IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 December 2023, accepted 5 February 2024, date of publication 14 February 2024, date of current version 21 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365425

RESEARCH ARTICLE

Detecting IP DDoS Attacks Using 3GPP Radio Protocols

LOAY ABDELRAZEK[1], RAMIN FULADI[2], JÁNOS KÖVÉR[1], LEYLI KARAÇAY[2], AND UTKU GÜLEN[2]

[1]Ericsson, 16480 Stockholm, Sweden
[2]Ericsson Research, 34467 Istanbul, Turkey

Corresponding author: Ramin Fuladi (ramin.fuladi@ericsson.com)

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) through the 1515 Frontier Research and Development Laboratories Support Program under Project 5169902, and in part by the Hexa-X II Project through the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe Research and Innovation Programme under Grant 101095759.

ABSTRACT Contemporary mobile networks, offering advanced services such as highly dependable and fast communication (URLLC) and extensive device-to-device connectivity (mMTC), are paving the way for the upcoming 6G era. These networks are expanding their capabilities beyond traditional voice and short messaging services, enabling diverse devices to connect to the cellular network. However, with this increased connectivity comes a heightened vulnerability at the radio interface, which is the primary access medium for mobile network communication. This research work focuses on safeguarding the availability of the radio interface in the face of emerging threats. Threats to radio interface availability can originate either directly from exploiting the 3GPP radio protocol stack within base stations or indirectly through the IP protocol stack carried over the user plane. In particular, this research paper delves into user plane DDoS attacks leveraging the IP protocol stack to generate excessive traffic. It introduces a novel detection method situated within the Radio Access Network (RAN). This method analyzes the patterns of radio protocols and their functionalities to identify user plane DDoS attacks initiated from User Equipment (UEs). Importantly, the method does not rely on directly inspecting user plane packets like IP packets but rather leverages the characteristics of 3GPP radio protocols (e.g., MAC, RLC, PDCP) to detect IP DDoS attacks closer to their origin. This early detection capability helps prevent DDoS traffic from propagating to the transport network. The implications of this research extend beyond the current generation of networks, as it lays a foundation for enhancing security in the forthcoming 6G networks. As 6G aims to deliver even more advanced services and connectivity across a diverse array of devices, the robust security measures proposed in this work will be instrumental in ensuring the reliability and availability of these cutting-edge networks. The analysis employed in this paper showcase the performance with accuracy of 98.9% for DDoS attack detection.

INDEX TERMS Cellular botnets, DDoS, machine learning, radio access network, security, XAI, 4G, 5G, 6G.

I. INTRODUCTION

The emergence of 5G and as well as the anticipated future generations like 6G (and beyond, XG), the most recent developments in mobile network technologies, as well as their associated capabilities like Massive Machine Type Communications (mMTC) and Machine-to-Machine (M2M) connectivity, have significantly changed the role of mobile networks [1], [2], [3]. They now serve as indispensable providers of connectivity for various industries. This transition has also introduced the concept of the Internet of Everything (IoE), wherein intelligent devices within the network seamlessly connect and collaborate with one another through the infrastructure facilitated by advanced mobile technologies [4], [5].

However, despite the numerous advantages presented by these technologies, they also bring forth potential vulnerabilities that can be exploited by malicious actors. These smart

The associate editor coordinating the review of this manuscript and approving it for publication was Chen Chen.

devices often lack reliable security safeguards, making them susceptible to exploitation and falling prey to malevolent purposes. An extensive array of vulnerable devices can be targeted by various threats, and one major worry with respect to network availability is the Distributed Denial of Service (DDoS) attack. This kind of attack involves gathering a large group of compromised devices, or "zombies," to besiege a target, which might be a server or the network itself. Through this attack, attackers drain the target's resources-such as bandwidth, memory, and computational power-making it unable to meet the needs of legitimate users [6], [7], [8].

Within the realm of mobile networks, various approaches have been suggested for identifying DDoS attacks [9], [10], [11], [12], [13]; nevertheless, the majority of these strategies focus on scrutinizing packets at the IP level. These examinations occur within the core network, where the malicious packets traverse the Radio Access Network (RAN) and are in close proximity to the targeted destination. Despite the merits of these proposed techniques for detecting DDoS attacks, the issue lies in their potential latency and the timing of detection. Given the pivotal role of mobile networks in meeting the extensive demands from users at high rates, this delay can result in service disruptions for legitimate users.

In light of the fact that the initial point of interaction in attacks on the network is often the radio interface, devising a method for DDoS detection at this level is reasonable and critical. Doing so makes it feasible to implement mitigation and countermeasures before malicious packets infiltrate critical junctures within the network. However, formulating a detection algorithm at this stage necessitates identifying and establishing new sets of features and methodologies. These novel approaches are indispensable for the creation of a new solution tailored to DDoS detection within this specific context. Such advances have the potential to significantly enhance the network's ability to fend off DDoS attacks more proactively, thereby ensuring uninterrupted service to legitimate users.

In this paper, we introduce a novel approach for the early detection of DDoS attacks, which operates directly within RAN, as opposed to the conventional method of inspecting packets within the core network. The significance of this shift arises from the fact that the 3GPP radio protocols are tasked with encapsulating and transmitting the IP packets dispatched by devices. Consequently, any alterations in the behavior of sending IP packets or modifications in the pattern of IP traffic would be manifested within the radio protocol stack. To illustrate this concept, consider the scenario of IP-based volumetric attacks. In such instances, changes in behavior can be observable within the radio stack. For instance, a surge in the number of packets sent per second would consequently increase the number of radio Packet Data Units (PDUs) sent per second. Similarly, variations in the size of the radio protocols could signify anomalous behavior. Another instance pertains to the repercussions of IP source address spoofing, which could adversely impact the ratio of radio uplink and downlink PDUs. By recognizing these patterns, we can identify instances where user-plane (i.e., IP traffic) volumetric attacks generate distinct radio protocol traffic patterns. This early detection capability, situated within the RAN, empowers us to promptly recognize these attacks before they infiltrate deeper into the network. Furthermore, this approach enables the application of localized policies for mitigation, thereby swiftly safeguarding the network against potential disruptions caused by DDoS attacks. In essence, our proposed method leverages the inherent characteristics of the radio protocol stack to proactively detect IP-based volumetric attacks at their point of entry into the network. This proactive approach not only enhances the network's resilience but also expedites the implementation of appropriate countermeasures, ensuring minimal impact on legitimate users' service experience.

In the existing body of literature, Machine Learning (ML) and Artificial Intelligence (AI) have become commonplace tools for identifying DDoS attacks. Furthermore, by advancing the infrastructure of mobile networks, especially in the context of 5G and 6G, it is becoming increasingly evident that an AI-centric approach will constitute the backbone of these cutting-edge mobile network systems. Therefore, in this research paper, we also delve into utilizing attributes extracted from the RAN protocol in conjunction with ML models. However, a significant challenge arises from the fact that these distinct attributes are acquired at different time points. This temporal misalignment prevents their direct utilization as features for ML models. To address this challenge, we introduce a designated time interval, denoted as t, and within this interval, we derive statistical features from the original attributes of the RAN protocol. This entails computing various statistical measures [14], such as the cumulative sum, mean value, standard deviation, skewness, kurtosis, second L-moment, L-skewness, L-kurtosis, and entropy for each attribute within the specified time-interval t. This method ensures that we generate a consistent set of features with synchronized temporal occurrences, making them suitable for employment by ML models. Moreover, our investigation extends to assessing the significance of these features. We execute several methods for feature analysis, including Analysis of Variance (ANOVA) [15], Mutual Information criterion [16], Recursive Feature Elimination [17], and SHAP (SHapley Additive exPlanations) analysis [18]. Through these techniques, we gain insights into the importance of individual features in contributing to the effectiveness of the ML models.

Our main contributions include:

- We analyzed the characteristics and behavior of both the user plane and control plane protocol stacks in comparison with the IP traffic generated from a malicious device.
- We identified the most important features in the 3GPP radio stack protocols (e.g., Medium Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol(PDCP)) that are sent on the radio

interface to detect malicious traffic within the radio network domain.

- We computed a range of statistical measures for each RAN protocol feature. This approach served a dual purpose: It not only tackled the temporal misalignment issue that existed among the features but also led to the creation of features that encapsulate information concerning the probability distribution of the underlying attributes.
- We developed a detection mechanism that enables us to detect user plane DDoS attacks (e.g., TCP SYN and UDP Flood) attacks with high accuracy before reaching the core network.
- We employed various techniques for feature engineering, seeking to identify the optimal feature set for the purpose of detecting DDoS attacks.

II. RELATED WORK

To our best knowledge, published DDoS detection techniques in conjunction with RAN revolve around protection against various types of DoS attacks against the eNB/gNB and the provided wireless service (e.g., 3GPP signaling storm, etc.) [9], [10], [11]. On the other hand, there are many existing solutions in the industry to detect different types of IP-based DDoS attacks (e.g., TCP SYN flood, UDP Flood, etc..) by looking at IP packets or IP flow information. These solutions are typically deployed at the core, where they only analyze incoming traffic to the RAN (i.e., traffic destined in the uplink direction from the UEs towards the network). If User Equipments (UEs) linked to a 3GPP network transform into origins of IP-based DDoS attacks, the existing detection approaches either fail to identify the attacks taking place in the user plane (because they disregard the user plane traffic or overlook outbound data) or accomplish this at a considerable distance from the origin, consequently resulting in only delayed detection. Additionally, if the analysis of traffic occurs multiple hops away from the origin in an attempt to pinpoint a malevolent traffic source, certain valuable details like exact timing might be altered, thereby diminishing the precision of detection

Software-Defined Networking (SDN) and Network Function Virtualization (NFV) have emerged as the fundamental components of both the 5G and 5GB frameworks, spanning both the RAN and the core network. Given the versatility and adaptability inherent in these advancements, a significant portion of the DDoS attack detection methods found in the literature are specifically tailored for these innovative technologies. In the study by Perez et al. [12], an architecture oriented towards SDN and NFV was proposed, encompassing components aimed at detecting and mitigating botnets within 5G networks. Their approach to detection relied on the implementation of two control loops, each functioning at a distinct level of abstraction to cater to the substantial number of anticipated 5G subscribers' User Equipments (UEs). The initial control loop encompassed a lightweight, high-level detection mechanism that rapidly scrutinized network flows to pinpoint potentially suspicious bot activities. Upon identifying potential bots, the second control loop triggered a more comprehensive, low-level Deep Packet Inspection (DPI) to validate the presence of a botnet. However, the authors did not provide any evaluation outcomes for their proposed architecture in their paper. In the work by Sridharan et al. [9], a framework was created to passively identify application layer attacks by analyzing encrypted wireless traffic through link-layer features. Experimental trials involving diverse IoT devices revealed that the framework successfully recognized 96.2% of IP camera attacks, achieving a 97% accuracy in their classification, and accurately pinpointed Mirai bot infections with a precision of 96.1%. Furthermore, it exhibited a 98.3% accuracy in detecting DDoS attacks on IoT devices initiated by a Mirai botnet. Notably, the framework achieved anomaly detection with a remarkable accuracy exceeding 98.8% for a TP-Link bulb utilizing WiFi, and it attained a detection accuracy of 99% for association flooding attacks on a Zigbee controller. In a separate study detailed in [10], authors introduced an independent security system designed to counteract UDP flooding DDoS attacks, providing safeguarding measures for 5G multi-tenant infrastructures. This concept was demonstrated using a container-based emulator. The researchers executed practical tests on their proposed system within the emulator environment. Their approach involved identifying harmful data flows through a security monitoring agent and subsequently mitigating these detrimental flows via a network flow control agent.

Ravi and Shalinie [11] introduced a technique termed Learning-Driven Detection Mitigation (LEDEM) aimed at identifying and mitigating DDoS attacks on IoT servers to enhance security within cloud and SDN environments. LEDEM underwent assessment within a test-bed and an emulated topology, with outcomes being juxtaposed against those of contemporary solutions. They attained an enhanced accuracy rate of 96.28% in the detection of DDoS attacks through the application of the Semi-supervised Deep Extreme Learning Machine (SDELM) model. Furthermore, they proposed a mitigation algorithm classified as an approximation algorithm, which they demonstrated to be a 2-approximation algorithm in their research. They carried out rigorous testing of their mechanism within their test-bed to validate its effective performance within an actual hardware network. They executed a sequence of experiments utilizing the benchmark UNB-ISCX dataset and contrasted the findings with those of state-of-the-art solutions.

In [13], the authors put forward an anomaly-based IDS that was capable of identifying and counteracting emerging DDoS attacks promptly within IoT networks. They showcased the efficacy of the IDS in detecting and mitigating surreptitious DDoS attacks, even when originating from sources with exceedingly small attack sizes. This was achieved through numerical assessments and trials conducted within a testbed. The authors introduced a technique for detecting and addressing these forms of attacks and examined the associated time and space complexities. They demonstrated the asymptotic optimality of the proposed detection mechanism in a minimax context, as the volume of training data increased. Additionally, they devised a solution to accommodate dynamic scenarios wherein the count of devices within the network fluctuated. An evaluation of the system's performance was executed through the utilization of a test-bed implementation, the N-BaIoT dataset, and simulations.

In the study conducted by Harada et al. [19], an attack suppression mechanism is introduced, aiming to decrease the unnecessary discarding of legitimate traffic within the IoT backhaul when its capacity is surpassed. This approach swiftly regulates frame priorities without introducing novel routes to accomplish the task. The system's Network Controller (NWC) outpaces conventional techniques in gauging attack traffic and adapts frame priorities to assign the lowest priority exclusively to suspicious data flows. Furthermore, the switches obstruct attack traffic as identified by a DDoS protector positioned ahead of the IoT server, thus preventing the IoT backhaul from exceeding its maximum capacity. The system rapidly estimates potentially malicious traffic based solely on its rate prior to the detection of a DDoS attack. Additionally, it quells questionable traffic by manipulating frame priorities until the DDoS protector recognizes the presence of attack traffic, obviating the necessity for extra switches and fibers. This work mainly focused on the rate of the traffic coming from the IoT devices, which we believe that relying on rate of traffic should not be the only feature to detect an attack, we believe that we have several 3GPP radio features that could outperform features related only to rate of traffic.

Most of the related works in the literature are related to detection of DDoS attacks in the core, and there are quite a few works on early detection of DDoS attacks at RAN. The closest work to our current proposed method can be found in [20]. They presented a technique for early detection of DoS attacks that harnessed the novel OpenRAN framework to amass data from the wireless communication interface, allowing for the early identification of attacks before their propagation through the network. They developed a nearly real-time RAN Intelligent Controller (RIC) compatible with open-source base stations, such as srsRAN, enabling the collection of measurements. Their approach capitalized on attributes from the physical and Medium Access Control (MAC) layers to spot diverse forms of DoS attacks. Employing various machine learning (ML) algorithms, they performed real-time analysis of data traffic, effectively classifying different DoS attacks. Empirical findings from their experimentation showcased their method's ability to accurately distinguish between legitimate and malicious traffic, yielding an impressive accuracy rate of 95%. These results were exhibited within a practical testbed environment.

However, their work did not delve into an assessment of the chosen features' benefits. For instance, relying on bitrates and the signal-to-noise ratio (SINR) could lead to false positives, as these metrics might indicate poor channel conditions unrelated to an attack. Furthermore, they did not explore how the characteristics of traffic impacted the chosen features. Not all features or combinations thereof might effectively work with all types of traffic for attack detection, especially in the case of web traffic. Web traffic heavily relies on human interactions with webpages, introducing intricacies in behavior. Additionally, selecting web traffic as a benchmark might be somewhat misleading. DDoS attacks primarily manifest as uplink traffic, and web traffic consists of more downlink traffic than uplink traffic.

Lastly, we contend that the features chosen in our approach are applicable to both monolithic and disaggregated deployment scenarios.

III. BACKGROUND

This section presents the basic concepts for the LTE radio protocol stacks used in this work. This is followed by brief description of cellular botnets and the type of attacks that is in the scope of this paper.

A. 3GPP RADIO PROTOCOLS

Mobile networks provide wireless communication service worldwide. The architecture and protocols of mobile networks are standardized by 3GPP. There are different generations of mobile network, i.e., 2G, 3G, 4G, and 5G, the latest generation being the 5G [21]. On a high-level, there are three distinct components in a mobile network, i.e., user equipment (UE), radio access network (RAN), and core network (CN). The UE is a mobile device used by users to wirelessly access the network. The RAN is responsible for providing wireless radio communication to the UE and connecting the UE to the CN. The CN is responsible for authenticating the UE, packet routing, and handling mobility of the UE, among other responsibilities. We will focus mainly on the RAN part in this paper. In 5G, the RAN is known as a new generation RAN (NG-RAN). An NG-RAN is either a gNB providing New Radio (NR) user plane and control plane protocol terminations towards the UE or an ng-eNB providing Evolved Universal Terrestrial Radio Access (E-UTRA) user plane and control plane protocol terminations towards the UE. The gNB hosts many functions including, radio resource management, routing of user plane data towards the core network, etc.

The Radio interface in a mobile network is called the Uu interface, regardless of the mobile generation it carries two types of Radio protocols stack: (1) Control Plane (CP), and (2) User Plane (UP). The user plane protocols, which are terminated in the gNB, are the protocols carrying user data through the Access Stratum (AS) on the Uu interface between the UE and the gNB. On the other hand, the control plane protocols control the user plane sessions and the connection between the UE and the network, for example, initial access, requesting services, mobility, etc.

As illustrated in Figure 1, the UP protocol stack is composed of different layers each providing different functionality. In NG-RAN, the Service Data Adaptation Protocol (SDAP), the Medium Access Control (MAC), Radio Link Control (RLC), and Packet Data Convergence Protocol

(PDCP) are known as layer 2 protocols. The MAC layer performs multiplexing, demultiplexing, and scheduling. The RLC sublayer handles sequence numbering, segmentation, and re-segmentation. PDCP handles header compression, in-sequence delivery, ciphering and integrity protection, and it is also responsible for transferring user plane data. The SDAP sublayer is responsible for quality of service (QoS) flow handling.

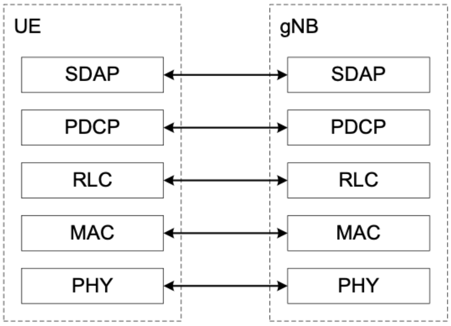

FIGURE 1. User plane protocol stack.

As illustrated in Figure 2, control plane stack protocols are handled and terminated differently, although they might be in the same sublayers as in user plane stack, but they have different functionalities. In the CP stack the PDCP, RLC and MAC sublayers are terminated in gNB. On top of the previously mentioned protocols, Radio Resource Control (RRC), known as layer 3, is a control plane protocol that is terminated in the gNB. RRC is responsible on establishment, maintenance and release of the RRC connection between the UE and NG-RAN, security functions, QoS, paging, broadcast of system information, UE measurement reports and other functions. Additionally, there is the Non-Access Stratum Protocol (NAS) that is terminated in the Authentication Management Function (AMF) in the 5G CN, and responsible for subscriber authentication, mobility management, security control among other functions as well.

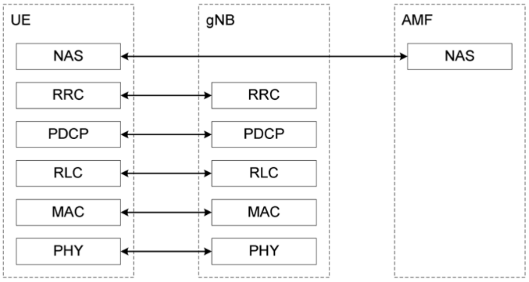

FIGURE 2. Control plane protocol stack.

B. CELLULAR BOTNETS

DDoS attacks are becoming more widespread [22], [23], relying on compromised hosts (botnet) connected to the Internet. It is expected that devices connected to the internet through 3GPP networks may become part of a botnet and as such, source of DDoS attacks. User plane DDoS attack leverage network protocols based on protocols (e.g., TCP, UDP) that are defined by the Internet Engineering Task Force (IETF) [24]. There are many flavors of DDoS attacks, possible main categories are network-layer attacks and application layer attacks. The network-layer attacks rely on the Open Systems Interconection (OSI) model layers, for example the network and transport layer protocols to conduct the attack, two prominent examples are the TCP SYN flood and the UDP flood volumetric attacks. The UDP protocol is used in time-sensitive communications, for example voice, video, and gaming traffic. These are some of the widely used traffic types in today's networks. Additionally, UDP is considered as a lightweight protocol (in comparison with TCP), it provides some advantages for IoT device developers; thus, it is mostly common among cellular IoT devices or lower power applications (e.g, smart agriculture sensors). UDP protocols can be used in spoofed network-layer volumetric attacks. Volumetric attacks are attacks with high packet rates that attempt to cause exhaustion of a server's or a network link's resources. UDP flood attack is a DDoS type in which many UDP packets are sent to a targeted server with the aim of overwhelming the server's ability to process and respond.

In a TCP SYN flood attack an attacker sends a huge number of packets with the SYN flag set to a victim server without the intention of ever completing any of the three-way handshakes. This attack implicitly puts strain on the network links but also has a protocol-level effect on the victim. The victim TCP server responds to the attacker's packets (with possibly spoofed source port) with SYN/ACK packets and maintains half-open connections until a timeout, since ACK packets will never arrive. Too many of these fake connection initiations will cause the victim server to be unable to establish connections with legitimate clients. The potential security issues of IoT devices make it easy for adversaries to exploit the IoT devices and make them part of a botnet that can launch DDoS attacks towards targets that reside on the internet. This doesn't only have an impact on the target machines, but indirectly on the network infrastructure as well, including the mobile infrastructure (in case of 3GPP access) that is delivering those packets to the Internet. This infrastructure includes the RAN, transport, and core domains.

IV. 3GPP RADIO PROTOCOLS FEATURE SELECTION

In this work, the method used to detect IP DDoS relies solely on analyzing the 3GPP radio protocols that are sent on the radio interface and processed by a base station (i.e eNB/gNB) function in RAN. The method does not inspect the IP packets that are encapsulated in the radio protocols. The relationship between the different sublayers of the radio protocol stack can be better understood in Figure 3. As shown, in the figure the MAC PDU transport block is formed of one or several MAC SDUs, where each MAC SDU encapsulates a RLC PDU that in turn encapsulates a PDCP PDU, where then the

PDCP PDU encapsulates the SDAP PDU. Lastly, the SDAP PDU encapsulates the IP packet being sent. A one-to-one mapping relationship can be drawn between PDCP, SDAP PDUs and the IP packets, where each PDCP and SDAP PDU corresponds to an IP packet.

We divided our features into two feature categories 1) Radio Protocol Traffic Pattern and 2) Radio Protocol Procedures and Functions. The following sections will describe those features and relevance to the attack detection.

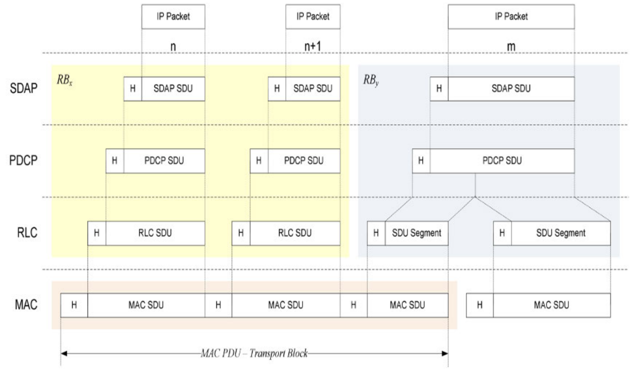

FIGURE 3. User plane data flow.

A. RADIO PROTOCOL TRAFFIC PATTERN

This feature category concentrates mainly on extracting features that are related to the traffic pattern behavior of the radio protocols in relation to the changes that occur in the IP based traffic behavior, in other words trying to draw the understanding if the radio protocol encapsulating the IP packet will change linearly in its pattern if the IP traffic changed as well or not. This category of features focuses on analyzing the PDCP and RLC sublayers, there are several reasons to choose those two sublayers. Firstly, due to the direct one-to-one mapping of the IP packet to the PDCP PDU, an increase in the IP packet size or increase in the rate of the IP packets per second would in turn reflect on the PDCP PDU length and rate. Secondly, in LTE there is a feature for RLC where it can concatenate multiple numbers of PDCP PDUs in one RLC PDU, thus increasing the size of a RLC PDU, and that can occur when sending high numbers of packets per second (pps). Moreover, we considered the physical layer mainly interested in the resources block utilization which lies in the frequency domain. The features are described more in details below:

1) PDCP TRAFFIC PATTERN FEATURES

The PDCP protocol [25] is a layer 2 protocol that is used to support procedures in the control plane or the user plane depending on which radio bearer (signaling radio bearer or data radio bearer) it is carrying data for. Several PDCP entities maybe defined for a UE. In our proposed detection method, we leverage PDCP that is associated with the transfer of user plane data, as this is where the IP packets are carried. We take into consideration PDCP PDUs features in the uplink and downlink direction. The PDCP PDU follows a specific structure as illustrated in Figure 4. The PDCP PDU data field encapsulates the SDAP PDU, where the SDAP PDU encapsulates the IP packet.

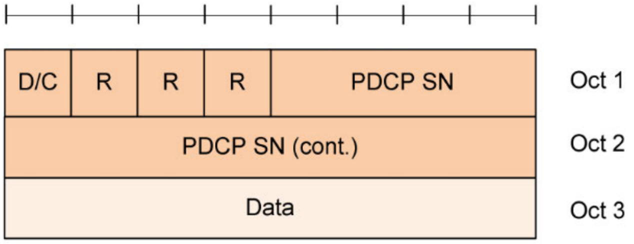

FIGURE 4. PDCP PDU user plane structure.

The IP packets are represented in the data field in the PDCP PDU format. As mentioned earlier, the PDCP PDUs have one to one mapping to IP packets, for example if a UE is sending 10 IP packets that in return will be proportional to 10 PDCP PDUs. From the characteristics and structure of PDCP, there are four possible features that can be derived: 1) PDCP count within a time window, 2) PDCP PDU size distribution, 3) PDCP UL to Downlink (DL) ratio, and 4) Time delay between PDCP PDUs.

PDCP count within a time window, is one feature that can be used for detection as it relies on the knowledge that each PDCP PDU corresponds to an IP packet. In the scenario of a DDoS attack, during a time window of 1 second, it could be determined if there is an ongoing attack or not by analyzing the PDCP PDUs count per second (or some established normal baseline), as that represents the packets per second (pps) characteristic of a DDoS attack. One example could be, within one second a TCP SYN flood attack that is generated from a computationally constrained device (e.g., raspberry pi) can generate 21,776 PDCP PDUs compared to the count of PDCP PDUs generated while rendering a webpage can be around 848 PDCP PDUs in the uplink direction (that is depending on the webpage and its contents, and the number of webpages rendered at the same time).

PDCP PDU size distribution, is yet another feature for detection as it reflects well a steady flow of small TCP packets of constant size. An example of a TCP SYN packet with options on the IP layer is 60 bytes long (TCP header: 20 bytes, TCP options: 20 bytes, IP header: 20 bytes), and it would add up to 63 bytes long in PDCP (SDAP header [1byte] + PDCP header [2bytes]). In a DDoS attack the TCP SYN packet length will remain the same during the attack period, while benign traffic might have a dynamic size distribution such as mobile broadband traffic or static size characteristic that occurs at distinct time intervals such as a MTC use case that we focus on in this work.

PDCP UL to DL ratio, can be used for detection, as in a DDoS attack some of the attacked targets may not be responsive, either they don't exist or the IP address exists but the service that is mapped to a specific TCP port that is not available. Thus, the number of UL PDCP PDUs will be higher than the DL PDCP PDUs.

Time delay between PDCP PDUs, is the final PDCP feature that can be used for detection, because for high packet rate traffic such as the case with DDoS attacks, the time arrival difference between packets will be very low. Thus, analyzing the time intervals between PDCP PDUs can give an indication if the incoming traffic is a part of an attack or not.

2) RLC TRAFFIC PATTERN FEATURES

The RLC sublayer [26] functionality in LTE is slightly different in 5G-NR, both share some common functionalities such as; error correction, segmentation and re-segmentation of RLC SDUs, duplicate detection, RLC SDU discard, RLC re-establishment. LTE RLC differs that it provides concatenation in RLC, which means that one RLC PDU contains one or more PDCP PDU. This is a feature that can be used in LTE networks, as an indicator of an attack as the higher the IP packets per second reflect on lower time intervals between PDCP PDUs that leads to concatenating multiple PDCP PDUs. This can be detected using the RLC PDU Size distribution for a given time.

3) PHYSICAL (PHY) LAYER RESOURCES PATTERN FEATURES

A resource block (RB) [27] is the smallest unit of resources that can be allocated to a subscriber. Where one resource block corresponds to a time slot that is represented in 0.5 ms. This means that each 0.5 ms a subscriber can be allocated some resources. Moreover, frequency units can be expressed in resource block, for instance 10MHz in frequency bandwidth can be described as 50 resource blocks. Depending on the reported buffer sizes that a UE needs to send to the network, the base station leveraging its scheduling algorithms allocated resource blocks that are suitable to be used by a UE to carry its data. The more data to be sent, the more resource blocks will be allocated for a single UE in different time slots. Thus, the more frequent the attacker sends IP packets, the base station will in return schedule continuous resources for the attacking device, thus it will impact the allocation and may lead to exhaustion of the physical resource.

B. RADIO PROTOCOL PROCEDURES AND FUNCTIONS

This category analyzes the features that are related to the functions and procedures for the different sublayer of the protocol stack, trying to understand if the change in IP traffic will affect the way radio protocols interact with the network through different functions like sending periodic reports about number of bytes to be sent, asking for resources from the network and other functions that are supported by the protocols. This category of features focuses on the MAC sublayer functions, which is composed of 1) MAC Control Element (CE) types, 2) Scheduling requests (SR), and 3) MAC buffer status (BSR) report.

1) MAC CONTROL ELEMENT (CE) TYPES

A MAC PDU consists of a MAC header, and a zero or more SDUs as shown in Figure 5, and zero or more MAC CEs. The MAC PDU header consists of one or more MAC PDU subheaders, where each subheader corresponds to either a MAC SDU, a CE or padding. The MAC CE is a special structure within the MAC PDU that carries control information.

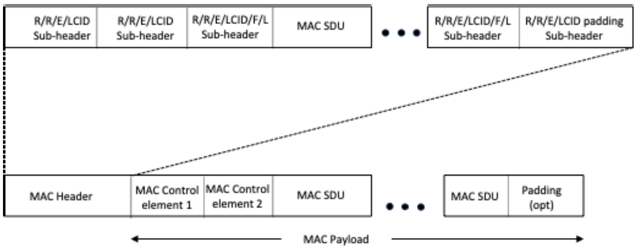

FIGURE 5. MAC PDU structure.

TABLE 1. Values of LCID for UL-SCH.

Index	LCID Values
00000	CCCH
00001-01010	Identity of the logical channel
11001	Extended Power Headroom Report
11010	Power Headroom Report
11011	C-RNTI
11100	Truncated BSR
11101	Short BSR
11110	Long BSR
11111	Padding

In our work, we focus more on the uplink MAC CE values that are sent from the malicious UE to the network, where snippet of the MAC CE values that are inscope are shown in Table 1 as specified in [28]

2) SCHEDULING REQUESTS

The MAC architecture, services and procedures are provided by 3GPP [28]. One of those procedures of interest is the scheduling procedure. Uplink Scheduling Requests (SR), are used by the UE to ask the network for uplink resources on the Uplink Shared Channel (UL-SCH) for a new transmission where the network in return sends back an UL Grant, so the UE will be able to transmit control or user plane PDUs. The scheduling requests are observed to be low in case of high number of PDCP PDUs per second at a given time, which in turn means high number of IP packets, compared to certain types of normal traffic.

3) MAC BUFFER STATUS REPORT

UL-BSR, is a yet another MAC layer procedure which is used by the UE to provide information about the amount of data available for transmission in the UL buffers to the serving gNB. Two main features of MAC BSR that can be useful for detection are: 1) Frequency of UL-BSRs (or time delta between two consecutive UL-BSRs) may be a valuable feature. In case of a DoS attack on the user plane it is expected that BSRs are more frequent due to constant sending of attack traffic, compared to a benign scenario with lower rate traffic. 2) Buffer sizes that are reported in BSR from the malicious UE to the gNB, due to short intervals between PDCP PDUs during an attack, the reported bytes will be constantly maintaining a high number during a given time.

C. STATISTICAL MEASURES

Our study focused on differentiating User Plane DDoS attacks from legitimate traffic by extensively analyzing traffic patterns. To achieve this, we continuously collected 3GPP radio protocol features from the Radio Access Network (RAN). We hypothesized that legitimate traffic and DDoS traffic would exhibit distinct probability distributions, enabling us to establish a discriminative plane for separating attack traffic from legitimate traffic. Statistical measures can effectively characterize the probability distribution for each specific feature within the traffic data. Therefore, rather than solely relying on the raw protocol features, we opted to extract multiple statistical measures for each feature and for each time interval t, we extract several statistical measures for every feature. This comprehensive approach allows us to gain a multi-faceted understanding of the traffic behavior within discrete time frames. By calculating these statistical measures, we can effectively capture temporal changes and variations in the data. The extracted statistical measures included:

1) Cumulative Sum: Providing insights into the temporal evolution of each feature's values within the interval t. for variable $\mathcal{X}$, the cumulative sum is obtained as:

$$CS = \sum_{x_i \in t} x_i \tag{1}$$

2) Mean: Representing the average value of each feature within the time interval, revealing its central tendency. The mean value is obtained as:

$$\mu = \frac{1}{n} \sum_{x_i \in t} x_i, \tag{2}$$

where n is the number of x_i within in the time interval t.

3) Standard Deviation: Quantifying the variability of traffic data by measuring the spread or dispersion around the mean. It is obtained as:

$$\sigma = \frac{1}{n} \sum_{x_i \in t} (x_i - \mu)^2 \tag{3}$$

4) Skewness: Assessing any asymmetry in the distribution of each feature's values within the interval, indicating tail tendencies. The skewness is obtained as:

$$Skew = \frac{1}{n\sigma^3} \sum_{x_i \in t} (x_i - \mu)^3 \tag{4}$$

5) Kurtosis: Analyzing the peakedness or tail behavior of the distribution, identifying heavy tails and it is obtained as:

$$Kurt = \frac{1}{n\sigma^4} \sum_{x_i \in t} (x_i - \mu)^4 \tag{5}$$

6) L2 (λ_2): Estimating L-moments to gain robust insights into the central tendency and spread of each feature's data. 1t is also so called L-scale and it is obtained as:

$$\lambda_2 = \frac{1}{2} \binom{n}{2} \sum_{x_1 < \cdots < x_i < \cdots < x_n} (-1)^{2-i} \binom{1}{i} x_i. \tag{6}$$

7) τ_3 and τ_4: Computing L-moments to capture information about skewness and kurtosis in a robust manner. The most useful of these ratios are τ_3 and τ_4 which are called L-skewness and L-kurtosis, respectively and are estimated as:

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \; where \; r \in \{3, 4\} \tag{7}$$

8) Entropy: Calculating Shannon entropy to quantify the information content and randomness of traffic patterns. The entropy of $\mathcal{X}$ is estimated as:

$$H(\mathcal{X}) = -\sum_{x_i \in t} p(x_i) \log_2(p(x_i)), \tag{8}$$

where $p(x_i) = Pr(\mathcal{X} = x_i)$ is the probability mass function of $\mathcal{X}$.

V. EXPERIMENTS

In order to evaluate different strategies for detection, we conducted a series of experiments with the aim of collecting data (candidate features) from the RAN stack under different conditions. We aimed to adhere to the following:

- Patterns of generated test traffic on IP level approximate selected real-world use cases and processes
- UE is resource constrained to limit the volume of the attack traffic
- Data collection is light-weight to enable data collection without disturbing the normal operation of the RAN implementation

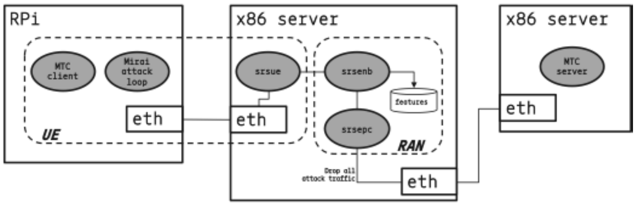

FIGURE 6. Experiments setup.

A. SETUP

Figure 6 shows the experiments setup that was used to obtain RAN features under the different kinds of IP traffic.

1) HARDWARE

The UE was represented in a split setup: the IP applications were run on a Raspberry Pi 3 Model B Rev 1.2 (ARM Cortex-A53, 1 GB RAM) and the UE software was running on an x86 server (Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz, 128 GB RAM). The same x86 server was used to run the ENB and EPC as well. The radio air interface was represented by a localhost connection between the UE software and the ENB, radio channel conditions were not modeled. The overhead of the Ethernet communication between the IP applications and the UE software was deemed negligible in terms of distorting the traffic patterns emanating from the applications.

2) SOFTWARE AND DATA COLLECTION

The radio stack (UE, ENB and EPC) was implemented in software using the srsRAN project, formerly known as srsLTE [29]. Relevant data from the ENB implementation, including candidate features were written into shared memory to enable parallel data processing/analysis or file-mapping for off-line analysis. Timing information was preserved by recording the CPU TSC (Time Stamp Counter) value for each data point for the different features in the respective event handlers. The following raw features were collected for each UE: PDCP PDU Sizes (uplink and downlink): `pdcp_pdu_sz_ul`, `pdcp_pdu_sz_dl`, reported buffer size by UE in MAC BSR: `mac_bsr_buf_sz`, size of the RLC pdu: `rlc_pdu_sz`, type of MAC Control Element: `mac_ce_type`, MAC PRB utilization: `mac_prb_util`.

3) TRAFFIC GENERATION

We developed SW that models traffic patterns emanating from an MTC (Machine Type Communication) application, as well as a version of a well-known DDoS attack tool. To generate benign MTC traffic, an application is used that mimics a non-time-critical industrial robot controller. Its payload traffic pattern can be characterized by a set of periodic communication channels over TCP with some variance in message size and inter-packet gap:

1) 160-320 Bytes messages (uniform distribution) at every 10 ms
2) 1514 Bytes messages every 200-500 ms (uniform distribution)

The attack traffic was modeled based on the Mirai botnet [30]. Only the attack loop (TCP SYN flood or UDP flood) was considered, running on a single core, traffic patterns related to scanning or communication with the CC server were not modeled. In the TCP SYN flood attack case, packets containing IP and TCP headers (with the SYN flag set) are prepared for each target (victim host or subnet). After the preparation, the attack tool loops through all targets multiple times. For each target, depending on configuration, it changes some fields (e.g., randomized source/destination IP, source/destination port, etc.), updates checksums and sends the packet to the target through a raw socket. The UDP flood attack is performed in a similar fashion, except there is an optionally randomized (per packet) UDP payload with configurable size (per attack). The attack loop may have slightly different runtime characteristics depending on the parameters of the attack such as whether to use randomization for ports or IPs and the list of target IPs/subnets.

B. EXPERIMENTS RUNS

As the main purpose of this detection method to try to detect DDoS attacks as fast as possible and close to its origination, we ran the experiments in short time periods, mainly 10, 30 with a maximum of 60 seconds. We noticed that when running the experiments, that it is possible to reach the same analysis if we ran the experiments for a long time (≈1 minute) or a short time period (≈1 second).

The drawn hypothesis is that even if the devices are compromised by an attacker and are part of a botnet, they will still keep on sending their normal traffic, as probably the attacker might be either periodic or the attacker wants to maintain their stealth. Thus, based on this hypothesis we combined both the benign MTC traffic and TCP SYN Flood malicious traffic in the experiments, where the attack runs in different traffic episodes at different time intervals. A summarization of the experiment runs can be found in Table 2.

TABLE 2. Traffic duration and attack episodes for each experiment run.

Traffic type	Total duration of traffic (in seconds)	Number of attack episodes	Duration of attack (in seconds per episode)	Target range
Mixed traffic (TCP SYN Flood and MTC benign traffic)	10	1	2	/24
	30	1	10	/24
	60	2	20	/24

VI. RESULTS

For each time interval t = 100 ms, we extracted several statistical measure features including the cumulative sum, the mean, the standard deviation, skewness, kurtosis, the second L-moment, the L-skewness, the L-kurtosis, and the entropy for each RAN protocol feature. In total, we obtained a dataset with 54 statistical measure features. To explore the potential utility of statistical features in identifying DDoS samples, we conducted an extensive analysis involving various machine learning (ML) models: K-nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), AdaBoost, Naive Bayes (NB), Linear Discriminant Analysis (LDA), and Logistic Regression (LGR). For our study, a total of 605 legitimate samples and 126 DDoS samples were utilized in the training phase. Each class represented 80% of its respective samples. Subsequently, these models were assessed for their performance in the testing phase using 152 legitimate samples and 31 DDoS traffic samples. The results of this analysis are summarized in Table 3 and shed light on the detection effectiveness of each ML algorithm. It is noteworthy that, with the exception of LGR, all other algorithms displayed robust performance in effectively distinguishing attack samples from legitimate traffic. Although LGR exhibited a zero False Positive Rate (FPR), the relatively lower True Positive Rate (TPR) of 0.65% indicates that a significant proportion of DDoS samples were erroneously classified as legitimate instances. This discrepancy led to a relatively modest accuracy of 17.5%. The subpar performance of LGR suggests the possible inclusion of irrelevant features within the feature set, a challenge effectively addressed by the other ML algorithms.

Despite its limitations, the inclusion of LGR in our analysis was deliberate for two main reasons. Firstly, logistic regression stands as a widely used and interpretable model for binary classification tasks. Its interpretability allows for the examination of the influence of individual features on classification decisions, thereby facilitating the understanding of each feature's discriminative power. Secondly, LGR is a fitting candidate for evaluating feature selection methods, as it is poised to perform better when extraneous features are eliminated from the model.

Through the evaluation of performance using the entire feature set, we established a baseline accuracy for comparative purposes, post the application of various feature selection techniques. As a result, subsequent sections of our study focus on the application of diverse feature selection methodologies with the intent of enhancing the classification performance of LGR.

TABLE 3. The performance result of applying ML algorithms on statistical feature set.

Classifier	FPr (%)	TPr (%)	ACC (%)	F1_score
KNN	0.0	98.7	98.9	0.97
SVM	0.0	98.7	98.9	0.97
DT	0.0	98.7	98.9	0.97
RF	0.0	98.7	98.9	0.97
AdaBoost	0.0	98.7	98.9	0.97
NB	0.0	98.7	98.9	0.97
LDA	0.0	97.4	97.8	0.93
LGR	0.0	0.65	17.5	0.29

First, to improve the classification performance, we experimented with different variance thresholds in the feature selection process. We applied variance analysis of features and tried several thresholds to discard features with low variance to see if the detection performance improved. However, the variance analysis did not lead to any significant improvement in the classification performance, suggesting that variance alone is not sufficient to capture the discriminative power of features. In light of this, we proceeded to use four more sophisticated feature selection methods: Analysis of Variance (ANOVA), Mutual Information, Recursive Feature Elimination (RFE), and Shapley Additive exPlanation (SHAP). These techniques are highly suitable for the process of selecting the most relevant features in classification tasks. By pinpointing the features that hold the most crucial information, these methods aid in improving the accuracy of classification between different classes, such as DDoS attacks and legitimate traffic.

ANOVA, for instance, plays a vital role in identifying features that exhibit substantial variations between distinct classes. Through evaluating the F-statistic for each feature, ANOVA effectively highlights those features contributing significantly to the differentiation between DDoS attacks and legitimate traffic. This technique focuses on capturing the differences in feature values that hold key discriminatory information.

Mutual Information, on the other hand, gauges the quantity of information a feature offers regarding the class labels. Its strength lies in capturing both linear and non-linear relationships existing between features and class labels. As a result, Mutual Information proves to be an invaluable criterion for feature selection, particularly in scenarios involving non-linear classification challenges. This method can effectively identify features that hold relevant insights for accurate classification.

In parallel, Recursive Feature Elimination (RFE) operates by iteratively removing the least influential features from the dataset. This process continues until an optimal subset of features is attained, ensuring that only the most informative features remain. RFE contributes to improving classification efficiency by eliminating noise and redundant information that might hinder accurate discrimination between different classes.

Lastly, Shapley Additive exPlanation (SHAP) is a sophisticated technique that provides a comprehensive understanding of feature importance within a model. It quantifies the contribution of each feature to the prediction and classification process, offering valuable insights into how individual features impact the final outcome. This transparency enables more informed decision-making regarding feature selection, leading to improved model accuracy.

In summary, ANOVA, Mutual Information, RFE, and SHAP offer distinct insights independently. We utilize these separate results to gather a broader understanding and compare their perspectives. By integrating these diverse insights, we gain a comprehensive view of feature importance. This approach strengthens our ability to differentiate between DDoS attacks and legitimate network traffic, enhancing the effectiveness of our analysis.

A. ANOVA

An F-statistic, also known as an F-test, is a statistical test that calculates the ratio between variance values, such as the variance from two different samples or the variance explained and unexplained by a statistical test, like ANOVA. In our study, we utilized the ANOVA method, which is a type of F-statistic called ANOVA F-test. The ANOVA F-test helps us identify the most important features among the statistical measures we extracted from time intervals (t = 100 ms). A higher F-test score indicates greater importance of the corresponding feature. Figure 7 displays the 20 most significant features identified through ANOVA analysis. The top two most important features are the sum of rlc_pdu_sz and the sum of pdcp_pdu_sz_ul, followed by the sum of mac_prb_util and the mean of rlc_pdu_sz. Once we determined the most important features, we utilized them to train a logistic regression model specifically for detecting DDoS traffic samples. Table 4 presents the confusion matrix for the test phase using the selected features. While the performance improved compared to the model trained with all features, it remained suboptimal.

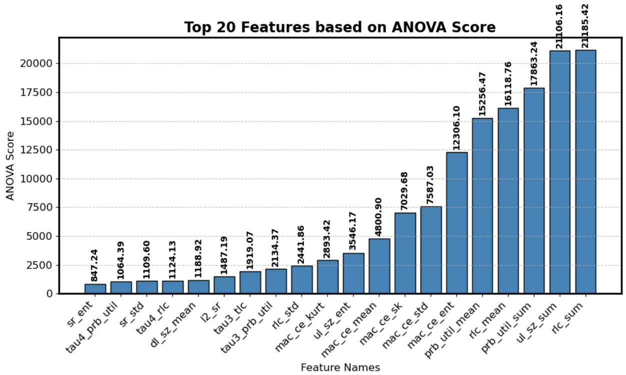

FIGURE 7. Top the first 20 features based on ANOVA score.

TABLE 4. The detection performance result using the first 20 features identified by ANOVA analysis.

Traffic	Legitimate	DDoS	Accuracy (%)
Legitimate	21	131	28.42
DDoS	0	31	

To enhance the performance systematically, we conducted a grid search. Instead of guessing the number of selected features, a grid search tests various combinations of features using ANOVA analysis to discover the best-performing model. The grid search identified the best combination of features, which turned out to be the sum of rlc_pdu_sz and the sum of pdcp_pdu_sz_ul - the two features with the highest scores in Figure 7. The scatter plot in Figure 8 displays the distribution of these two features with estimated kernel density estimation contours for both DDoS and legitimate traffic. As depicted in the figure, the density contours for the two classes are well-separated, indicating how effectively these two features can distinguish DDoS samples from legitimate traffic. We re-trained the logistic model using these two features and conducted the test phase accordingly. The results, summarized in Table 5, show a significant improvement in detection performance. The model successfully identifies all DDoS samples with a low false positive rate, demonstrating the effectiveness of the selected feature set for DDoS attack detection.

B. MUTUAL INFORMATION

Mutual information, derived from information theory, is a technique used in feature selection, drawing inspiration from information gain typically employed in decision tree construction. It quantifies the reduction in uncertainty for one variable when the value of another variable is known. While it is straightforward to calculate mutual information for two discrete variables (e.g., categorical input and categorical output data), it can be adapted for numerical input and categorical output scenarios. Similar to ANOVA analysis, we utilized mutual information to identify the 20 most significant features, as depicted in Figure 9, the two most influential features were found to be the L-skewness of mac_ce_type and the sum of rlc_pdu_sz, followed by the L-kurtosis of rlc_pdu_sz and the entropy of pdcp_pdu_sz_ul.

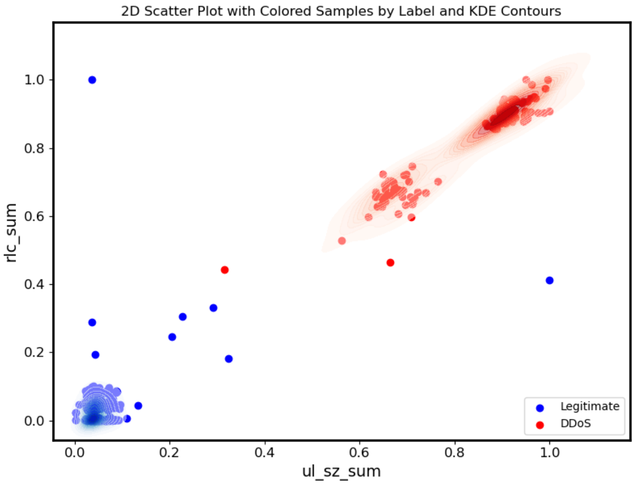

FIGURE 8. The scatter plot for the most important features obtained from the grid search utilizing ANOVA analysis.

TABLE 5. The Detection performance using the most important features obtained from the grid search utilizing ANOVA analysis.

Traffic	Legitimate	DDoS	Accuracy (%)
Legitimate	150	2	98.91
DDoS	0	31	

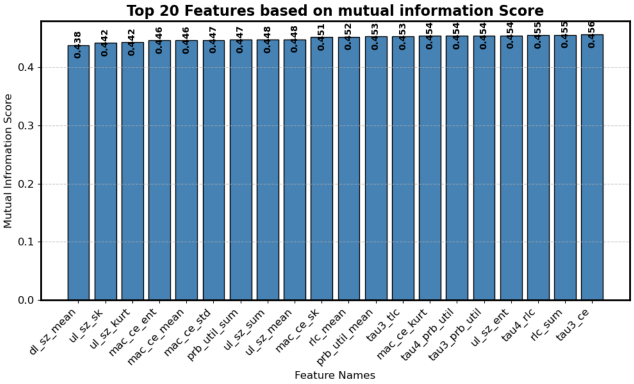

FIGURE 9. Top the first 20 features based on mutual information score.

Having determined these 20 crucial features, we proceeded to re-train the logistic model and evaluate its performance and we got the same perfromance result as the same as ANOVA analysis. Subsequently, we employed the grid search method with mutual information to identify the best feature combination. The results indicated that the L-skewness of mac_ce_type performed better than other feature combinations. In Figure 10, we visualize the KDE for the L-skewness of mac_ce_type for both DDoS and legitimate traffic. Observing the overlap of L-skewness values in certain regions, we anticipated a lower detection performance compared to ANOVA analysis. Table 6 presents the confusion matrix for the test phase, obtained using the best feature combination identified through mutual information. Although the performance is deemed high and acceptable, as expected, it falls short of the results summarized in Table 5 obtained through ANOVA analysis.

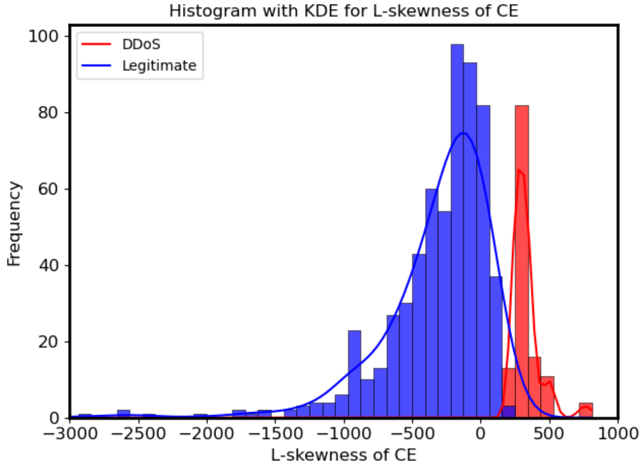

FIGURE 10. The KDE for the L-skewness of CE for both DDoS and legitimate traffic.

TABLE 6. The detection performance result using the best feature identified by grid search utilizing mutual information analysis.

Traffic	Legitimate	DDoS	Accuracy (%)
Legitimate	148	4	97.81
DDoS	0	31	

C. RFE

Recursive Feature Elimination (RFE) is a feature selection technique used in machine learning to identify and select the most relevant features from a dataset. It involves iteratively training a model on the full feature set, ranking the features based on their importance or contribution to the model, and eliminating the least important features until a desired number of features is reached. We employed the RFE technique to prioritize the most relevant features for distinguishing DDoS traffic from legitimate traffic. The first 20 ranked features are displayed in Figure 11, showcasing their importance according to the RFE process. Among these, the L-skewness of mac_prb_util, the L-skewness, and the L-kurtosis of pdcp_pdu_sz_dl stand out as the selected features for DDoS detection. Initially, we trained and tested a logistic regression model using the top 20 ranked features. The outcomes are summarized in Table 7, which presents confusion matrix for the test phase. In contrast to ANOVA and Mutual Information, where many samples are classified as attack samples, RFE incorrectly classified all attack samples as legitimate ones.

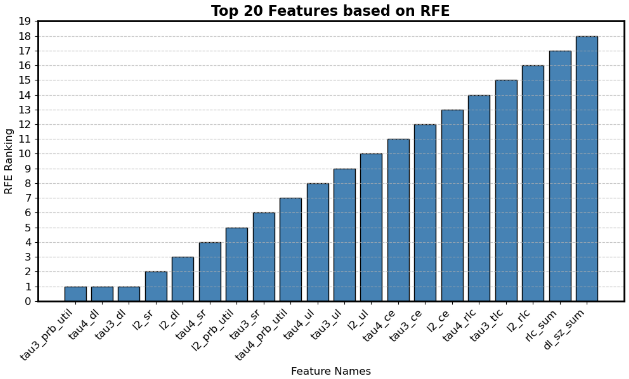

FIGURE 11. Top the first 20 features based on RFE.

TABLE 7. The detection performance result using the first 20 features identified by RFE.

Traffic	Legitimate	DDoS	Accuracy (%)
Legitimate	152	0	83.06
DDoS	31	0	

Before proceeding with training and testing the logistic regression model using the three selected features, we created a scatter plot of the three features (depicted in Figure 12). This scatter plot visually illustrates how the DDoS samples and legitimate samples are distinct in the three-dimensional space. As indicated by the figure, we anticipate higher detection performance using these three features compared to the top 20 ranked features. Table 8 provides a summary of the confusion matrix for the test phase of the logistic model using the selected three features. Notably, during the training phase, RFE outperforms ANOVA and mutual information. Furthermore, RFE achieves comparable performance to ANOVA during the test phase.

As Table 3 indicates, the LGR model initially exhibits suboptimal detection performance when considering the entirety of statistical features. However, the integration of feature selection techniques, namely ANOVA, Mutual Information, and RFE, serves as an avenue to enhance the model's classification proficiency. This augmentation leads LGR to attain a level of accuracy equivalent to that of other mentioned ML algorithms. Overall, the cumulative analysis conducted thus far underscores the effectiveness of the statistical metrics extracted from the RAN protocol statistics in detecting DDoS samples.

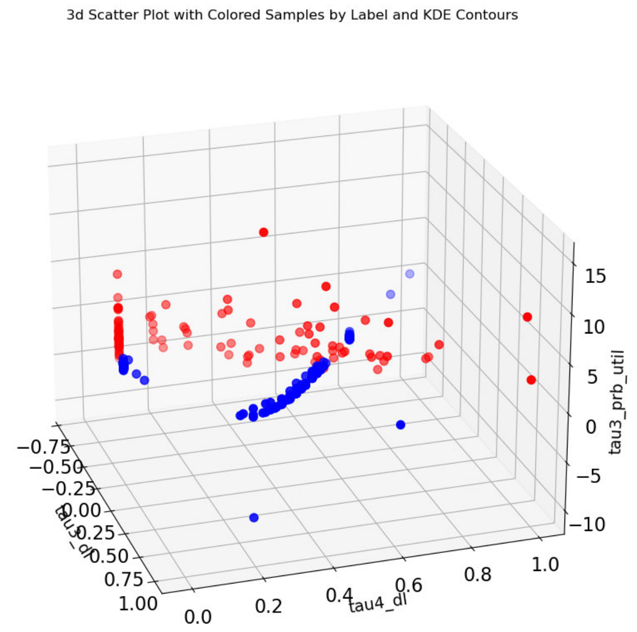

FIGURE 12. The scatter plot for the three first ranked features obtained from RFE.

TABLE 8. The detection performance using three first ranked features obtained from RFE.

Traffic	Legitimate	DDoS	Accuracy (%)
Legitimate	150	2	98.91
DDoS	0	31	

In the subsequent section, we intend to further deepen our understanding of the classification process. To achieve this, we employ SHapley Additive exPlanations (SHAP) analysis-a technique within the realm of explainable AI (XAI). Through SHAP analysis, we seek to attain a more comprehensive comprehension of the features contributing to the identification of both DDoS samples and legitimate traffic instances. This approach is pivotal in unraveling the intricate relationships between features and classification outcomes, providing valuable insights into the model's decision-making process.

D. SHAP

SHAP stands for SHapley Additive exPlanations, which is a versatile technique used to interpret the outcomes of machine learning models, regardless of the model type. This method works by assigning a value to each feature, indicating its contribution towards the final prediction. Positive SHAP values indicate that a feature positively impacts the output, whereas negative values suggest a negative impact. Two figures, namely Figure 13 and Figure 14, visually depict SHAP's significance.

The first figure, the SHAP feature importance plot, showcases the ten most crucial features, ranked by their importance. Importance is determined based on the average absolute SHAP value across the dataset. This plot serves as a rapid means of identifying key model features and comparing their relative importance. As shown in Figure 13, the contribution of each feature appears balanced for both Legitimate and DDoS traffic. Specifically, the features the sum of rlc_pdu_sz and the sum of pdcp_pdu_sz_ul emerge as the most influential contributors, a finding that aligns with ANOVA analysis.

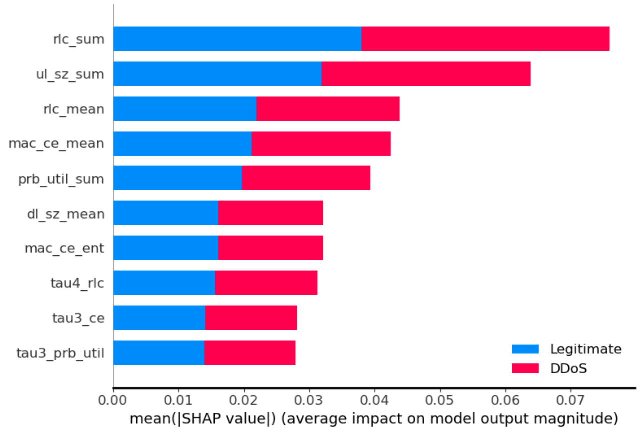

FIGURE 13. SHAP feature importance plot.

Conversely, the second figure, the SHAP summary plot, is a scatter plot illustrating the influence of each feature across all dataset instances. The x-axis represents SHAP values, while the y-axis signifies feature values. This summary plot visually summarizes the distribution of SHAP values for each feature and their relationship with predictions. As presented in Figure 14, this figure portrays the summary plot for legitimate traffic. For DDoS traffic, the plot would be a mirror image of this one. For instance, the feature the sum of rlc_pdu_sz negatively affects predicting normal traffic for higher values (highlighted in red), while lower values (indicated in blue) positively impact the prediction. Similarly, the positive impact of the L-skewness of mac_prb_util on legitimate traffic classification is observed for higher values, with lower values exerting a negative effect.

VII. DISCUSSION & FUTURE WORK

The paper focuses on enabling RAN to detect compromised UEs belonging to cellular botnets during DoS attacks at the very edge of the network. The work presents a methodology to select powerful features that can serve as a basis for inferring on-going attacks. The benign and attack traffic used in the data analysis represent only a fraction of use cases expected in real-world scenarios. However, the features identified by the analysis align well with the intuition based on the expected radio protocol impact of the two classes of IP traffic. As an example, the uplink RLC PDU size is an aggregate view of the uplink traffic volume which is rather high during an attack compared to the benign scenario. Data from a more diverse set of benign cases would be necessary to design features that generalize well. The set of collected features from RAN can also be expanded, also taking into consideration the cost of obtaining the selected features from actual implementations.

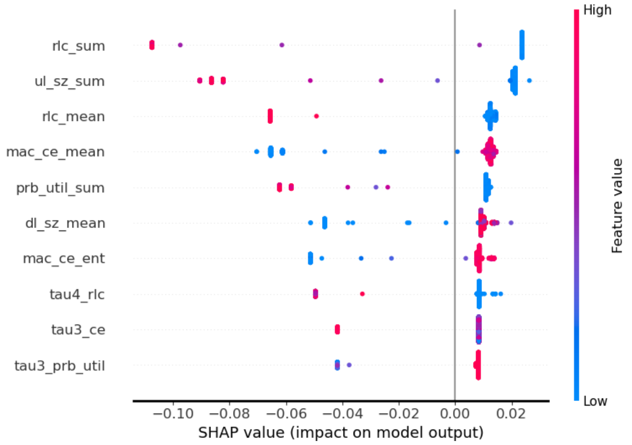

FIGURE 14. SHAP summary plot for Legitimate traffic.

Additionally, when compared to previous studies, such as the methodology outlined in [13], which relies on counting packets per second and CUSUM, our method performs notably better. The number of packets per second, typically an IP-based network statistic, can be correlated with the sum of PDCP PDUs, which serves as a statistical feature within our proposed solution. Despite this mapping, the feature analysis approaches (ANOVA, SHAP etc.) employed in our paper reveal that other statistical features obtained from radio protocol statistics exhibit superior performance compared to merely using the sum of PDCP PDUs. As our proposed solution is model-agnostic and primarily focuses on leveraging statistical features derived specifically from radio protocol statistics, it's safe to assert that our proposed features showcase a performance edge over both the method and the features utilized in [13]. Furthermore, since our method centers on detecting anomalies at the radio level, preempting malicious packets before their entry into the network to potentially overwhelm it, our solution offers the advantage of early detection compared to the proposed method in [13]. This proactive stance allows for the identification of malicious packets at an earlier stage, thereby fortifying the network against potential disruption or damage.

We intend to continue the research on this topic, our future work is planned to extend the simulations and experiments to include other types of bengin traffic types as part of our traffic generation, an example could be upload traffic. The reason behind that is the importance to find the suitable features for the different type of traffic to ensure detection accuracy and provide an advantage to differentiate between high load traffic and attack in some traffic use cases. An additional item is to collect data from live networks to help on fine tuning the detection algorithms and feature engineering process.

Additionally, we will investigate other DoS attack categories, other than volumetric attack, for example slow resource exhaustion attacks to validate that our approach presented in this paper can detect other types of DoS attacks as well. Moreover, continuing with the Mirai botnet use case expand the detection capabilities to detect other phases of the attack, for example the infection phase, where the infected device tries to search for other vulnerable devices in order to comproise and initiate the attack, we believe this can be possible since even Machine to Machine (M2M) traffic within the same cell coverage has to be processed by the basestation, thus there is an opportunity to analyze the traffic and probably be able to detect the attack even in earlier phases than the described in this paper.

On the other hand, there is further work to be done on the detection and analysis part. Notably, all evaluations have been conducted in the time-domain. Consequently, as a potential next step, we intend to explore the characteristics of RAN protocol statistics in the frequency domain. This exploration aims to ascertain whether incorporating frequency domain-based attributes can enhance the current solution's performance.

VIII. CONCLUSION

In this research paper we presented a novel detection method to detect user plane DDoS attacks more closer to the originating attack source, thus can enhance the time to detect and react to those type of attacks before exhausting radio resource, or propagating more into the network. This method relies on 3GPP radio protocols analysis without the need to inspecting the encapsulated user plane packets (e.g, IP packets).

Numerous statistical attributes were derived from the statistics of the RAN protocol. These attributes were examined to determine their effectiveness in distinguishing between DDoS samples and legitimate ones. Various analyses were conducted to evaluate the significance of these attributes. Moreover, the acquired features were employed as inputs for multiple machine learning algorithms, resulting in a notable achievement of strong performance in detecting DDoS attacks. Although all experiments and analyses were conducted within the framework of a 5G mobile network, the method proposed in this paper possesses the adaptability to be seamlessly applied within the context of 6G and beyond (XG). This adaptability underscores its potential as a valuable roadmap for designing secure and resilient networks to counter DDoS attacks, not only in the present but also in the ever-evolving landscape of future mobile network generations.

ACKNOWLEDGMENT

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) through the 1515 Frontier Research and Development Laboratories Support Program under Project 5169902, and in part by the Hexa-X II Project through the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe Research and Innovation Programme under Grant 101095759.

REFERENCES

[1] S. Guo, B. Lu, M. Wen, S. Dang, and N. Saeed, "Customized 5G and beyond private networks with integrated URLLC, eMBB, mMTC, and positioning for industrial verticals," *IEEE Commun. Standards Mag.*, vol. 6, no. 1, pp. 52–57, Mar. 2022.

[2] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, "Towards enabling critical mMTC: A review of URLLC within mMTC," *IEEE Access*, vol. 8, 2020.

[3] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 905–929, 2nd Quart., 2020.

[4] D. Mishra, N. R. Zema, and E. Natalizio, "A high-end IoT devices framework to foster beyond-connectivity capabilities in 5G/B5G architecture," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 55–61, Jan. 2021.

[5] W. Shi, W. Xu, X. You, C. Zhao, and K. Wei, "Intelligent reflection enabling technologies for integrated and green Internet-of-Everything beyond 5G: Communication, sensing, and security," *IEEE Wireless Commun.*, vol. 30, no. 2, pp. 147–154, Apr. 2023.

[6] R. F. Fouladi, O. Ermiş, and E. Anarim, "A DDoS attack detection and countermeasure scheme based on DWT and auto-encoder neural network for SDN," *Comput. Netw.*, vol. 214, Sep. 2022, Art. no. 109140.

[7] G. Nayak, A. Mishra, U. Samal, and B. K. Mishra, "Depth analysis on DoS & DDoS attacks," in *Wireless Communication Security*. Hoboken, NJ, USA: Wiley, 2022, pp. 159–182.

[8] M. K. Hasan, T. M. Ghazal, R. A. Saeed, B. Pandey, H. Gohel, A. A. Eshmawi, S. Abdel-Khalek, and H. M. Alkhassawneh, "A review on security threats, vulnerabilities, and counter measures of 5G enabled Internet-of-Medical-Things," *IET Commun.*, vol. 16, no. 5, pp. 421–432, Mar. 2022.

[9] R. Sridharan, R. R. Maiti, and N. O. Tippenhauer, "WADAC: Privacy-preserving anomaly detection and attack classification on wireless traffic," in *Proc. 11th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, Jun. 2018, pp. 51–62.

[10] A. Serrano Mamolar, P. Salvá-García, E. Chirivella-Perez, Z. Pervez, J. M. A. Calero, and Q. Wang, "Autonomic protection of multi-tenant 5G mobile networks against UDP flooding DDoS attacks," *J. Netw. Comput. Appl.*, vol. 145, Nov. 2019, Art. no. 102416.

[11] N. Ravi and S. M. Shalinie, "Learning-driven detection and mitigation of DDoS attack in IoT via SDN-cloud architecture," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3559–3570, Apr. 2020.

[12] M. G. Pérez, A. H. Celdrán, F. Ippoliti, P. G. Giardina, G. Bernini, R. M. Alaez, E. Chirivella-Perez, F. J. G. Clemente, G. M. Pérez, E. Kraja, G. Carrozzo, J. M. A. Calero, and Q. Wang, "Dynamic reconfiguration in 5G mobile networks to proactively detect and mitigate botnets," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 28–36, Sep. 2017.

[13] K. Doshi, Y. Yilmaz, and S. Uludag, "Timely detection and mitigation of stealthy DDoS attacks via IoT networks," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 5, pp. 2164–2176, Sep./Oct. 2021.

[14] R. Fuladi, T. Baykas, and E. Anarim, "The use of statistical features for low-rate denial of service attack detection," in *Proc. 2nd Int. Conf. 6G Netw. (6GNet)*, Oct. 2023, pp. 1–6.

[15] L. Stahle and S. Wold, "Analysis of variance (ANOVA)," *Chemometrics Intell. Lab. Syst.*, vol. 6, no. 4, pp. 259–272, Nov. 1989.

[16] S. Lall, D. Sinha, A. Ghosh, D. Sengupta, and S. Bandyopadhyay, "Stable feature selection using copula based mutual information," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107697.

[17] X.-W. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Proc. 6th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2007, pp. 429–435.

[18] S. Mane and D. Rao, "Explaining network intrusion detection system using explainable AI framework," 2021, *arXiv:2103.07110*.

[19] R. Harada, N. Shibata, S. Kaneko, K. Honda, J. Terada, Y. Ishida, K. Akashi, and T. Miyachi, "Quick suppression of DDoS attacks by frame priority control in IoT backhaul with construction of Mirai-based attacks," *IEEE Access*, vol. 10, pp. 22392–22399, 2022.

[20] B. M. Xavier, M. Dzaferagic, D. Collins, G. Comarela, M. Martinello, and M. Ruffini, "Machine learning-based early attack detection using open RAN intelligent controller," 2023, *arXiv:2302.01864*.

[21] *NR and NG-RAN Overall Description*, document TS 38.300, Version 17.5.0, 3GPP, 2023.

[22] Netscout. (2022). *Threat Intelligence Report 2022*. [Online]. Available: https://www.netscout.com/threatreport/wp-content/uploads/2023/04/Threat-Report-2H2022.pdf

[23] Lumen. (2022). *Lumen Quarterly DDoS Report Q4*. [Online]. Available: https://assets.lumen.com/is/content/Lumen/lumen-quarterly-ddos-report-q-4-22

[24] (2023). *Internet Engineering Task Force*. [Online]. Available: https://www.ietf.org

[25] *NR; Packet Data Convergence Protocol (PDCP) Specification*, document TS 38.323, Version 17.5.0, 3GPP, 2023.

[26] *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) Protocol Specification*, document TS 36.322, Version 17.0.0, 2023.

[27] *NR; Physical Channels and Modulation*, document TS 38.211, Version 17.5.0, 2023.

[28] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification*, document TS 36.321, Version 17.5.0, 2023.

[29] I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "srsLTE: An open-source platform for LTE evolution and experimentation," in *Proc. 10th ACM Int. Workshop Wireless Netw. Testbeds, Exp. Eval., Characterization*, 2016, pp. 25–32.

[30] J. Gamblin. (2017). *Mirai-Source-Code: Leaked Mirai Source Code for Research/IoC Development Purposes*. [Online]. Available: https://github.com/jgamblin/Mirai-Source-Code/tree/master

LOAY ABDELRAZEK received the B.Sc. degree in electromechanical engineering from Alexandria University, Alexandria, Egypt, in 2012, and the M.Sc. degree in information security from Nile University, Cairo, Egypt, in 2018. In 2019, he joined Ericsson Standards and Technology, Sweden, as a Researcher of concepts security. His research interests include automated detection and response for radio attacks, security automation and management, and adaptive security capabilities in radio access networks.

RAMIN FULADI received the B.Sc. degree from the Amirkabir University of Technology, Tehran, Iran, in 2004, and the M.Sc. and Ph.D. degrees in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, in 2014 and 2021, respectively. He is currently an AI/Security Researcher with Ericsson Research. His research interests include 5G/6G, physical layer security, anomaly detection, intrusion detection, network-based detection, data analysis, machine learning, and artificial intelligence.

JÁNOS KÖVÉR, photograph and biography not available at the time of publication.

LEYLI KARAÇAY received the M.Sc. and Ph.D. degrees in computer science and engineering from Sabancı University, Turkey, in 2012 and 2020, respectively. She was a Teaching Assistant with Sabancı University for seven years. Since October 2019, she has been a Security Researcher with Ericsson Research, Turkey. Her research interests include AI/ML security and privacy, data privacy, homomorphic encryption, and machine learning.

UTKU GÜLEN received the B.Sc. degree in electronics and communication engineering from Yıldız Technical University, Istanbul, Turkey, in 2012, and the M.Sc. and Ph.D. degrees in computer engineering from Bahcesehir University, Istanbul, in 2014 and 2022, respectively. In 2020, he joined Ericsson Research, Turkey, as an experienced Security Researcher. His research interests include applied cryptography, network and computer security, public-key cryptography on the IoT devices, and embedded systems.