Signaling Storm in O-RAN: Challenges and Research Opportunities

Azadeh Tabiban, Hyame Assem Alameddine, Mohammad A. Salahuddin, Raouf Boutaba

***Abstract*—O-RAN enables agile network architecture through programmable disaggregated units. However, the complexity and disaggregation of O-RAN may increase the risk of security incidents. Signaling storm is one of such incidents that disrupts network services through excessive control signals, and is yet unexplored in the context of O-RAN. In this article, we provide a holistic picture of open challenges and opportunities to address signaling storms in O-RAN. Specifically, we discuss different threat models for signaling storm, and explore their applicability to O-RAN. We also survey and categorize existing detection and mitigation solutions. Finally, we provide insights on leveraging key benefits of O-RAN to address signaling storms, and conclude by outlining future directions in this area.**

***Index Terms*—Signaling storm, security, O-RAN, mobile networks, 5G**

I. Introduction

Open Radio Access Network (RAN), and its embodiment O-RAN, embrace the third Generation Partnership Project (3GPP) disaggregation of base stations into cloud-native functions, standardize the interfaces, and introduce intelligence to the management of RAN [1]. Despite their advantages, the new features of O-RAN increase the risk of security incidents that hinder the availability of RAN resources.

Signaling storm is one of such incidents in which attackers increase the intensity of control signals (e.g., the communication overhead of authentication, changing connectivity state, send/receive data, and handovers) to an extent that cannot be handled by the network. For instance, attackers may trigger malicious/compromised User Equipments (UEs) to aggressively attach to the network, i.e., at a rate of thousands of times per hour (as opposed to the standard maximum allowed limit of 20 attachments per hour) [2]. According to the O-RAN Alliance specifications [2], signaling storm is a serious threat to mobile networks. Based on a recent survey [3], around 40% of operators believe that signaling storm raises a significant challenge to 5G and beyond networks. Moreover, the growing demand for various services enabled by 5G and an increasing number of vulnerable UEs, including Internet of Things (IoT) devices that lack sophisticated security mechanisms, render O-RAN more prone to signaling storms.

The naïve solution of increasing network resources temporarily reduces the impact of signaling storms, but the cost of maintaining its effectiveness is unpredictable and depends on the strength of the attack. Most existing works (e.g., [4, 5]) focus on detecting signaling storms based on key performance indicators, such as the number of registration requests, specifically in traditional mobile networks (i.e., prior to O-RAN). However, such works are oblivious to the challenges and opportunities in O-RAN. Similarly, the few existing surveys (e.g., [6]) focus solely on signaling storm in mobile networks prior to O-RAN. Indeed, signaling storm in the context of O-RAN is rather unexplored. The only work that proposes a detection mechanism in O-RAN [7] is based on the number of registration requests, and does not avail the key benefits provided by O-RAN (e.g., integration with Artificial Intelligence/Machine Learning (AI/ML)).

In this article, we conduct a comprehensive study of signaling storms involving the RAN and provide our insights on their applicability in O-RAN. Ultimately, we seek to answer the following research questions:

- What are the threats leading to signaling storms and how are they applicable to O-RAN?
- How effective are existing solutions in addressing signaling storms?
- What are the opportunities introduced by O-RAN that we can leverage to effectively detect and mitigate signaling storms?
- What are the challenges of building an effective solution that can be applied to real-world O-RAN deployment?

II. O-RAN—A Primer

A simplified view of the O-RAN architecture [1] is shown in Fig. 1. O-RAN disaggregates Base Station (BS) functionalities into an O-RAN Central Unit (O-CU), Distributed Unit (O-DU) and Radio Unit (O-RU), which allows the deployment of corresponding functionalities at different locations and on different platforms. For instance, O-DU and O-CU are virtualized and deployed at the edge, while O-RU is typically deployed on field programmable gate arrays and close to the antenna. O-CU manages the life cycle of connections, O-DU is responsible for baseband processing, and O-RU converts radio signals into digital signals [1].

O-RAN connects the disaggregated units to the two O-RAN Intelligent Controllers (RICs), namely near-Real Time (near-RT) and non-Real Time (non-RT) RICs, that leverage AI/ML algorithms to perform management and control of the network at near-RT (i.e., 10 ms to one second) and non-RT (i.e., more than one second) time scales, respectively. The near-RT RIC is deployed at the edge of the network, and enables multiple applications, namely xApps. xApps receive data (e.g., Key Performance Measurements (KPMs), such as load and resource utilization) from E2 nodes (i.e., O-CUs, O-DUs, and O-RAN-compliant Long Term Evolution (LTE) eNodeBs) through E2 interfaces, to compute and provide control actions for different services. xApps deliver specific services such

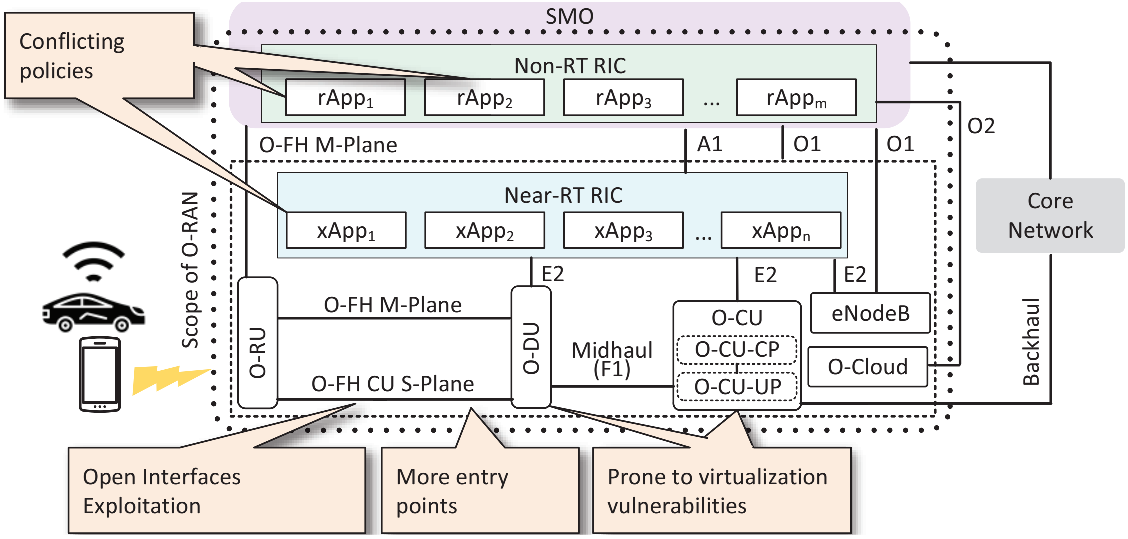


Fig. 1. O-RAN components and vulnerabilities.

as inference, classification, and prediction to optimize users' quality of experience, control handover processes, etc. The non-RT RIC enables another group of applications, namely rApps, and influences the Service Management and Orchestration (SMO) operations. Therefore, it indirectly governs all O-RAN components by making decisions and enforcing policies.

The new paradigm of O-RAN comes with several challenges and opportunities. For instance, although virtualizing O-CU and O-DU enables more agile network deployment, it renders those units prone to threats that generally exist in virtualized environments (e.g., infected images). Moreover, the split functionality of O-RAN enables improved adaptability and resiliency, while it increases the threat surface due to the larger number of interfaces between components that can be exploited as entry points by attackers. Further, as each xApp and rApp is typically designed for specific services, in some cases, they may generate policies that could lead to conflicting requests (cf., Section VII). Finally, as synchronizing O-DU and O-RU requires low latency in the Open-FrontHaul (O-FH), integrity and confidentiality protection via encryption has been removed on corresponding interfaces, which may allow the attacker to impersonate O-DU and O-RU [1].

III. Signaling Storm in Traditional Mobile Networks Prior to O-RAN

Signaling storm typically refers to the overload of control signals that surpasses the available network resources. Radio Resource Control (RRC) is a network layer protocol used between UEs and BSs, which is typically exploited for signaling storm attacks. Therefore, in this section, we introduce the RRC protocol and then explain different signaling storm attacks that exploit its vulnerabilities to target RAN resources.

A. Radio Resource Control and Signaling Storm

RRC protocol manages resources in the RAN by associating a state machine to each UE, where states represent phases during RRC connection establishment and release between a UE and the network. Fig. 2 shows the RRC state machine in a 5G network.

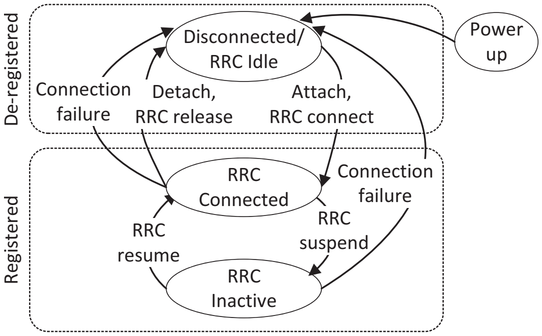


Fig. 2. RRC state machine in 5G networks [8].

In 5G, the RRC state machine consists of three states: RRC_{Idle}, $\text{RRC}_{\text{Connected}}$, and $\text{RRC}_{\text{Inactive}}$ [8]. When a UE is turned on, it is in RRC_{Idle} state, and it subsequently enters $\text{RRC}_{\text{Connected}}$ state to establish radio connection and radio bearers with gNodeB (i.e., O-CU in O-RAN) to send/receive data. Once gNodeB detects that the UE does not send or receive data in a certain time interval (which we refer to as inactivity_timeout), it triggers the UE to enter $\text{RRC}_{\text{Inactive}}$ state, in which the UE and gNodeB store the context of the connection. Therefore, to send/receive data at a later time, the UE transitions from $\text{RRC}_{\text{Inactive}}$ to $\text{RRC}_{\text{Connected}}$ state via only three signaling messages exchanged between the UE and RAN. This is in contrast with LTE networks, where in the absence of $\text{RRC}_{\text{Inactive}}$ state, the UE transitions to RRC_{Idle}, loses connection context and requires seven signaling messages for connection re-establishment.

RRC state machines are exploited by attackers to cause signaling storms through excessive transitioning of UEs from one state to another. While the transition from $\text{RRC}_{\text{Inactive}}$ to $\text{RRC}_{\text{Connected}}$ state generates fewer signaling messages at the RAN, it does not prevent signaling storms in 5G networks [8]. In fact, depending on the threat model, an attacker can force a sufficiently large number of compromised UEs to simultaneously transition from RRC_{Idle} to $\text{RRC}_{\text{Connected}}$ state, or $\text{RRC}_{\text{Inactive}}$ to $\text{RRC}_{\text{Connected}}$ state, or vice versa, to cause a signaling storm.

B. Threat Models

In this section, we introduce different threat models involving the RRC protocol to cause a signaling storm.

UE Impersonation (T1). Attackers can impersonate and de-register subscribed UEs by exploiting the lack of integrity protection of de-registration requests [9], which could inherently lead to a high signaling load upon UE re-registration [9]. To this end, an attacker can first obtain the Globally Unique Temporary User Equipment Identity (GUTI) of many legitimate UEs (e.g., via eavesdropping). GUTI is a unique identifier that is temporarily assigned by the Core Network (CN) to each UE upon registration, which is used for subsequent identification and communication. Therefore, after obtaining the GUTIs, an attacker can send de-registration requests from their botnet by impersonating legitimate UEs. The CN validates the GUTIs and accepts the de-registration requests. Subsequently, all the de-registered UEs will re-initiate the registration process, causing an increase of control signals, which may lead to signaling storms.

False Base Station (T2). Attackers can cause signaling storms by constructing a False Base Station (FBS) using a Universal Software Radio Peripheral device and the 5G protocol stack [9]. Specifically, to launch this attack, attackers first set an FBS with a stronger signal than the one generated by the legitimate BS. This triggers all nearby UEs to detach from the genuine BS and attach to the attacker-controlled BS. The FBS can then broadcast RRC de-registration requests, which will be followed by re-initiating the registration requests from the disconnected UEs. Such simultaneous initiation of registration requests could potentially lead to signaling storms.

Group Handover (T3). Handover is a process in which data transmission is transferred from one BS to another. Simultaneous handover of a large number of moving UEs (e.g., devices on a high-speed train) may cause signaling overload and lead to the disconnection of many UEs due to failed handover. Subsequently, the re-registration of the moving, disconnected UEs may lead to a signaling storm.

Additionally, attackers may intentionally trigger signaling storms by simulating a handover condition for a large number of UEs [10]. To this end, attackers exploit the lack of integrity protection of broadcast messages. Broadcast messages transmit information (e.g., radio resource configurations) that are necessary for a UE to access the mobile network [10]. However, protecting the integrity of most broadcast messages (e.g., using public key infrastructure trust model across different operators) is challenging, which enables the attackers to modify them, indicate that UEs have entered a new tracking area, and should perform handover. To simulate a handover condition in a stealthy manner, an attacker may generate only a subframe of the broadcast message with a strong signal to induce legitimate UEs to read a spoofed Tracking Area Code (TAC). Subsequently, the victim UEs send Tracking Area Update requests to the network for updating their location based on the received TAC. However, as the location of the UE has not changed, the network sends a list of valid TACs, which does not contain the spoofed TAC, and causes UEs to repeat their requests and generate excessive signaling load [10].

UE Rebooting (T4). Signaling storm may also be caused by attackers creating a botnet of compromised UEs (e.g., IoT devices). To this end, attackers infect a large number of vulnerable devices with a remote-reboot malware, and then remotely instruct the malware to simultaneously reboot all devices in a targeted 5G coverage area. This subsequently generates excessive control signals caused by registration requests triggered from all disconnected devices leading to a signaling storm.

Chatty Application (T5). Many mobile applications (e.g., instant messaging applications) periodically send short messages, namely heartbeat messages, to remote servers to indicate their availability. These heartbeat messages cause frequent changes to RRC states by establishing and releasing connections, which can lead to unintentional signaling overload.

Collateral Damage (T6). Signaling storm may also occur as a consequence of other malicious activities, which may not necessarily target the disruption of network services. For instance, many infected UEs (e.g., by SMS/email spammers) generate frequent but small amounts of data, which repeatedly establishes and releases RRC connections. Moreover, network outages are typically followed by a large number of UEs attempting to reconnect after the restoration of the network.

IV. Signaling Storm in O-RAN

Signaling storms can seriously hinder the security of O-RAN, especially considering its disaggregated architecture and relaxed security. In this section, we discuss the realization of signaling storm (cf., Section III-B) in O-RAN with respect to its different properties, and present our preliminary evaluations.

A. Signaling Storm Realization in O-RAN

Additional Signaling (R1). The disaggregation of gNodeB requires additional signaling messages between its disaggregated units in comparison to a monolithic gNodeB to fulfill the requested 5G procedures. For instance, the UE registration procedure involved in all our enumerated threat models (cf., Section III-B) requires additional messages between O-DU and O-CU. Therefore, attackers need fewer UE-initiated registration requests (e.g., through T1 and T4) to generate a sufficiently high signaling load. Moreover, such additional signaling increases the contribution of passive threats (e.g., T5) that can indirectly cause signaling storms.

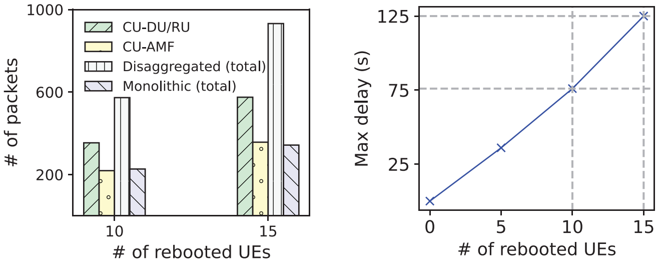
Virtualization Vulnerabilities Exploitation (R2). Attackers may exploit virtualization vulnerabilities to compromise O-DU or O-CU and use them to cause a signaling storm. For instance, a compromised O-DU can be used to de-register (T1) or excessively transition the state of UEs by illegitimately requesting the release of UEs' resources or sending fake UE's inactivity notifications to the O-CU, respectively. Similarly, attackers can launch UE registration and handover procedures by forging RRC messages from a compromised O-DU towards the O-CU to cause a signaling storm.

O-RAN Open Interfaces Exploitation (R3). The limited security measures at the O-FH allows attackers to impersonate O-DUs, and subsequently, initiate de-registration (T1 and T2) or simulate a handover condition by broadcasting a spoofed TAC (T3) for groups of UEs. Impersonating O-DUs also allows attackers to store and delay the received registration requests and simultaneously release them, which can lead to signaling overload. Additionally, attackers may tamper with the information exchanged over E2, A1, and O1 interfaces (Fig.1) to evade detection and mitigation signaling storm xApps and rApps.

O-RU-based False Base Station (R4). In O-RAN, attackers can also cause signaling storms through FBS attacks (T2) launched from compromised O-RU equipment of legitimate vendors. To this end, attackers can either: (i) disable the operational O-RU access to the O-FH and replace it by an FBS, (ii) use a standalone non-operational O-RU, or (iii) gain an unauthorized access to the O-RU and connect it to an FBS system [11].

B. Preliminary Evaluations

To evaluate the impact of disaggregation (R1), we simulate signaling storms by rebooting various numbers of UEs during a small time interval in two versions of Open Air Interface (OAI) testbed (i.e., with monolithic and disaggregated RAN). As Fig. 3a shows, the number of packets exchanged between disaggregated gNodeB units (CU and DU/RU) and the CN is almost 2.5 times those exchanged between gNodeB and the CN for monolithic RAN. Moreover, Fig. 3b shows that the re-registration delay experienced by benign UEs when rebooted within a small time interval before/after the malicious ones may be significant (e.g., 125 seconds for just 15 UEs).



(a)

(b)

Fig. 3. Signaling impact during UE reboot with respect to: (a) number of packets exchanged in disaggregated vs. monolithic RANs, and (b) re-registration delay of UEs.

Note that the above results are specific to OAI, which currently collocates DU and RU. Evaluations on a fully compliant O-RAN testbed is left for future work.

V. Signaling Storm Detection and Mitigation in Traditional Mobile Networks Prior to O-RAN

In this section, we categorize and discuss different existing solutions for detecting and mitigating signaling storms in traditional mobile networks.

A. Detection Mechanisms

Back-off Timer (D1). This mechanism is standardized by 3GPP as the main defense against signaling storms [12]. It is based on trusting the compliance of devices with mobility standards, which restricts repeated registration requests. Nevertheless, attackers may remotely exploit vulnerabilities in a large number of UEs to cause aggressive attachments.

Counter-based (D2). Some of the existing solutions (e.g., [5]) detect suspicious UEs when the number of their bandwidth allocation requests is greater than a threshold n during time interval t_{window}. Typically, the t_{window} is greater than n × inactivity_timeout (cf., Section III-A). This is primarily because after establishing each connection, attackers must wait for inactivity_timeout to expire before re-triggering a new connection and ensuing signals. Therefore, the sequential establishment and release of n connections can be detected throughout the minimum period of n × inactivity_timeout.

Bandwidth-based (D3). There is typically little data that is transmitted between the RAN and a UE contributing to a signaling storm. Therefore, some existing solutions [5] detect suspicious UEs based on the ratio (r_{BW}) between the time that each UE is and is not transmitting data. A threshold (TH) is determined for each UE based on its usual traffic, and a UE is detected to contribute to a signaling storm if the corresponding r_{BW} is smaller than TH.

Data Plane-based (D4). This mechanism detects signaling storms using data plane information collected from packet headers [4]. Signaling storm exhibits certain features (e.g., an increase in the number of destination IP addresses) that can be extracted from packet headers. Relying on such features enables easier deployment compared to other solutions, which require decoding lower radio layers, and handling encryption.

Core Network-based (D5). Attackers may succeed to reach the CN, for example, by generating negligible signaling load at different cell sites to by-pass detection at the RAN. To address this, AI/ML techniques detect signaling storms at the CN based on traffic statistics and identity-based entropy features of subscribers collected at different interfaces of the CN [13].

B. Mitigation Mechanisms

Random Blocking (M1). As the back-off timer at the UEs can be by-passed by attackers (cf., Section V-A), randomly rejecting the incoming registration requests is performed to mitigate the signaling storm attack [2]. However, random rejection of registration requests can lead to denial of service for legitimate UEs.

Blocking Suspicious UEs (M2). Most existing solutions (e.g., [5]) block detected suspicious UEs for a period of t_{block}. Determining the optimum t_{block} requires considering the trade-off between protecting the network and guaranteeing subscribers' satisfaction. This approach may block legitimate UEs that are either compromised or may have accidentally triggered excessive requests.

Configuring inactivity_timeout (M3). Transmitting data through existing connections does not trigger RRC signaling messages. Therefore, increasing the inactivity_timeout of

TABLE I
SUMMARY OF EXISTING SOLUTIONS TO SIGNALING STORM.

#	Mechanism	Brief description	Pros (+) and open issues (-)	Trad. mobile network	O-RAN
S1	Back-off timer [12] (D1, M1)	Tracks the # of registration requests of each UE	+ Direct monitoring of control signals - False positives; reliance on UE compliance to mobility standards	T1, T4-T6	R1
S2	Counter-based [5] (D2, M2)	Tracks the # of same bandwidth allocation requests of each UE	+ Direct monitoring of control signals - False positives	T1, T4-T6	R1
S3	Bandwidth-based [5] (D3, M2)	Tracks the bandwidth usage of each UE	+ Direct monitoring of control signals - False positives	T1, T4-T6	R1
S4	Data plane-based [4] (D4)	Tracks data plane behavior of UEs (e.g., packet size)	+ Easier deployment - Less accurate than control plane-based solutions	T1, T4-T6	R1
S5	Core Network-based [13] (D5)	Based on probabilistic and stochastic techniques	+ Applicable to attacks by-passing the RAN - Effective after CN is already involved	T1-T6	R1-R4
S6	Configuring inactivity timeout [8] (M3)	Modifies the inactivity_timeout once the signaling load increases	+ No direct blocking of registrations - Longer resource allocation	T1, T4-T6	R1
S7	Dynamic res. allocation [14] (M4)	Increases control signals resources	+ No intended blocking of registrations - Reducing data resources	T1-T6	R1-R4
S8	Device-to-Device-based [15] (M5)	Sends heartbeat messages collectively	+ Avoids missing heartbeat messages deadline - Requires code modification	T5	R1

UEs ensures that attackers can trigger RRC signaling less frequently. However, such an approach retains RRC connection resources for a longer time, which increases the overall resource consumption.

To prevent attackers from identifying the precise time of connection release, some existing solutions (e.g., [8]) randomize the inactivity_timeout of UEs. To this end, UEs are assigned with random inactivity_timeout based on the volume of their transmitted traffic.

Dynamic Resource Allocation (M4). Some existing solutions (e.g., [14]) borrow resources from the data channel and allocate them to control channel based on the predicted signaling load. However, this approach may lead to exhausting data resources. The alternative solution of increasing all resources guarantees quality of service, while it may cause an economic denial of sustainability against strong attackers.

Device-to-Device-based (M5). To prevent the contribution of chatty applications to signaling storms, developers modify the code of applications such that applications send heartbeat messages through WiFi Device-to-Device connections to a relay device [15] instead of establishing RRC connection for each heartbeat message. The relay transmits the collected messages to the BS in a single established RRC connection.

C. *Effectiveness of Existing Solutions*

The applicability of existing solutions to different threat models (cf., Section III-B) and their realization in O-RAN (cf., Section IV) are summarized in Table I. Solutions S1-S4 and S6 can render the attack through impersonated (T1) and compromised UEs (T4 and T6) more challenging, as enforcing a blocking time requires the attacker to prepare a new group of malicious UEs for maintaining the high signaling load. Attackers can de-register (T2) or simulate a handover condition (T3) for different groups of UEs at different time intervals, such that each UE generates a negligible load to evade S1-S4 and S6. S5 is applicable to all threats once the attack reaches the CN, and S7 can reduce the impact of all threats by allocating more resources to control signals. In O-RAN, all solutions can partially alleviate R1 by decreasing the signaling load, while S1-S4 and S6 can be evaded by involving different groups of UEs at varying time intervals from compromised O-DU and O-CU (R2 and R3) or O-RU (R4).

D. *Shortcomings of Existing Solutions*

The effectiveness of many existing solutions highly relies on selecting an optimum threshold, which is typically determined either heuristically or analytically based on the rate of transitions to different RRC states. An improper threshold may either lead to high blocking rates of legitimate UEs (i.e., due to false positives) or allowing the progress of the attack (i.e., due to false negatives). For instance, we can infer from our preliminary results (cf., Fig. 3b) that increasing the threshold from 10 to 15 (i.e., blocking registrations upon the reboot of 10 vs. 15 UEs) may increase the re-registration delay of a benign UE from 75 to 125 seconds in our testbed. Further, attackers can generate a negligible signaling load in different areas, which by-passes local detection and mitigation solutions, while the aggregation of the generated signals may overwhelm the CN [2]. Finally, relying on approaches detecting the attack at the CN (e.g., [13]) is not sufficient, as such solutions are effective once the core is already affected by the attack, which may have devastating impact on the entire network. Therefore, a network-wide view prior to the involvement of the CN is essential. In the next sections, we offer our insights on addressing such shortcomings in O-RAN.

VI. SIGNALING STORM DETECTION AND MITIGATION OPPORTUNITIES IN O-RAN

O-RAN Alliance [2] has identified signaling storm as a use case of non-RT and near-RT RICs, and introduced a protection schema against signaling storm (Fig. 4), which is referred to as Signaling Storm Protection Schema (SSPS) in this article. However, the integration and enhancement of the existing solutions as xApps and rApps have not yet been explored.

As Fig. 4 illustrates, the non-RT RIC maintains a network-wide view by probing the CN, collecting information from

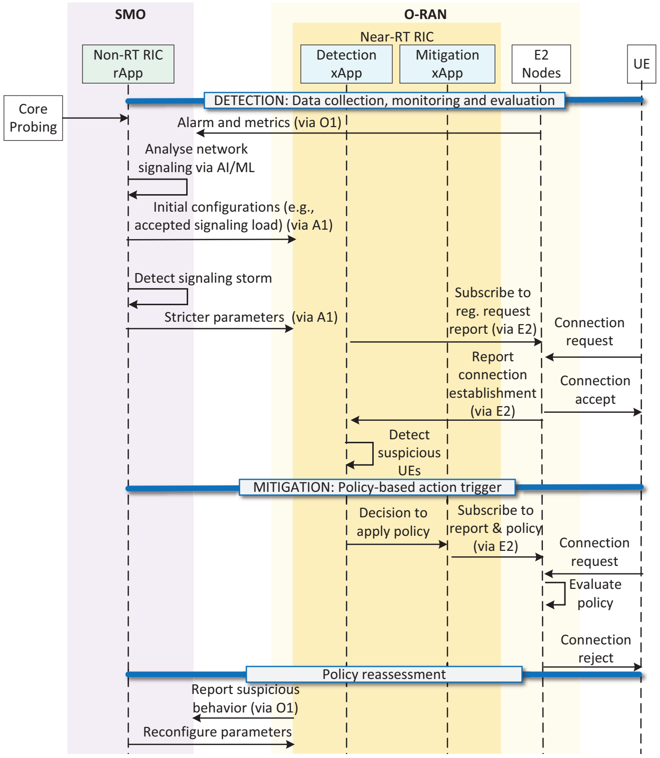

SMO
O-RAN
Near-RT RIC
Non-RT RIC rApp
Detection xApp
Mitigation xApp
E2 Nodes
UE
Core Probing
DETECTION: Data collection, monitoring and evaluation
Alarm and metrics (via O1)
Analyse network signaling via AI/ML
Initial configurations (e.g., accepted signaling load) (via A1)
Detect signaling storm
Stricter parameters (via A1)
Subscribe to reg. request report (via E2)
Connection request
Report connection establishment (via E2)
Connection accept
Detect suspicious UEs
MITIGATION: Policy-based action trigger
Decision to apply policy
Subscribe to report & policy (via E2)
Connection request
Evaluate policy
Connection reject
Policy reassessment
Report suspicious behavior (via O1)
Reconfigure parameters

Fig. 4. The O-RAN signaling storm protection schema (SSPS).

E2 nodes at different geographical locations, and leveraging AI/ML to analyse the signaling behavior of the network. Accordingly, non-RT RIC provides signaling storm detection policies to near-RT RIC by configuring detection and mitigation parameters, which can be modified based on the observed signaling load. Detection xApp monitors the connection establishment requests received at E2 nodes, and informs the mitigation xApp to reject connection requests of the detected suspicious UEs. Finally, the detection xApp reports the suspicious behavior to the SMO. The SMO sends the received reports to the non-RT RIC, which accordingly re-configures detection and mitigation parameters of relevant near-RT RICs. In the following, we discuss the opportunities provided by this schema, O-RAN architecture and enabling technologies for hardening O-RAN against signaling storms.

Easier Deployment. Standardizing security solutions as standalone xApps and rApps will enable faster development and integration of signaling storm xApps and rApps. The latter can leverage their interactions with E2 nodes to easily access the intercepted signaling messages for more efficient solutions to handle signaling storms.

Aggregated Network View. The standardized and open interfaces in O-RAN enable network-wide data exchange between its distributed nodes. This provides signaling storm xApps and rApps with a wide view of the network, which can enhance O-RAN security against signaling storms. For instance, aggregating the data received from xApps at different locations enables the detection of stealthy attackers who may generate signaling load that is spread across the network.

Intelligent Mechanisms. Although [7] proposes an xApp to detect malicious UEs based on the deviation in the number of registration requests from a threshold, it lacks a systematic mechanism for continuous network-wide monitoring to dynamically adjust the threshold based on the network load and behavior of UEs. In O-RAN, the different time granularities of non-RT and near-RT RICs enable intelligent threshold updates based on KPMs from various O-RAN interfaces. For instance, non-RT RIC rApps can dynamically update the threshold by leveraging the alarms and metrics as well as suspicious behavior reported by E2 nodes and near-RT RIC, respectively.

VII. Future Research Directions

To secure O-RAN against signaling storms, future research directions leveraging O-RAN opportunities need to extend but yet be in compliance with the O-RAN SSPS (cf., Fig. 4) and AI/ML workflow [1]. In this section, we highlight our related suggestions.

Practical Study. Although signaling storm attacks are not new, their realization and impact on O-RAN may differ from traditional mobile networks (cf., Section IV). Therefore, a first step towards securing O-RAN against signaling storm attacks is to conduct simulations on an O-RAN testbed and studying their impact on different components (e.g., O-DU, O-CU, and O-RU). However, most existing testbeds are not fully compliant with O-RAN standards.

Early Detection. Most of the existing solutions that are based on RRC messages can be implemented in the O-CU. However, an early detection of this attack is important, especially to harden the network against attackers that generate high signaling load spread across geographically distributed O-RUs and O-DUs. Therefore, it is important to explore the possibility of signaling storm detection at the O-RU and O-DU by correlating their corresponding logs.

RIC Interactions. As shown in Fig. 4, O-RAN Alliance considers the interaction between signaling storm rApps and xApps only for policy enforcement and reassessment through the A1 interface from rApps towards xApps [1, 2]. However, an effective solution can benefit from collaborative and distributed AI/ML techniques to enhance signaling storm detection and mitigation policies. A feedback loop (i.e., from rApps to xApps and vice versa) for reinforcement learning, periodic ML model re-training and policy updates are intrinsic for better security against signaling storms in O-RAN.

Securing O-RAN Interfaces. O-RAN Alliance recently required the support of protocols such as IEEE 802.1x at the O-FH for authentication and authorization, and IPSec and TLS for confidentiality, integrity and replay protection at E2, A1 and O1 interfaces. Media Access Control Security (MACSec) is suggested to secure the O-FH against O-DU impersonation. Operators can configure the bounded_delay field of MACSec to prevent attackers from delaying and collectively releasing registration requests (cf., Section IV, R3). Despite their advantages, these protocols cannot entirely prevent signaling storm attacks exploiting O-RAN interfaces (cf., Section IV, R3), especially considering their latency requirement which is yet to be addressed.

Signaling Storm Prevention. Signaling storm can be prevented by forecasting future RRC requests. The forecasted request patterns can be used in xApps and rApps, which are integrated with the O-RAN SSPS, to conduct an early adjustment of detection and mitigation policies.

Conflict Mitigation. To mitigate signaling storm, potential conflicting policies must be addressed. For instance, two different xApps focusing on maintaining the availability of resources and mitigating signaling storms may request a smaller inactivity_timeout (i.e., to release resources faster) and a longer inactivity_timeout, respectively. It is important to resolve such a conflict by considering the trade-off between signaling load and availability of resources under different network conditions.

VIII. Conclusion

In this article, we discussed signaling storm threat models and solutions in monolithic RAN, and explored their applicability and envisioned impact on O-RAN. Further, we analysed some of the key features of O-RAN, and concluded that O-RAN can provide promising opportunities to build more effective solutions for hardening real-world O-RAN deployments against signaling storms.

Acknowledgments

This research was supported by Ericsson Canada, University of Waterloo, and the Mitacs Accelerate program. We thank Dr. Makan Pourzandi for his insightful comments.

References

[1] M. Polese *et al.*, “Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges,” *IEEE Communications Surveys & Tutorials*, 2023.

[2] “O-RAN Working Group 1, Use Cases Detailed Specification,” *O-RAN.WG1.Use-Cases-Detailed-Specification-v10.00*, 2023.

[3] Cellusys, “Signalling Security - Sizing The Operator Threat,” 2022, accessed March 27, 2023, https://www.cellusys.com/resources/kaleido-security-report-download-from-resources/.

[4] O. H. Abdelrahman *et al.*, “A Data Plane Approach for Detecting Control Plane Anomalies in Mobile Networks,” in *Internet of Things*. Springer, 2016, pp. 210–221.

[5] M. Pavloski, “Detecting and Mitigating Storm Attacks in Mobile Access to the Cloud,” in *ICFC*. IEEE, 2019, pp. 53–58.

[6] M. Q. Khan, “Signaling Storm Problems in 3GPP Mobile Broadband Networks, Causes and Possible Solutions: A Review,” in *iCCECE*, 2018, pp. 183–188.

[7] M. Hoffmann *et al.*, “Signaling Storm Detection in IIoT Network based on the Open RAN Architecture,” *arXiv preprint arXiv:2302.08239*, 2023.

[8] R. Ettiane *et al.*, “Mitigating Denial of Service Signaling Threats in 5G Mobile Networks,” *IJACSA*, vol. 12, no. 2, 2021.

[9] X. Hu *et al.*, “A Systematic Analysis Method for 5G Non-access Stratum Signalling Security,” *IEEE Access*, vol. 7, pp. 125 424–125 441, 2019.

[10] H. Yang *et al.*, “Hiding in Plain Signal: Physical Signal Overshadowing Attack on LTE,” in *USENIX Security Symposium*, 2019, pp. 55–72.

[11] “O-RAN Security Threat Modeling and Remediation Analysis,” *O-RAN.WG11.Threat-Model.O-R003-v06.00.00*, 2023.

[12] 3GPP, “Procedures for the 5G System (5GS); Stage 2,” *3GPP TS 23.502 version 18.0.0 Release 18*, 2022.

[13] S. Park *et al.*, “Machine Learning Based Signaling DDoS Detection System for 5G Stand Alone Core Network,” *Applied Sciences*, vol. 12, no. 23, 2022.

[14] T.-H. Chen *et al.*, “Dynamic Inter-Channel Resource Allocation for Massive M2M Control Signaling Storm Mitigation,” in *IEEE VTC-Fall*, 2016, pp. 1–5.

[15] X. Yi *et al.*, “eDirect: Energy-efficient D2D-assisted relaying framework for cellular signaling reduction,” *IEEE/ACM Transactions on Networking*, vol. 28, pp. 860–873, 2020.

Biographies

Azadeh Tabiban is currently an industrial postdoctoral fellow at the University of Waterloo in collaboration with Ericsson, Canada.

Hyame Assem Alameddine is currently a senior security researcher at Ericsson, Canada.

Mohammad A. Salahuddin is currently a Research Assistant Professor of Computer Science at the University of Waterloo.

Raouf Boutaba is currently a University Chair Professor and the director of the School of Computer Science at the University of Waterloo.