that traditional machine learning methods can effectively separate the nonsignificant anomalies, such as MAL_DEL (C3), MAL_MOD1 (C4), and MAL_MOD2 (C5). Neural network models are able to distinguish between NORMAL (C1) and EST_EST (C2). Our FLAD model combines the different strengths of these two types of models, aiming to maximize the separation of significant and nonsignificant anomalies. (a) KNN confusion matrix. (b) DT confusion matrix. (c) RF confusion matrix. (d) SVM confusion matrix. (e) CNN confusion matrix. (f) MLP confusion matrix. (g) KAN confusion matrix. (h) Transformer confusion matrix. (i) FLAD confusion matrix.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS—UPPER LAYER ANALYSIS

Method								Parameter Count	Accuracy	FPR	F1
KNN	MLP	CNN	Transformer	SVM	Decision tree	Random Forest	KAN				
✓								✗	0.573	0.1273	0.558
✓	✓							151173	0.667	0.0998	0.645
✓		✓						14930	0.668	0.0981	0.654
✓			✓					595077	**0.688**	**0.0919**	**0.678**
✓				✓				✗	0.662	0.1011	0.634
✓					✓			✗	0.650	0.1034	0.627
✓						✓		✗	0.645	0.1052	0.619
✓							✓	**14720**	**0.688**	**0.0919**	**0.678**

single-point core network anomaly detection in terms of time and detection capability on both small-scale IID datasets and large-scale extended IID datasets. The large-scale extended IID dataset is larger and more complex, allowing us to evaluate the performance of our proposed method under different data complexities.

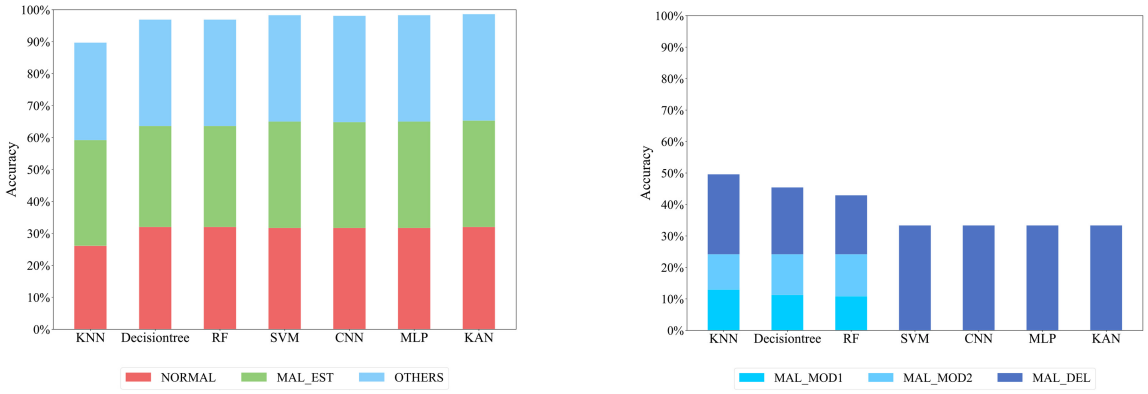

Fig. 7. Upper and lower layer algorithm analysis chart: a) in the upper layer, algorithms like KAN and CNN perform excellently in detecting significant anomalies and b) in the lower layer, the KNN algorithm performs well in detecting nonsignificant anomalies. (a) Accuracy contribution for each category of upper level. (b) Accuracy contribution for each category of Lower level.

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT METHODS—LOWER LAYER ANALYSIS

Method								Parameter Count	Accuracy	FPR	F1
KAN	MLP	CNN	Transformer	SVM	Decision tree	Random Forest	KNN				
✓								29440	0.588	0.1375	0.492
✓	✓							165893	0.558	0.1269	0.519
✓		✓						29650	0.588	0.1352	0.504
✓			✓					609797	0.618	0.1146	0.611
✓				✓				**14720**	0.558	0.1265	0.521
✓					✓			**14720**	0.650	0.0984	0.651
✓						✓		**14720**	0.653	0.0997	0.654
✓							✓	**14720**	**0.688**	**0.0919**	**0.678**

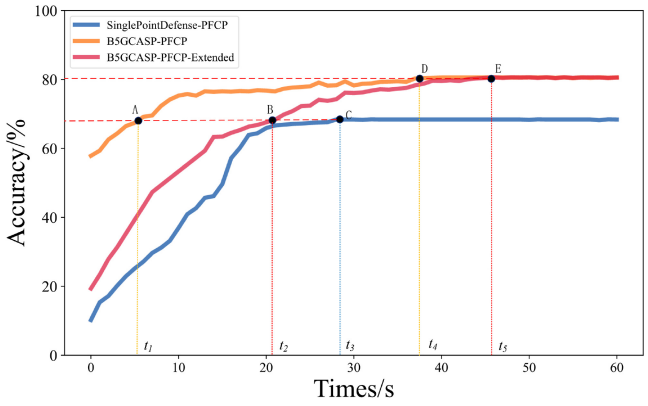

Fig. 8. Comparison of the B5GCASP with traditional single-point core network anomaly detection in terms of time and detection capability.

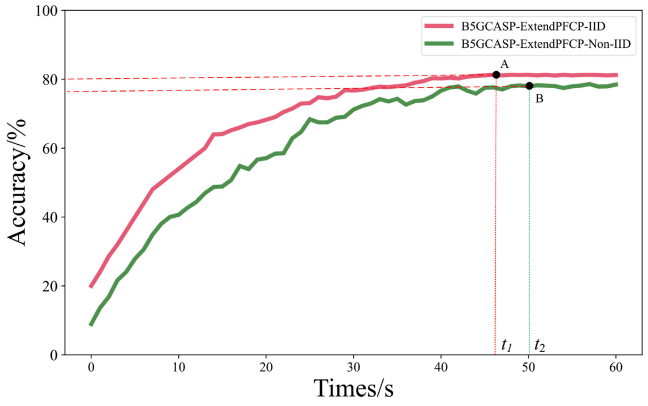

Fig. 9. Convergence comparison graph of the B5GCASP model on large-scale IID and non-IID datasets.

During the experiment, as the detection time progressed, the accuracy of all three curves showed an upward trend, eventually reaching a peak and gradually stabilizing. Notably, the detection accuracy of B5GCASP was consistently higher than that of single-point defence at various time points. Specifically, by leveraging decentralized federated learning principles and sharing the same anomaly weights among peer core networks, B5GCASP (represented by the orange and red curves) reached points A and B at times t_1 and t_2 with an accuracy of 68.8%. The traditional single-point defence mechanism (blue curve) also reached the same accuracy of 68.8% at time t_3 (point C). However, points A and B were reached earlier than point C, showing that B5GCASP can achieve the same detection performance as traditional single-point methods but in a shorter time. Furthermore, B5GCASP can detect anomalous signaling faster than single-point defence in both small-scale and large-scale extended datasets.

To validate the generalization capability of the proposed B5GCASP model, we conducted experiments using large-scale independent and identically distributed (IID) and non-IID datasets, as shown in Fig. 9. The experimental results show that under the IID data condition (red curve), where the data

from each client follows IID characteristic, the accuracy of B5GCASP rapidly increases and reaches a high-level stable state relatively early. In contrast, under the non-IID data distribution (green curve), B5GCASP requires more iterations to achieve a stable state, and the convergence process exhibits more significant fluctuations, resulting in a longer time to reach a stable accuracy. Specifically, the B5GCASP model achieves its highest accuracy of 80.0% at iteration t_1 on IID data and reaches a peak accuracy of 78.5% at iteration t_2 on non-IID data. Although there is a slight difference in final accuracy between the two data distributions (a reduction of only 1.5%), these results show that B5GCASP maintains strong generalization and robustness when handling data with different distributions, sustaining high-performance levels in complex real-world applications. Overall, the experimental results further show that B5GCASP performs well under ideal data distribution conditions and effectively maintains its performance in the face of data heterogeneity challenges, demonstrating excellent adaptability and generalization capabilities.

VI. Conclusion

To address the challenge of effectively detecting anomalous signaling attacks in the user plane of B5G networks, this article introduced a decentralized, federated abnormal signaling protection architecture based on functionally layered networks (B5GCASP). In addition, we proposed an FLAD model to identify both data anomalies efficiently. Extensive experiments on the PFCP dataset demonstrate that the FLAD model outperforms the existing detection methods in identifying data features under the N4 interface, showcasing its accuracy, FPR, and F-score effectiveness.

In the future, we plan to evaluate how real-time (time series) feature extraction impacts the state of the edge core network control plane. For large-scale datasets, our goal is to optimize the balance between accuracy and processing time while maintaining a unified level of point-to-surface protection.

References

[1] Y. Lu and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues," *J. Ind. Inf. Integr.*, vol. 19, Sep. 2020, Art. no. 100158.

[2] S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng, "Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 218–228, Apr. 2020.

[3] Z. Lv and N. Kumar, "Software defined solutions for sensors in 6G/IoE," *Comput. Commun.*, vol. 153, pp. 42–47, Mar. 2020.

[4] P. Radoglou-Grammatikis et al., "5GCIDS: An intrusion detection system for 5G core with ai and explainability mechanisms," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 353–358.

[5] V. Ziegler, H. Viswanathan, H. Flinck, M. Hoffmann, V. Räisänen, and K. Hätönen, "6G architecture to connect the worlds," *IEEE Access*, vol. 8, pp. 173508–173520, 2020.

[6] X. Wang et al., "RadioDiff: An effective generative diffusion model for sampling-free dynamic radio map construction," *IEEE Trans. Cogn. Commun. Netw.*, early access, Nov. 22, 2024, doi: 10.1109/TCCN.2024.3504489.

[7] X. Chen, W. Feng, Y. Chen, N. Ge, and Y. He, "Access-side DDoS defense for space-air-ground integrated 6G V2X networks," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 2847–2868, 2024.

[8] J. Shen et al., "RingSFL: An adaptive split federated learning toward taming client heterogeneity," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5462–5478, May 2024.

[9] L. Zhao et al., "A survey on open-source-defined wireless networks: Framework, key technology, and implementation," 2022, *arXiv:2209.01891*.

[10] G. Amponis et al., "Threatening the 5G core via PFCP DoS attacks: The case of blocking UAV communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2022, no. 1, p. 124, 2022.

[11] S. Park, S. Kwon, Y. Park, D. Kim, and I. You, "Session management for security systems in 5G Standalone network," *IEEE Access*, vol. 10, pp. 73421–73436, 2022.

[12] T. Fei and W. Wang, "LTE is vulnerable: Implementing identity spoofing and denial-of-service attacks in LTE networks," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[13] D. Rupprecht, K. Kohls, T. Holz, and C. Pöpper, "Call me maybe: Eavesdropping encrypted LTE calls with ReVoLTE," in *Proc. 29th USENIX Secur. Symp. (USENIX Secur.)*, 2020, pp. 73–88.

[14] Y. Shi and Y. E. Sagduyu, "Adversarial machine learning for flooding attacks on 5G radio access network slicing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2021, pp. 1–6.

[15] J. L. C. Bárcena et al., "Enabling federated learning of explainable ai models within beyond-5G/6G networks," *Comput. Commun.*, vol. 210, pp. 356–375, Oct. 2023.

[16] Y.-E. Kim, Y.-S. Kim, and H. Kim, "Effective feature selection methods to detect IoT DDoS attack in 5G core network," *Sensors*, vol. 22, no. 10, p. 3819, 2022.

[17] A. Rajak and R. Tripathi, "Analysis of network failure detection using machine learning in 5G core networks," in *Proc. Int. Conf. Innov. Comput. Commun.*, 2023, pp. 53–61.

[18] K. Sood, M. R. Nosouhi, D. D. N. Nguyen, F. Jiang, M. Chowdhury, and R. Doss, "Intrusion detection scheme with dimensionality reduction in next generation networks," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 965–979, 2023.

[19] S. S. Tripathy, S. Bebortta, C. Chakraborty, D. Senapati, S. K. Pani, and M. Guduri, "Leveraging resource-aware deep collaborative learning toward secure B5G-driven IoT-fog-based consumer electronic systems," *IEEE Trans. Consum. Electron.*, early access, Jun. 10, 2024, doi: 10.1109/TCE.2024.3411869.

[20] R. Pell, M. Shojafar, and S. Moschoyiannis, "LSTM-based anomaly detection of PFCP signaling attacks in 5G networks," *IEEE Consum. Electron. Mag.*, vol. 14, no. 1, pp. 56–64, Jan. 2025.

[21] Y. Liu, J. Kang, Y. Li, and B. Ji, "A network intrusion detection method based on CNN and CBAM," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.

[22] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, 1993.

[23] Z. Liu et al., "KAN: Kolmogorov-Arnold networks," 2024, *arXiv:2404.19756*.

[24] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *Proc. OTM Confed. Int. Conf. Move Meaningful Internet Syst. (CoopIS, DOA, ODBASE)*, 2003, pp. 986–996.

[25] Q. Chen, Z. Wang, W. Zhang, and X. Lin, "PPT: A privacy-preserving global model training protocol for federated learning in P2P networks," *Comput. Secur.*, vol. 124, Jan. 2023, Art. no. 102966.

[26] S. Ying et al., "An improved KNN-based efficient log anomaly detection method with automatically labeled samples," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 3, pp. 1–22, 2021.

[27] C. Liu, Z. Gu, and J. Wang, "A hybrid intrusion detection system based on scalable K-means+ random forest and deep learning," *IEEE Access*, vol. 9, pp. 75729–75740, 2021.

[28] L. Kong, G. Huang, and K. Wu, "Identification of abnormal network traffic using support vector machine," in *Proc. 18th Int. Conf. Parallel Distrib. Comput., Appl. Technol. (PDCAT)*, 2017, pp. 288–292.

[29] M. Wang, Y. Lu, and J. Qin, "A dynamic MLP-based DDoS attack detection method using feature selection and feedback," *Comput. Secur.*, vol. 88, Jan. 2020, Art. no. 101645.

[30] H. Liu and H. Wang, "Real-time anomaly detection of network traffic based on CNN," *Symmetry*, vol. 15, no. 6, p. 1205, 2023.

[31] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

Cong Li received the B.Eng. degree in electronic information technology from Xidian University, Xi'an, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Cyber Engineering.

His research interests are in the areas of next-generation mobile communication and privacy preservation.

Xingxing Liao received the Ph.D. degree in circuits and systems from the University of Chinese Academy of Sciences, Beijing, China, in 2017.

His research interests are in the areas of B5G/6G security and network resilience.

Zilong Wang (Member, IEEE) received the B.S. degree in mathematics from Nankai University, Tianjin, China, in 2005, and the Ph.D. degree in mathematics from Peking University, Beijing, China, in 2010.

From 2008 to 2009, he was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Since 2010, he has been with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China, where he is currently a Professor with the School of Cyber Engineering. His research interests are in the areas of sequence design, cryptography, and information security.

Prof. Wang received a Postdoctoral Fellowship from the University of Waterloo in 2012.

Xinsheng Ji received the B.S. degree from Fudan University, Shanghai, China, in 1991, and the M.S. degree from National Digital Switching System Engineering and Technological Research Center (NDSC), Zhengzhou, China, in 1994.

He has been a Professor with NDSC since 2005. His current research interests include next-generation mobile communication and cyber space security.

Guoqiang Mao (Fellow, IEEE) received the Ph.D. degree in telecommunications engineering from Edith Cowan University, Joondalup, WA, Australia, in 2002.

He is a Distinguished Professor and the Dean of the Research Institute of Smart Transportation, Xidian University, Xi'an, China. He has published more than 200 papers, which have been cited more than 10 000 times. His research interests include intelligent transport systems, Internet of Things, and wireless-localization techniques.