Understanding the Security Implications in O-RAN with Abusive Adversaries

Mark Megarry, Antonino Masaracchia, Muhammad Fahim, Vishal Sharma(✉), and Trung Q. Duong

School of Electronics, Electrical Engineering and Computer Science (EEECS), Queen's University Belfast (QUB), Belfast, NI, UK
{mmegarry04,a.masaracchia,m.fahim,v.sharma,trung.q.duong}@qub.ac.uk

Abstract. Open-Radio Access Network (O-RAN) is considered the next scalable solution, which aims to devolve the network into Near-real-time RIC and Non-real-time RIC to have far more flexibility in services with adaptable components. This disaggregation, however, will have broader security implications, primarily arising because of the use of legacy systems in the new architecture. Current threat models take a lighter tone towards the evaluation of security measures. Thus, strict adversarial methods must be adopted, which can consider scenarios of cyber-vandalism in such networks. Based on this ideology, the article presents security implications posed by abusive adversaries in the offloading procedures. This methodology provides a viewpoint on how an adversary forms predictive methods on when to attack the system, which is followed by mitigation mechanisms for the network to avoid it from happening. The work is based on the Markov Decision Process (MDP) and a Fuzzy Inference System (FIS), which uses Synthetic Data Augmentation for Tabular Data (SMOTE) to generate a set of metrics that can offer a high probability of attack in the transition mode to the adversary. The implications are presented using a synthetic dataset created on the backbone of the simulated scenario in NS3 and followed by mitigation strategies.

Keywords: Security · O-RAN · Abusive Adversary · Simulations

1 Introduction

Open Radio Access Network (O-RAN) is a radio access network (RAN) architecture specified by the O-RAN Alliance, which emphasises disaggregation, virtualization, open interfaces, and interoperability between vendors [1]. The potential benefits of O-RAN include provisions for a multivendor ecosystem, reduced cost, use of commercial off-the-shelf (COTS) hardware, improved scaling flexibility, energy efficiency, and enhanced security [2,3]. It is expected that O-RAN will

N.-S. Vo et al. (Eds.): INISCOM 2024, LNICST 595, pp. 215–235, 2024.
https://doi.org/10.1007/978-3-031-67357-3_16

considerably impact the capital expenditures (CapEx) and operating expenses (OpEx) due to the multi-vendor environment, open-source software and hardware designs/implementation, and scalability enabled by the architecture and offered services via xApps [3]. The O-RAN architecture is built on the principles of disaggregation, intelligent/closed-loop control, virtualization, and open interfaces [1]. An overview of the network architecture, as detailed in [4,5], is illustrated in Fig. 1 along with the security functions of the 5G core network [6–8]. The following points summarise the operation of key components of O-RAN architecture [1,4]:

- O-RU: The O-RAN Radio Unit (O-RU) is a physical node which carries out low physical layer functions of the radio interface with the user equipment (UE) and connects to the Open Fronthaul interface [4].
- O-DU: The O-RAN Distributed Unit (O-DU) is a logical node (implemented either by virtualized or non-virtualized means) which may be connected to one or more O-RUs, and is compatible with the functions of a gNB Distributed Unit (gNB-DU) [4]. The O-DU may also support the management of O-RUs [4].
- O-CU: The O-RAN Central Unit (O-CU) comprises two logical nodes/planes: The O-CU Control Plane (O-CU-CP) and the O-CU User Plane (O-CU-UP) [4]. This unit implements high 3GPP layers, including the Radio Resource Control (RRC), Service Data Adaptation Protocol (SDAP), and Packet Data Convergence Protocol (PDCP) layers [1].
- Near-RT RIC: The Near-Real-Time RAN Intelligent Controller (Near-RT RIC) is a network function which allows for the control and optimisation of E2 nodes through the use of control loops with periods between 10ms and 1 s [4]. A vital component of the Near-RT RIC is a number of microservice-based applications known as xApps, which may process data generated by the RAN and generate control actions for the E2 nodes connected to the Near-RT RIC [1,4].
- Non-RT RIC: The Non-Real-Time RAN Intelligent Controller (Non-RT RIC) is part of the Service Management and Orchestration (SMO) framework, and it is responsible for optimizing RAN operation via control loops lasting longer than 1 s [1]. A key component of the Non-RT RIC is a collection of applications known as rApps, which provide added value services for RAN optimization and operations support [1]. rApps may generate policy guidance for the Near-RT RIC, provide enrichment information to the Non-RT RIC, and carry out configuration management and data analytics [1].
- SMO: The SMO contains the Non-RT RIC, and enables RAN support services such as a fault, configuration, accounting, performance, security (FCAPS) interface, and O-RAN Cloud (O-Cloud) management [4].
- O-Cloud: The O-Cloud is a cloud computing platform which hosts O-RAN functions, supporting software components, and management and orchestration functions [4].
- O-eNB: The O-RAN eNB (O-eNB) hosts functions of an O-DU and an O-RU, which are connected by an Open Fronthaul interface [4].

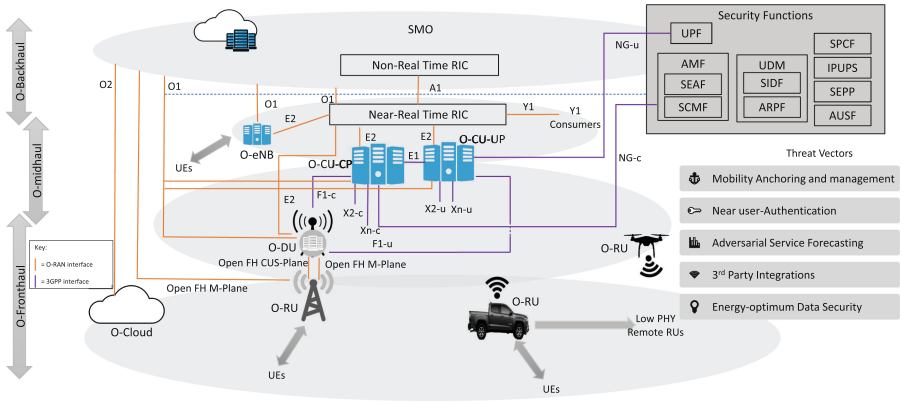

Fig. 1. An overview of the O-RAN architecture with 5G core security functions [4,6–8]

Despite the benefits of O-RAN discussed previously, making it an attractive RAN architecture for network operators, the specifications of O-RAN present new security challenges. Previous work has identified the following categories of technology-related threats to O-RAN networks [9–11]:

- Architecture openness: The open nature of O-RAN potentially creates threats relating to human error, supply chain software, supply chain hardware, vulnerable open source software and API exploitation [11].
- Radio/open interface: Threats related to the fronthaul, the CU/DU, and the 5G radio network [9].
- Intelligence: As artificial intelligence (AI)/machine learning (ML)-enabled control loops are to be an integral part of O-RAN architecture, there are threats specific to O-RAN due to this provision which do not affect existing RAN architectures such as V-RAN or C-RAN [9].
- Virtualisation: This category describes threats related to network functions virtualisation (NFV). As functions in the RAN are increasingly implemented as virtual network functions (VNFs) or cloud network functions (CNFs) rather than physical network functions (PNFs), the security of the components hosting these functions, e.g., the virtual machines (VMs), containers (CNs), and hypervisors, becomes more prevalent than in previous RAN architectures [9]. A notable exception is that most of the threats in this category also apply to V-RAN and/or C-RAN [9].
- Threats against O-RAN system: Threats related to the O-RAN system architecture, including implementation of insufficient security features on the interfaces between components [10].
- Threats against O-Cloud: Threats against the O-Cloud platform are largely derived from previous studies on virtualization and containerization security, mostly falling under the previously-mentioned virtualisation category [10].

- Physical threats: These threats are related to the use of hardware, and include a user with physical access to a site accessing O-RAN components, or accessing the fronthaul cable network [10].
- REST protocol stack threats: These threats relate to the use of the representational state transfer (REST) protocol stack (which includes JSON, HTTP, TLS, TCP, IP) over the A1 and R1 interfaces [10]. According to O-RAN Alliance Work Group 11 (WG11), each of these protocols has known vulnerabilities [10].

The O-RAN architecture specification allows for the CU of the O-RAN network to be connected to the 5G core network via the NG-c and NG-u interfaces, which connect to the Access and Mobility Management Function (AMF), and User Plane Function (UPF) of the 5G core network respectively [4]. Specifications in [6–8] and [12] describe the Authentication Server Function (AUSF), Authentication credential Repository and Processing Function (ARPF) − which is a component of the Unified Data Management (UDM), Inter-Public-Land-Mobile-Network User Plane Security (IPUPS), Security Context Management Function (SCMF) − which is a component of the AMF, Subscription Identifier De-concealing Function (SIDF) − which is a component of the UDM, Security Anchor Function (SEAF) − a component of the AMF, Security Edge Protection Proxy (SEPP), and Security Policy Control Function (SPCF) as the critical security entities in the 5G Core network.

In this domain, the impact of adversaries increases as the number of attack surfaces in O-RAN's architecture is high. Furthermore, it becomes critical to evaluate the security of the networks using stronger non-simulated adversarial models, which can help realise the true potential of the security solutions. One of the adversarial models in this direction would be abusive adversaries, which is a model of an adversarial agent who aims to cause damage to a target system without any reward, regardless of the resources they must expend to achieve this goal [13]. The threat model of an abusive adversary considers network entities coming under the control of the adversary randomly [14]. An abusive adversary may use their resources to attempt to evade detection by the system under attack [13], or influence misbehaviour detection rules to ensure that network entities under their control remain undetected [14]. In the context of the 5G and beyond IoT network discussed in [14], an abusive adversary is capable of carrying out a broad range of attacks, including host-impersonation, replay attacks, denial of service, distributed denial of service, and hidden-terminal attacks [14]. When targeting O-RAN networks, an abusive adversary may attack the Quality of Service (QoS) management functions, xApps, AI/ML models, and mobility management functions hosted in the Near-RT RIC as illustrated in Fig. 2.

To categorise the threats posed by an abusive adversary towards O-RAN, six threat agent models identified by WG11 of the O-RAN Alliance in [10] are considered:

- Cyber-criminals: These threat agents use computers to commit crimes and/or criminally target computer systems [10]. A cyber-criminal may have the goal

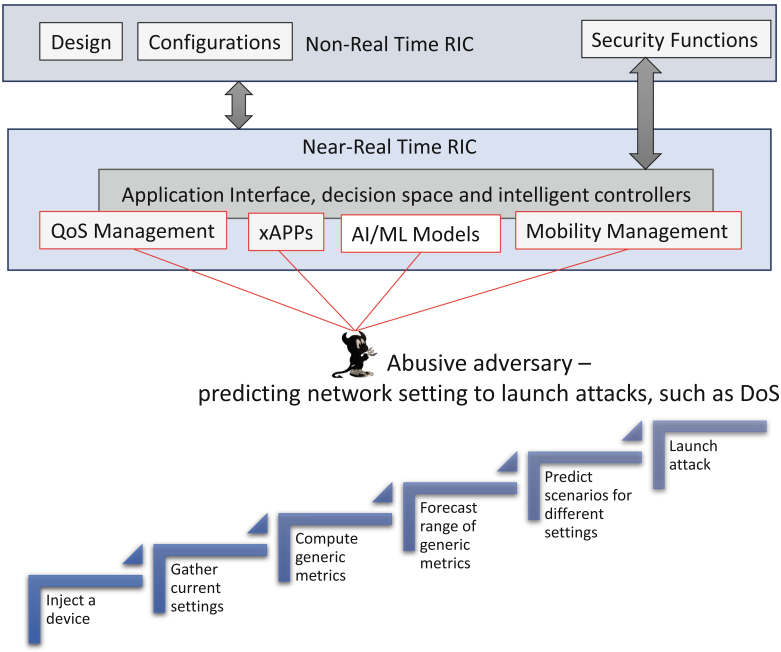

Fig. 2. An exemplary illustration of abusive adversary's region of interest in Near-RT RIC in O-RAN.

of short-term personal gain from attacking a system, e.g., selling compromised user data.

- Insiders: These threat agents are trusted individuals who abuse privileged access to a system to carry out attacks related to the system [10].
- Hacktivists: These threat agents carry out attacks with the goal of some political or social gain [10].
- Cyber-terrorists: These threat agents use computers with the aim to carry out violence, cause fear, or cause financial damage [10,15]. They may be politically motivated and may target subnational groups [10,15].
- Script kiddies: These threat agents lack computational resources and technical knowledge [10]. However, they may take advantage of existing tools to probe for and exploit known vulnerabilities in a system.
- Nation-state: These are threat agents supported by a nation-state [10]. They may seek to gain persistent access to networks, possibly in adversarial countries, to acquire, manipulate or destroy information [10]. As they are supported by a nation-state, they may have access to significant computational resources and expertise [10].

Due to their purely malicious goals (i.e., they are only trying to cause damage to a system without any monetary reward) and large amounts of computing resources available, the abusive adversary model could potentially be used as a

basis for modelling well-equipped hacktivists, cyber-terrorists, and nation-state actors with the aim of harming network performance or infrastructure.

1.1 Our Contributions

This article helps understand the security implications in O-RAN with abusive adversaries [16], thus helping to check the network's security measures. The key contributions are listed below:

- The understanding of the abusive adversary is enhanced specifically in an offloading scenario where the number of attack surfaces increases, and the probability of an attack is high.
- A combination of literature-based Markov Decision Process (MDP) [16–18], Fuzzy Inference System (FIS) and Synthetic Data Augmentation for Tabular Data (SMOTE [19]) are considered to form intelligence mechanisms for abusive adversaries to form the strategy of attack and further understand the impact it may have on the network by forming synthetic dataset on the backbone of the traffic modelled via NS3.
- The method utilised in the article can help form the threat models for O-RAN as it can expose the network settings which are adversary-friendly and must be avoided. Such mechanisms can be modelled for different scenarios and not just offloading.

2 Related Works

This section presents some of the most recent works related to O-RAN security and abusive adversaries. The works related to O-RAN security include broad overviews of the topic, security analyses, investigations, and proposals for novel technology implementations to support the security of future O-RAN networks. In terms of work related to abusive adversaries/abusive modelling, it has been investigated in terms of the impact in 5G-IoT networks [14], cyber-physical systems [13], and blockchain systems [16] with details compared in Table 1.

Liyanage et al. [9] provided an overview of Open RAN architecture and gave a taxonomy of threats relating to O-RAN networks. At the highest level, they divided these threats into the three categories of "Process", "Technology", and "Global". For each threat detailed, the authors also stated whether it is specific to O-RAN or if it is applicable to other RAN architectures such as C-RAN or V-RAN. Polese et al. [1] provided a broad overview of O-RAN. In terms of security, their paper discussed the relevant stakeholders, the threat surface presented by O-RAN, and security principles and opportunities.

Abdalla and Marojevic [20] split the O-RAN architecture into a number of security domains, described the characteristics of each domain, and provided a table of O-RAN security risks identified by the O-RAN Alliance Security Work Group (SWG). The paper then moves to focus on the Open Fronthaul security threats. Soltani et al. [21] investigated an attack against bearer context migration in an O-RAN-based 5G network. This attack causes network anomalies,

resulting in a significant decrease in cell performance. These works are pivotal in understanding the impact of adversaries in open network settings.

Klement et al. [22] presented the key stakeholders in the O-RAN system, and discussed methods to mitigate the threats associated with each stakeholder. In [23], Wen et al. proposed a solution which gathers telemetry data from network entities to support security services running as xApps in the Near-RT RIC. In a similar direction, Abdalla et al. [24] provided an overview of the use cases and limitations of O-RAN, as well as a discussion of the results of a survey on O-RAN distributed by the authors.

Ramezanpour and Jagannath [25] introduced intelligent zero-trust architecture for use in next-generation communications networks, allowing for the implementation of AI engines to provide security functions in untrusted networks. Groen et al. [26] provided an investigation into the impact of implementing encryption on the E2 interface on delay and throughput. In [27], Shen et al. proposed a security threat analysis and treatment strategy system, and provided an example test to verify whether or not the SMO correctly authenticates a Near-RT RIC node.

Mimran et al. [11] proposed an ontology for security evaluation of O-RAN, and discussed the current state of O-RAN security, providing a taxonomy and a map of relevant cybersecurity threats. Soltani et al. [28] discussed a risk assessment carried out by the O-RAN Alliance Security Focus Group (SFG), AI threats against O-RAN including data poisoning, evasion attacks, and attacks on API-based AI models. This article also discussed the potential for security countermeasures in O-RAN.

Dik and Berger [29] provided an analysis of the vulnerabilities of each data plane of the Open Fronthaul interface under a Man-in-the-Middle attack, and discussed the suitability of MACsec to secure each plane of this interface. Liao et al. [30] described a tool they developed to carry out denial-of-service attacks on the C-Plane of the Open Fronthaul interface for testing purposes in line with guidance from the O-RAN Alliance Test and Integration Focus Group. In [31], Haas et al. have proposed hardware-enforced capabilities as an enabling technology of security in future O-RAN networks, and presented an approach for hardware/operating system co-design to implement these capabilities. In [32], Groen et al. utilised the Colosseum radio frequency (RF) emulator to investigate the performance cost of implementing security features on the E2 interface.

Rahman et al. [33] described potential artificial intelligence attacks against future 6G networks, and discussed enabling technologies for 6G (including O-RAN) and security threats. Giupponi and Wilhelmi [34] proposed integrating blockchain technologies in O-RAN for RAN-sharing, and provided an example O-RAN-based architecture for utilising this technology. Huang et al. [35] investigated the detection of rogue base stations in a software-defined radio (SDR)-enabled O-RAN environment using data generated by the UE, and an xApp carrying out machine-learning techniques on this data. Motalleb et al. [36] proposed a method for energy-efficient service admission control in O-RAN using deep reinforcement learning techniques, and a moving target defence strat-

egy to secure this service admission control. These works are instrumental in understanding the consequences of incorrectly configured networks as well as the requirement of having security as a part of the architecture to ensure services and users are secured from bad actors.

Table 1. A summary of related works.

Article	Theme	Parameters
Liyanage et al. (2023) [9]	O-RAN security	–
Polese et al. (2023) [1]	O-RAN security and O-RAN overview	–
Abdalla and Marojevic (2023) [20]	O-RAN security	Open Fronthaul decrypted packets, simulation time
Soltani et al. (2023) [21]	O-RAN security	signaling cost, average bearer migration rate, throughput, time, ARP request, packet loss rate
Groen et al. (2023) [26]	O-RAN security	Delay, packet size, actual throughput, attempted transmission rate, encryption algorithm, throughput, ratio of different outputs, noise standard deviation, euclidean distance
Groen et al. (2023) [32]	O-RAN security	Probability, delay, packet type, packet size, actual throughput, attempted transmission rate
Huang et al. (2023) [35]	O-RAN security	Signal strength, classifier, accuracy, precision, recall, F1-measure
Motalleb et al. (2023) [36]	O-RAN security	Mean reward, episode, service admission rate, normalized power consumption, service arrival rate, number of VNFs, extra power consumption
Gaur et al. (2023) [13]	Abusive adversary	Corrections to the sensor readings, the time elapsed for a successful attack, the difference of spoofed values compared to the original sensor readings, instances over the attack duration
Klement et al. (2022) [22]	O-RAN security	–
Wen et al. (2022) [23]	O-RAN security	–
Abdalla et al. (2022) [24]	O-RAN security and O-RAN overview	Community survey results
Ramezanpour and Jagannath (2022) [25]	O-RAN security	–
Shen et al. (2022) [27]	O-RAN security	–
Mimran et al. (2022) [11]	O-RAN security	–
Soltani et al. (2022) [28]	O-RAN security	–
Liao et al. (2022) [30]	O-RAN security	Data rate of U-plane message reception, attack rate
Haas et al. (2022) [31]	O-RAN security	Latency
Rahman et al. (2022) [33]	O-RAN security	–
Giupponi and Wilhelmi (2022) [34]	O-RAN security	Sharing mechanism performance, UE request rate, number of operators, overhead, block size
Sharma et al. (2022) [14]	Abusive adversary	Time required by an adversary to replicate states with brute-force, number of behaviour rules, Signaling overheads at unit message size, number of hops failed towards core
Sharma et al. (2022) [16]	Abusive adversary	Relative rewards to the adversary, probability of induced fork, probability of delay, probability of block generation by an adversary, compliance degree of a mining pool
Dik and Berger (2021) [29]	O-RAN security	Payload overhead, payload size

Furthermore, in the direction of the abusive adversary, the authors in [13] modelled attacks carried out by an abusive adversary on a supervisory data and control (SCADA) system consisting of an on-site system and a control server. Four attack strategies were presented involving spoofing sensor values relayed to the control server. In their work, the abusive adversary attacks stealthily and attempts to evade detection, and is able to steer the corrective action taken by the actuator. In another work by Sharma et al. [14], abusive adversaries in the context of 5G-IoT networks were elaborated, in which an abusive adversary attempts to compromise network nodes or the security functions of the 5G network. These works were inspired by other work of Sharma et al. [16], where the authors discussed the impact of self-defying adversaries utilizing zero expectation-based reward abuse in the context of blockchain systems.

3 Network Modelling with Abusive Adversaries

This article considers O-RAN architecture, as discussed in the initial section, with two critical components − Near-RT RIC and Non-RT RIC. The system model considered is based on understanding the performance and adversarial impact when the adversary is operating under an 'Abusive' ideology, as highlighted in [14,16]. In this case, the adversary is considered to be operating as an insider threat, specifically having insights into implementing the RIC. The scenario targeted by the adversary is the resource utilisation and offloading where the adversary intentionally starts a race condition to cause deadlocks, thereby impacting the performance of the system as it increases the attack surface by identifying the list of components that can be compromised when the traffic is offloaded which could be between the near Near-RT RIC components or the control messages at the Non-RT RIC. The article details the implementation of MDP to let the adversary operate with this ideology as in [16–18], and then utilise prediction to reverse engineer the state that could have led to the attack on the network under the said scenario of resource allocation and data offloading. This could be further supported by FIS to understand the probabilities and their impact as expressed in Sect. 3.2.

To examine this further, consider a Near-RT RIC component with a CU that manages several DUs, and each DU can handle RUs that manage several UEs. *To help understand the scenario*[1], consider a specific resource consumption case where a set K of servers are supporting a set M of UEs depending on the capacity and offloading rate along with the priority of the application. A specific network implementation may allow a single RU to control multiple servers amongst the given set K. Each server, k_1, k_2, $\cdots$, k_i can handle m_k number of UEs. Thus, the total requests handled by the $|K|$ servers will be given as,

[1] The generic settings in the system model are used to understand the problem, whereas, in real settings, the number of UEs can be determined based on the measuring reports but cannot be fixed.

$$R_{T,K} = \sum_{i=1}^{|K|} (m_k)_i \leq |M|, \tag{1}$$

which implies each UE generates a single request; thus, the requests handled by each server are, in general, equal to the number of UEs in its periphery or zone as configured in the planning phase of the network. Thus, if the number of UEs that the server can accommodate increases beyond m_k, there is a need to balance the load and check for resource utilisation - and this is where the adversary can target the system and lead to an attack, such as denial of service (DoS).

3.1 Resource Utilisation and Offloading

This work uses the resource utilisation and offloading model in [37] to decide how to share the tasks amongst the network. Specifically, in this work, the mobility of UEs is considered across the RUs. It is also considered that on-demand RUs can be utilised when the number of UEs exceeds the desired limit. Alongside this, the offloading will depend on the type of the application as it significantly impacts resource utilisation, and over-utilisation may lead to network blackout, which would be the intentional target for the abusive adversary. Thus, the work considered offloading on fixed RUs and on-demand RUs by checking the capacity along with the resource utilisation time as in [37], with resource utilisation time expressed as:

$$T_{F,R} = \frac{\eta \times \alpha}{\beta}, \tag{2}$$

where η is the number of applications (can be at bit level as in [37]) to be transferred, α is the cycles per application and β is the cycles per second. Here, if the required utilisation time, $T_{R,T} \geq T_{F,R}$, the UEs can be moved across the RUs depending on the order of required computations.

Considering an adversarial scenario, an adversary will attempt to predict and evaluate the requirements of offloading across the network and then identify the RUs that are in maximum demand. The adversary can use this information to generate multiple fake requests to launch a DoS attack across the Near-RT RIC. To identify such an impact, we first need to understand if such a scenario is encountered, what the associated implications are, and how the network would respond to prevent malicious intent. To further understand this, we design an abusive adversary inspired by [16], which has sufficient knowledge of the network or has the examining ability of the following components:

- UEs' joining and leaving rate (λ)
- Requests handled ($R_{T,K}$)
- Available servers ($|K|$)
- RU switching and deployment time (ϑ)
- Resource utilisation time ($T_{F,R}$)

Based on these metrics, the adversary can calculate a reward and launch an attack on the network without getting detected. These can be determined as action and no action phases for the adversary that are denoted by α_1 and α_2, respectively. The associated actions for these phases feed into the MDP and include

- No offloading (No (N)): This action is for the scenario where the available servers can handle the requests at the RU, and no switching of the requests is needed. This action also represents the situations where requests can be handled even after the new RUs are added to the network or additional resources are made available, but offloading is not required.
- Offload (Off (Y)): This action represents the cases where services from the UEs are offloaded to different servers or RUs. An exemplary case can be that of handover or when the number of UEs increases more than the expected number (such as during a sports event). All the available states drive this action.
- Rollback (Roll (R)): There is a likelihood of having several on-demand RUs, which will help RIC to make intelligent decisions on resource utilisation. Now, this can lead to managing and saving a lot of resources during idle time or when the demand drops considerably. In certain cases, some of the RUs can be run to their capacity to manage network utilisation. In such cases, the rollback of the offloaded services is considered. Rollback is the specific action of the offload - where the decisions are not to be taken based on the computations; rather, the previous state is attained and continued until the desired configuration is hit

The adversarial system relies on MDP as expressed earlier, which can be defined as MDP $= \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$, denoting the space, action, probability and reward, respectively. The state change is defined based on three out of five components where Δ refers to the adjustments for λ, $R_{T,K}$, $|K|$ and the final values after the variations are denoted with λ', $R'_{T,K}$, and $|K|'$. For the sake of simplicity, the MDP, in our case, considers static for ϑ and $T_{F,R}$, and transitions are referred to as $(\lambda, R_{T,K}, |K|, \vartheta, T_{F,R}, .) \longrightarrow (\lambda \pm \Delta\lambda, R_{T,K} \pm \Delta R_{T,K}, |K| \pm \Delta|K|, \vartheta, T_{F,R}, \langle \alpha_1 | \alpha_2 \rangle)$. These can be examined by the following Action $\times$ Space matrix in Table 2 using the MDP modelling in [16], [17] and [18]. The reward, $Ar_{<state>}$, for the adversary is calculated as the latency factor induced by the adversary, and it would vary depending on the choice of attack by the abusive adversary. For example, in the case of DoS, the adversary would be concerned about not letting ACKs reach either party and allowing consistent traffic generation, thus choking the network channel. If $\mathcal{O}_f$ is the service offloading denoted by the triplet as $\mathcal{O}_f(\lambda, R_{T,K}, |K|)$, then the latency is defined in time to handle the adversarial requests and general network requests, denoted by $\tau_{A,R}(\mathcal{O}_f)$ and $\tau_{G,R}(\mathcal{O}_f)$, respectively. The adversary's aim may be to predict service offloading (p) to identify the slots when the attacks would have the maximum impact. This implies the time to handle the adversarial requests would offer more control on the network and a higher probability of offloading and associated rewards, which the adversary would aim to maximise and given as:

$$Ar_{<state>} = \left\{ \begin{array}{l} \max\left(\frac{\tau_{G,R}(\mathcal{O}_f)-\tau_{A,R}(\mathcal{O}_f)}{\tau_{G,R}(\mathcal{O}_f)}\right), \tau_{G,R}(\mathcal{O}_f) > \tau_{A,R}(\mathcal{O}_f) \\ 0, \text{otherwise} \end{array} \right\} \quad (3)$$

If the adversary is able to inject devices into the network and is able to receive ACK from the RU, the requirements of prediction become negligible, and so as the rewards would lean towards a minimisation problem of the time required to handle the adversarial requests, given as $\min(\tau_{A,R}(\mathcal{O}_f))$. This can maximise the chances for the adversary to launch the attack due to increased device time in the network.

Table 2. An overview of the MDP process for the possible state x action space along with transitions, probability and rewards along with fuzzy inference system. Here, the probability considers attack (α_1) or no-attack phase (α_2), and rewards are for adversarial actions associated with the attack phase (α_1).

$\mathcal{S} \times \mathcal{A}$	Next State	$\mathcal{P}$	$\mathcal{R}$ × F-Output
$(\lambda, R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, .) \times N$	$(\lambda - \Delta\lambda, R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_0(.)$	0 x Low
	$(\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_1(.)$	0 x Low
	$(\lambda - \Delta\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_2(.)$	0 x Low
	$(\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_3(.)$	0 x Low
	$(\lambda, R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_4(.)$	0 x Low
$(\lambda', R'_{T,K}, \lvert K\rvert', \vartheta, T_{F,R}, \alpha_2) \times N$	$(\lambda + \Delta\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert - \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_1)$	$P_0(\alpha_1)$	Ar_0 x Very High
	$(\lambda + \Delta\lambda, R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_1)$	$P_1(\alpha_1)$	Ar_1 x High
	$(\lambda + \Delta\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_1)$	$P_2(\alpha_1)$	Ar_2 x High
	$(\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert - \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_1)$	$P_3(\alpha_1)$	Ar_3 x Very High
	$(\lambda, R_{T,K}, \lvert K\rvert - \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_1)$	$P_4(\alpha_1)$	Ar_4 x High
$(\lambda', R'_{T,K}, \lvert K\rvert', \vartheta, T_{F,R}, \alpha_1) \times Y$	$(\lambda - \Delta\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_0(\alpha_2)$	0 x Low
	$(\lambda - \Delta\lambda, R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_1(\alpha_2)$	0 x Low
	$(\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_2(\alpha_2)$	0 x Low
	$(\lambda, R_{T,K} + \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_3(\alpha_2)$	0 x Low
	$(\lambda, R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_4(\alpha_2)$	0 x Low
$(\lambda', R'_{T,K}, \lvert K\rvert', \vartheta, T_{F,R}, \alpha_2) \times R$	$(\lambda - \Delta\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_0(\alpha_2)$	0 x Neutral
	$(\lambda - \Delta\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_1(\alpha_2)$	0 x Neutral
	$(\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_2(\alpha_2)$	0 x Neutral
	$(\lambda, R_{T,K} - \Delta R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_3(\alpha_2)$	0 x Neutral
	$(\lambda - \Delta\lambda, R_{T,K}, \lvert K\rvert + \Delta\lvert K\rvert, \vartheta, T_{F,R}, \alpha_2)$	$P_4(\alpha_2)$	0 x Neutral

3.2 Probabilities and Impact Modelling

Now, considering the reverse engineering component, the adversary can have a range of rewards and the probability to utilise the knowledge of the network to identify state variables and vice versa to launch a DoS attack. This attack can be done on several low-cost IoT, which will shift the paradigm of the network services. An example is that a new device included over the xApps, which was unplanned during the O-RAN configuration, now dominates the number of requests over the network. Here, an adversary can predict the next state to enter and launch a strategic attack.

The expectation from such modelling is that once the network has a similar threat model for Near-RT RIC, vendors can set a range for fluctuation in each of the values of components to identify when the potential risk of an attack in

the network is, especially when the data is not acknowledged by the DU and CU based on what is expected from the RUs.

Probabilities in MDP define the possibilities of the transition between the states and earning rewards for the adversary, which will help it to decide the most impactful attack strategy in O-RAN. This will depend on the prediction rate, which means if the time to predict the next state is low, then the probability of switching states will be high. In general, non-zero values for $\Delta\lambda$, $\Delta R_{T,K}$ and $\Delta|K|$ would mean high attack surfaces if the adversary can learn about these settings. In the current state, the probability is to be determined for three actions, and the model considers a probability of dominance for the three accountable parameters denoted by ζ_1, ζ_2 and ζ_3, for λ, $R_{T,K}$ and $|K|$, respectively. In the given settings, $\zeta_1 \geq \zeta_2 \geq \zeta_3$, which relates the request handling capacity to the number of available servers, and $\zeta_1 + \zeta_2 + \zeta_3 = 1$. This defines the order of probability in which the adversary can attain and launch a successful attack by considering the present and next feasible state in the network. Using this, the probabilities in Table 2, for three sets of State $\times$ Action can be put in an order for possible probabilities, which will be defined as $P_0 = \zeta_1$, $P_1 = \zeta_1.(1 - \zeta_3)$, $P_2 = \zeta_3.(1 - \zeta_1)$, $P_3 = \zeta_2.\zeta_3$ and $P_4 = \zeta_3.\zeta_1$. For example, consider a scenario where $\zeta_1 > \zeta_2 \geq \zeta_3$, based on which the probabilities will follow $P_0(.) > P_1(.) > P_2(.) \geq P_4(.) > P_3(.)$. Similarly, if $\zeta_3 > \zeta_1 > \zeta_2$, the probability order will be be $P_2(.) > P_0(.) > P_4(.) > P_3(.) > P_1(.)$.

However, it is challenging to identify all such combinations – be it the adversary or a mitigation approach, and inferencing needs to be as close as possible to understand the impact of the adversary. Here, several approaches, like the approximation techniques, can be applied. An alternative can be in the form of a fuzzy-inference engine, which allows converting a possible range of inputs for ζ_1, ζ_2, and ζ_3, to be converted into a particular probability value based on the rules obtained by converting the next state in Table 2. For the sake of simplicity and general applicability to understand adversarial impact, this work considers type-1 Mamdani FIS, which is built using MATLAB^{TM}considering Gaussian Distribution considering that discrete values of Poisson distribution for the users can be translated into continuous data for adversary when it operates for a longer duration in the network.

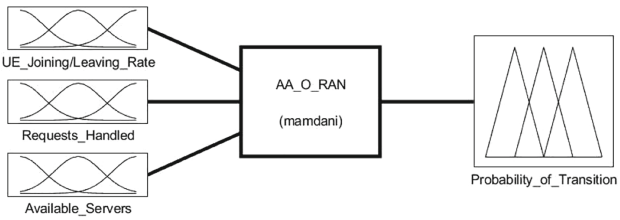

Fig. 3. An illustration of the FIS for an abusive adversary with considered network settings.

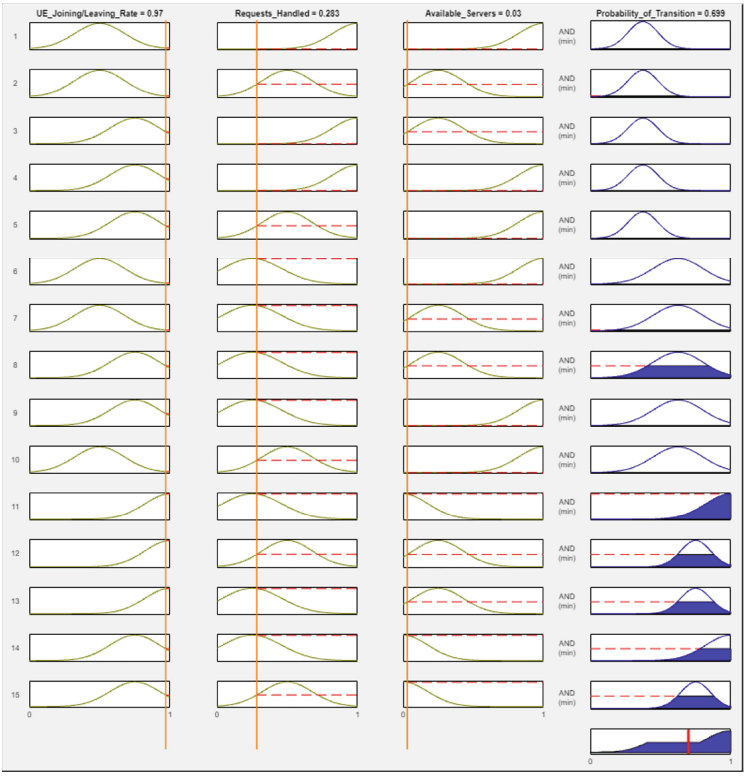

Fig. 4. An exemplary illustration of the attack phase with fuzzy rules for an abusive adversary with considered network settings based on Table 2.

The FIS is shown in Fig. 3, and the associated rules generated using the MDP modelling are illustrated in Fig. 4. This figure helps to understand the probabilistic combinations that would allow an adversary to plan its next set of actions, whether to attack or not, and understand the negative impact it can create on the network. Understanding the abusive adversary rules allows network operators to develop an intelligent intrusion detection system that can secure the network against cyber vandalism. Further details on the impact and associated probability of attack with transitions are shown in Figs. 5, 6 and 7. These graphs help visualise the rules in Fig. 4 and understand the co-relations between the independent and dependent variables when plotted for ζ_1, ζ_3, and ζ_3. It is worth noting that the rules would change depending on the probability of dominance for variables in state $\times$ action space. The model helps to associate probabilities to available discrete values, which can then be extended into a continuous range of values, which are then evaluated for scenarios and impact on the system.

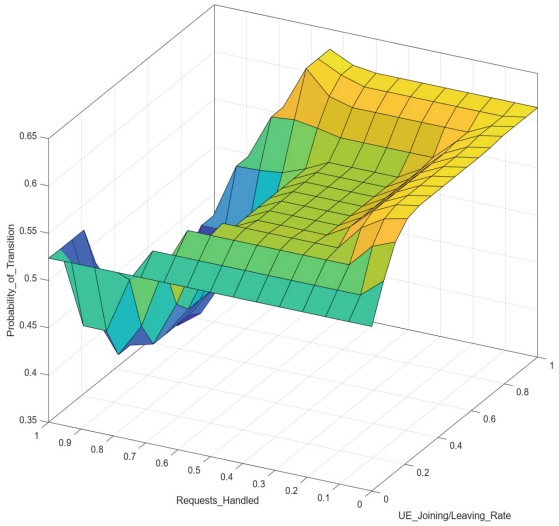

Fig. 5. A comparative view of the rules showcasing the probability of transition leading to an attack using user joining and leaving rate (ζ_1) vs requests handled (ζ_2).

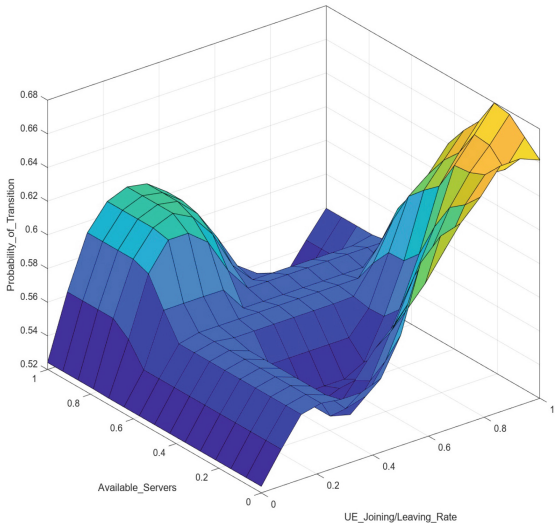

Fig. 6. A comparative view of the rules showcasing the probability of transition leading to an attack using user joining and leaving rate (ζ_1) vs available servers (ζ_3).

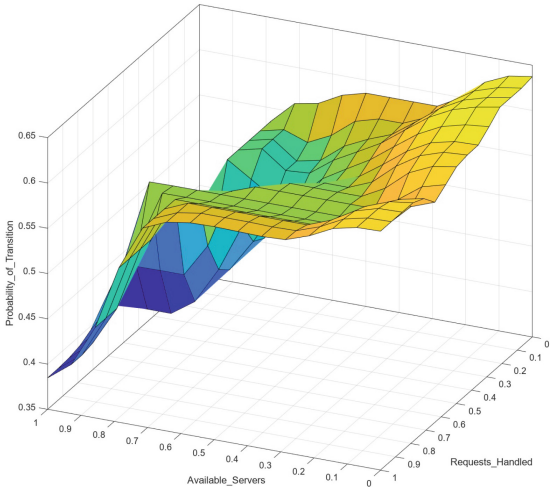

Fig. 7. A comparative view of the rules showcasing the probability of transition leading to an attack using available servers (ζ_3) vs requests handled (ζ_2).

3.3 Impact Evaluations

This section details the impact an abusive adversary can have specifically when several attack surfaces are available in the O-RAN settings. The article expresses results by using near-identical traffic as expected over O-RAN without deploying specific control sequences. Though implementing similar work using OpenRAN Gym could be more informative, it is beyond the scope of this work in its current form. Other details on traffic generation, adversarial scenarios with impacts and mitigation strategies are expressed below:

- Traffic and Scenario Generation: The traffic is generated considering a set of M UEs uniformly distributed over the simulated area. This area is served by $|K| = 4$ BSs, each placed at the corner of that area. Without loss of generality, we assumed that all the users within the are generating connection requests toward the same server. For simplicity and without loss of generality, each user is assumed to send a service request using the nearest BSs. The request inter-arrival times (considered equivalent to the resource utilisation time) and the service time at each BS are exponentially distributed. In addition, we assumed that each BS can accept a certain number of connection requests between 10 and 25. Service requests are denied to users when the queue is full. If so, users re-schedule service requests after a back-off time, which follows the same distribution of the inter-arrival times. Simulations have been carried out using the open-source discrete-event network simulator NS3 [38] by varying the number of users and request rate, as shown in Table 3.

The traffic available from NS3 is then customised to a synthetic dataset to match the requirements of running an MDP in Table 2 and identifying associated

Table 3. Simulation parameters for generating the traffic scenario for impact evaluation of abusive adversary.

Parameter	Value
Distance between BS [m]	500
Number of UEs	[20, 50, 100]
Average inter-arrival Times [s]	[1, 2, 5]
Resource utilisation time [s]	60
Queue Size at BS	[10; 25]

attack probability from FIS as in Fig. 3, and deliberately induce an adversary. Here, $\tau_{A,R}(\mathcal{O}_f)$ is calculated as a difference between the average of the request start time and finish time for a general UE, considering that similar instances would be required for an adversary to inject its device into the network. The total requests are calculated as $(\max(\text{Queue size}) + (\max(\text{Queue size}) \times (\lambda[1, 5])))$. The handled requests are calculated as the difference between max(Queue size) and the average drop of requests, which is 11 in the case of scenario with 100 UEs, which leads to the calculation of ζ_2 as the ratio between the handled requests and the total requests, and ζ_1 is calculated as the ratio of the leaving and joining rate of UEs and the sum of the leaving and joining rate of UEs and current UEs in the network; and injected five adversarial data. In the final readings for the probability of transition from the FIS, the points of local maxima are used for every membership function.

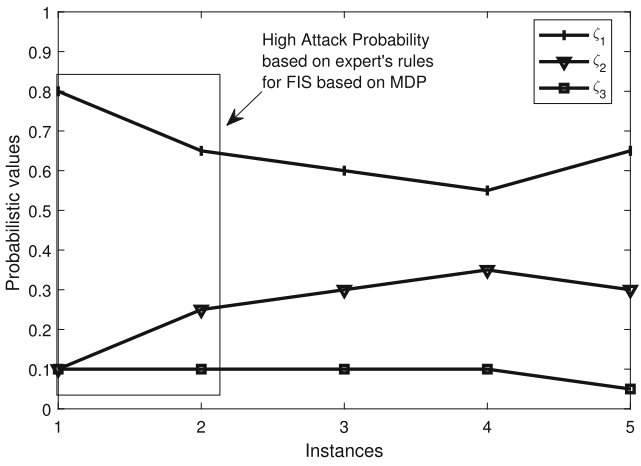

Fig. 8. An exemplary illustration of the attack scenarios selection based on expert's inputs as rules into the FIS.

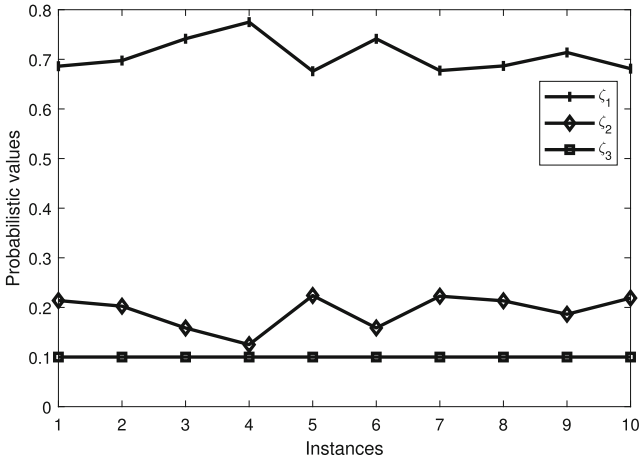

Fig. 9. An exemplary illustration of the attack scenarios selection available to adversary after executing a SMOTE model and generating possible settings of ζ_1, ζ_2 and ζ_3 that can lead to reward.

Once the adversary has the preliminary set of values for the parameters it is trying to utilise to identify attack scenarios, it can identify values which will yield maximum reward and calculate the associated probability of transition. Alternatively, an adversary can deploy the known probability of transition along with partial instances from the network, which can obtained via measurement reports to generate data via models such as SMOTE [19]. Figure 8 shows the settings that allow adversaries to have a high probability of attack on the network based on expert inputs. Figure 9 shows how an adversary can utilise partial knowledge of the network and identify settings it needs to attain to have a scenario most suited for an attack based on the combination of MDP, FIS and SMOTE.

- Mitigating Abusive Ideology: Executing the adversarial mode to regenerate the associated set of values for ζ_1, ζ_2 or ζ_3 for a set of probability can lead network managers to identify potential scope of attack and develop mitigation strategies. These modes must be run whenever major offloading activities are initiated in the network and checked for security compromises and attack surfaces. These modes can be executed in parallel or in the background system. Additional measures to be adopted are listed below:

– Having a pre-determined offloading mechanism must be avoided, and decisions must be taken by the entities involved in the offloading rather than the centralised node.
– Controlling ϑ and $T_{F,R}$ to ensure that the switching time is lower than the service time for the minimum application.

- Intelligent systems can be implanted alongside the Near-RT RIC that can run such strategic scenarios to identify the probability of attack based on service classification.
- Servers must never be run to the maximum capacity, and the idle phase must be planned using intelligent algorithms that can pick any misconfiguration in the system.
- Stateless computing could be another potential avenue to consider to move away from the threats posed by the legacy systems in the Near-RT RIC and Non-RT RIC when interoperability between xApps is expected to impact the network configurations significantly.

4 Conclusions

This article presented a background understanding of the O-RAN and its security, followed by some prominent literature examining current security standards and approaches. Following this, a particular case of abusive modelling is considered where an abusive adversary is designed to understand the impact of Near-RT RIC. The idea of having such an adversary is to offer a strict threat model which can check the security mechanisms of the network in the offloading scenario. The article used Markov Decision Process (MDP) and a Fuzzy Inference System (FIS), which uses SMOTE to generate a set of metrics that can offer a high probability of attack in the transition mode to the adversary. This methodology provides a viewpoint on how an adversary forms predictive methods on when to attack the system, which is followed by mitigation mechanisms for the network to avoid it from happening. In future, we aim to consider parameters from the live measurement report and formation of the MDP in real-time with different traffic scenarios to examine security threats in O-RAN.

Acknowledgement. This work is partly supported by the UK Department for Science, Innovation and Technology under the Future Open Networks Research Challenge project TUDOR (Towards Ubiquitous 3D Open Resilient Network) and partly by NICYBER2025. The views expressed are those of the authors and do not necessarily represent the project or the funding agency.

References

1. Polese, M., Bonati, L., D'Oro, S., Basagni, S., Melodia, T.: Understanding O-RAN: architecture, interfaces, algorithms, security, and research challenges. IEEE Commun. Surv. Tutor. **25**(2), 1376–1411 (2023)
2. Wypiór, D., Klinkowski, M., Michalski, I.: Open RAN - radio access network evolution, benefits and market trends. Appl. Sci. **12**(1), 408 (2022)
3. O-RAN Alliance: O-RAN use cases and deployment scenarios, white paper, February 2020. https://www.o-ran.org/resources. Accessed Nov 2023
4. O-RAN Working Group 1: O-RAN architecture Description, June 2023. https://specifications.o-ran.org/specifications. Accessed Nov 2023

5. Masaracchia, A., Sharma, V., Fahim, M., Dobre, O.A., Duong, T.Q.: Digital twin for open RAN: towards intelligent and resilient 6G radio access networks. IEEE Commun. Mag. **61**, 112–118 (2023)
6. 3GPP: "3GPP TS 23.501 V18.3.0," technical specification, September 2023. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144. Accessed Nov 2023
7. 3GPP: "3GPP TS 33.501 v18.3.0," technical specification, September 2023. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169. Accessed Nov 2023
8. Penttinen, J.: 5G security evolves, May 2021. https://www.gsma.com/security/wp-content/uploads/2021/06/T-ISAC_5G-Security_PenttinenJ-GSMA-public.pdf. Accessed Nov 2023
9. Liyanage, M., Braeken, A., Shahabuddin, S., Ranaweera, P.: Open RAN security: challenges and opportunities. J. Netw. Comput. Appl. **214**, 103621 (2023)
10. O-RAN Working Group 11: O-RAN security threat modeling and remediation analysis, June 2023. https://orandownloadsweb.azurewebsites.net/specifications. Accessed Nov 2023
11. Mimran, D., et al.: Security of open radio access networks. Comput. Secur. **122**, 102890 (2022)
12. Sharma, V., You, I., Guizani, N.: Security of 5G–V2X: technologies, standardization, and research directions. IEEE Netw. **34**(5), 306–314 (2020)
13. Gaur, V.S., Sharma, V., McAllister, J.: Abusive adversarial agents and attack strategies in cyber-physical systems. CAAI Trans. Intell. Technol. **8**(1), 149–165 (2023)
14. Sharma, V., Varghese, B., McAllister, J., Mohanty, S.P.: Abusive adversaries in 5G and beyond IoT. IEEE Consum. Electron. Mag. **11**(4), 11–20 (2022)
15. Oxford English Dictionary: Cyberterrorism. Oxford University Press, Oxford (2023)
16. Sharma, V., Szalachowski, P., Zhou, J.: Evaluating blockchain protocols with abusive modeling. In: Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security, ASIA CCS 2022, pp. 109–122, Association for Computing Machinery, New York (2022)
17. Sapirshtein, A., Sompolinsky, Y., Zohar, A.: Optimal selfish mining strategies in bitcoin. In: Grossklags, J., Preneel, B. (eds.) FC 2016. LNCS, vol. 9603, pp. 515–532. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54970-4_30
18. Gervais, A., Karame, G.O., Wüst, K., Glykantzis, V., Ritzdorf, H., Capkun, S.: On the security and performance of proof of work blockchains. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 3–16 (2016)
19. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
20. Abdalla, A.S., Marojevic, V.: End-to-end O-RAN security architecture, threat surface, coverage, and the case of the open fronthaul, arXiv preprint arXiv:2304.05513 (2023)
21. Soltani, S., Shojafar, M., Brighente, A., Conti, M., Tafazolli, R.: Poisoning bearer context migration in O-RAN 5G network. IEEE Wirel. Commun. Lett. **12**(3), 401–405 (2023)
22. Klement, F., et al.: Open or not open: are conventional radio access networks more secure and trustworthy than Open-RAN?, arXiv preprint arXiv:2204.12227 (2022)

23. Wen, H., Porras, P., Yegneswaran, V., Lin, Z.: A fine-grained telemetry stream for security services in 5G open radio access networks. In: Proceedings of the 1st International Workshop on Emerging Topics in Wireless, pp. 18–23 (2022)
24. Abdalla, A.S., Upadhyaya, P.S., Shah, V.K., Marojevic, V.: Toward next generation open radio access networks: what O-RAN can and cannot do! IEEE Netw. **36**(6), 206–213 (2022)
25. Ramezanpour, K., Jagannath, J.: Intelligent zero trust architecture for 5G/6G networks: principles, challenges, and the role of machine learning in the context of O-RAN. Comput. Netw. **217**, 109358 (2022)
26. Groen, J., et al.: Implementing and evaluating security in O-RAN: interfaces, intelligence, and platforms, arXiv preprint arXiv:2304.11125 (2023)
27. Shen, C., et al.: Security threat analysis and treatment strategy for ORAN. In: 2022 24th International Conference on Advanced Communication Technology (ICACT), pp. 417–422 (2022)
28. Soltani, S., Shojafar, M., Taheri, R., Tafazolli, R.: Can open and AI-enabled 6G RAN be secured? IEEE Consum. Electron. Mag. **11**(6), 11–12 (2022)
29. Dik, D., Berger, M.S.: Transport security considerations for the open-RAN fronthaul. In: 2021 IEEE 4th 5G World Forum (5GWF), pp. 253–258 (2021)
30. Liao, S.-H., Lin, C.-W., Bimo, F.A., Cheng, R.-G.: Development of C-plane DoS attacker for O-RAN FHI. In: Proceedings of the 28th Annual International Conference on Mobile Computing and Networking, pp. 850–852 (2022)
31. Haas, S., et al.: Trustworthy computing for O-RAN: security in a latency-sensitive environment. In: 2022 IEEE Globecom Workshops (GC Wkshps), pp. 826–831 (2022)
32. Groen, J., Kim, B., Chowdhury, K.: The cost of securing O-RAN. In: IEEE International Conference on Communications (ICC) (2023)
33. Rahman, T.F., Abdalla, A.S., Powell, K., W., AlQwider, K., Marojevic, V.: Network and physical layer attacks and countermeasures to AI-enabled 6G O-RAN. arXiv preprint arXiv:2106.02494 (2021)
34. Giupponi, L., Wilhelmi, F.: Blockchain-enabled network sharing for O-RAN in 5G and beyond. IEEE Netw. **36**(4), 218–225 (2022)
35. Huang, J.-H., Cheng, S.-M., Kaliski, R., Hung, C.-F.: Developing xApps for rogue base station detection in SDR-enabled O-RAN. In: IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1–6 (2023)
36. Motalleb, M.K., Benzaïd, C., Taleb, T., Shah-Mansouri, V.: Moving target defense based secured network slicing system in the O-RAN architecture, arXiv preprint arXiv:2309.13444 (2023)
37. Zhou, Y., et al.: Offloading optimization for low-latency secure mobile edge computing systems. IEEE Wirel. Commun. Lett. **9**(4), 480–484 (2019)
38. Piro, G., Baldo, N., Miozzo, M.: An LTE module for the NS-3 network simulator. In: Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques, pp. 415–422 (2011)