2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Modeling 5G Threat Scenarios for Critical Infrastructure Protection

Gerrit Holtrup
Kudelski IoT Security
Cheseaux-sur-Lausanne, Switzerland
gerrit.holtrup@nagra.com

William Blonay
NATO CCDCOE
Tallinn, Estonia
william.blonay@ccdcoe.org

Martin Strohmeier
armasuisse Science and Technology
Thun, Switzerland
martin.strohmeier@ar.admin.ch

Alain Mermoud
armasuisse Science and Technology
Thun, Switzerland
alain.mermoud@ar.admin.ch

Jean-Pascal Chavanne
Federal Department of Justice and Police
Bern, Switzerland
jean-pascal.chavanne@
isc-ejpd.admin.ch

Vincent Lenders
armasuisse Science and Technology
Thun, Switzerland
vincent.lenders@ar.admin.ch

Abstract: Fifth-generation cellular networks (5G) are currently being deployed by mobile operators around the globe. 5G is an enabler for many use cases and improves security and privacy over 4G and previous network generations. However, as recent security research has revealed, the 5G standard still has technical security weaknesses for attackers to exploit. In addition, the migration from 4G to 5G systems takes place by first deploying 5G solutions in a non-standalone (NSA) manner, where the first step of the 5G deployment is restricted to the new radio aspects of 5G. At the same time, the control of user equipment is still based on 4G protocols; that is, the core network is still the legacy 4G evolved packet core (EPC) network. As a result, many security vulnerabilities of 4G networks are still present in current 5G deployments. To stimulate the discussion about the security risks in current 5G networks, particularly regarding critical infrastructures, we model possible threats according to the STRIDE threat classification model. We derive a risk matrix based on the likelihood and impact of eleven threat scenarios (TS) that affect the radio access and the network core. We estimate that malware or software vulnerabilities on the 5G base station constitute the most impactful threat scenario, though not the most probable. In contrast, a scenario where compromised cryptographic keys threaten communications between

network functions is both highly probable and highly impactful. To improve the 5G security posture, we discuss possible mitigations and security controls. Our analysis is generalizable and does not depend on the specifics of any particular 5G network vendor or operator.

Keywords: *5G, next-generation networks, threat scenarios, critical infrastructures, cyber defense, security*

1. INTRODUCTION

The arrival of the fifth generation of cellular networks (5G) enables new use cases compared to previous mobile telecommunications standards. Examples range from the support of stationary devices in the Internet of Things (IoT) to highly mobile settings in vehicular networks. Power, latency, and data rate requirements vary widely across these different device classes. The introduction of the network slice and network function virtualization concepts in 5G are expected to address these differences in functional requirements.

Currently, the migration from 4G to 5G systems is taking place by first deploying 5G solutions in a non-standalone (NSA) manner, where the first step in 5G deployment is restricted to the new radio aspects of 5G (5G-NR). At the same time, the control of user equipment is still based on 4G protocols; that is, the core network is still the legacy 4G network.

Previously unsolved privacy concerns in 4G are addressed in the 5G standard. Contrary to the previous generation, the analysis of the security of the 5G system, as defined in [1], was already an active concern of researchers before the wide deployment of the standard [2]. A formal analysis of the security procedures by Basin et al. [3] has revealed weaknesses that may potentially still be fixed before 5G standalone systems are deployed.

While previous work focuses on the radio interface, this paper analyzes a full standalone system, including the 5G core network (5GC) architecture [4]. However, given the reality that immediate deployments of 5G in the field is NSA deployments, these will also be covered where appropriate.

We build our security analysis on existing literature focusing on the use of 5G in critical infrastructures [5]–[8], including recent research papers published by the CCDCOE

[9]. We first present the STRIDE methodology to achieve this. Then, various threat scenarios (TS) are analyzed in more detail, as well as the associated security controls to address them. Our work lays the foundation for risk analysis of 5G networks in critical infrastructure protection.

2. BACKGROUND

A. STRIDE Methodology

Our threat analysis follows the STRIDE (spoofing, tampering, repudiation, information disclosure, denial of service, and the elevation of privileges) classification [10], [11] of threats developed by Microsoft, which requires data flows between different components to be formalized. The threat assessment methodology is illustrated by six steps in Figure 1. Each component, process, data flow, external entity, and data store is exposed to a subset of threat categories, as described in Table I.

FIGURE 1: STRIDE THREAT ASSESSMENT METHODOLOGY

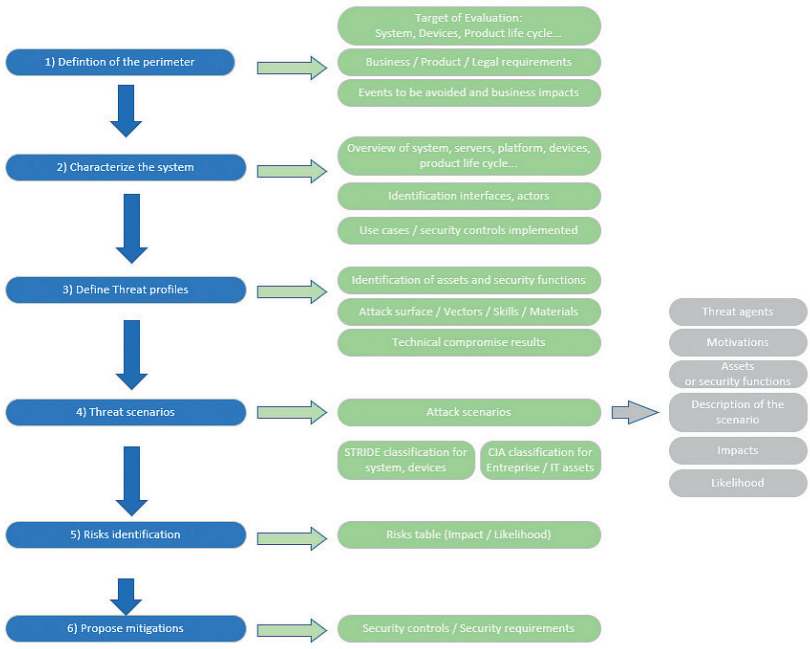


TABLE I: THREATS AFFECTING COMPONENTS WITH STRIDE CLASSIFICATION

Components	Spoofing	Tampering	Repudiation	Information disclosure	Denial of service	Elevation of privileges	STRIDE
External entity or interactors	X		X				SR
Process	X	X	X	X	X	X	STRIDE
Data / Keys storage		X		X	X		TID
Data flow		X		X	X		TID
Devices	X	X		X	X	X	STRIDE

1) 5G System Overview

There are several foundational changes in the 5G architecture compared to 4G. First, the 5G system extends to new frequency spectra, which increase data rates and are well suited for massive MIMO (multiple-input multiple-output) applications and micro-cells. Indeed, transmitters for frequencies in the mm-wave range have intrinsically high directivity, thereby also providing spatial multiplexing capabilities with more ease than at lower frequencies. However, power generation within these frequency ranges is still difficult, and absorption rates by the atmosphere tend to be high. They are therefore unsuitable for macro-cells, which are expected to continue to use frequency bands previously allocated to 3G and 4G cellular networks.

2) 5G New Radio (5G-NR)

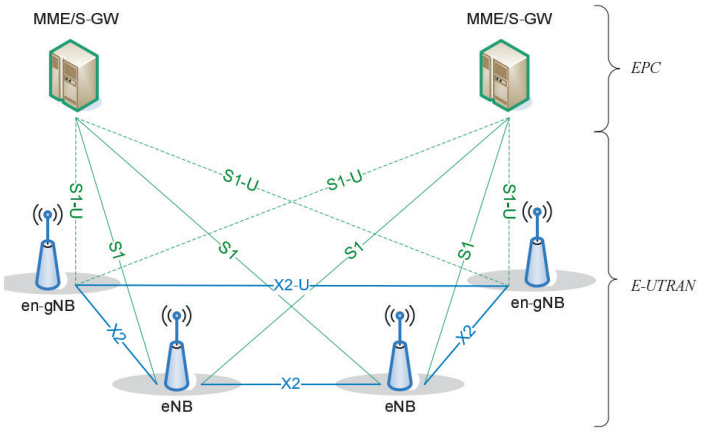
The 5G radio interface uses the same frequency ranges as 4G plus additional frequency bands. This includes frequencies in the sub-6GHz band, particularly the newly attributed frequencies around 3.5 GHz and frequencies around 24–26 GHz. The frequency bands above 6 GHz offer inherently higher bandwidth but present higher absorption rates and thus limit the size of a single cell. Furthermore, at these frequencies it is getting more complicated to use antennas with wide beamwidth as the antenna-to-wavelength ratio has the tendency to result in more directive antennas than at lower frequencies. To adapt to the higher frequency bands and ensure adequate coverage while meeting the increasing demands for end-user performance in uplink and downlink, mobile network operators deploy advanced antenna array systems with beamforming and MIMO capabilities. The frequency bands below 1 GHz still offer the means of achieving coverage with a minimum number of cells (thus achieving coverage in rural areas where the high-density deployment of nano-cells would be too costly).

3) 5G Non-standalone

The first stages in 5G deployment focus on the integration of 5G-NR base stations (known as gNodeBs or gNBs) into the existing 4G system in the context of a multi-radio dual connectivity implementation (see Figure 2). This is done by adhering to standard TS 37.340 [12]. The core network is still the 4G evolved packet core (EPC), and the master nodes for dual connectivity are 4G base stations (eNBs). The 5G base station is integrated as an en-gNB into the system and acts as a secondary node. It only exchanges user plane data with the core network. All control data is exchanged with the eNB over the X2 link. From a user equipment (UE) perspective, the control plane is located in the eNB, while user plane data are transmitted over the gNB. This dual connectivity system also implies that the UEs that support this mode have to integrate concurrent 4G and 5G radio interface support. The increased power consumption might be unsuitable for low-power applications in the IoT context.

Finally, UEs supporting this mode of operation must use the standard 4G network attach procedures, which implies sending their unique international mobile subscriber identity (IMSI) clear to the network during the first attach. This means that the identity concealment feature introduced for 5G is not usable in non-standalone deployments, and IMSI catching is still possible without any increased difficulty.

FIGURE 2: NSA 5G NETWORK ACCORDING TO [12], DEPICTING THE INTERACTION BETWEEN THE 4G EPC USING THE MOBILITY MANAGEMENT ENTITY / SERVING GATEWAY (MME/S-GW) AND THE 5G EVOLVED UNIVERSAL TERRESTRIAL RADIO ACCESS NETWORK (E-UTRAN)



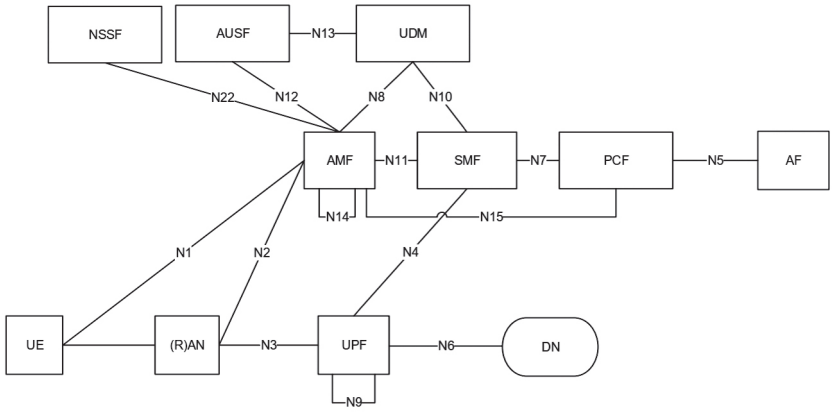
4) 5G Standalone

In the case of a standalone 5G deployment (or of a dual connectivity deployment using a 5G core network), the radio interface and core network differ from 4G. In 5G, the architecture has been designed to achieve a cleaner separation of control and user planes. The core network has been redesigned using a service-based architecture,

which makes the virtualization of some network functions easier. Once virtualized, the network functions can be implemented as cloud instances. To guarantee the security of virtualized network functions, the operator of the 5G system has to pay attention to the isolation mechanisms between the virtual machines. The implicit level of trust in a serving network has also been reduced, and some new security features have been implemented. Authentication and access management functions are now in two different building blocks of the system.

Figure 3 shows the various reference points in the 5G system architecture if no roaming is involved, that is if the serving network corresponds to the home network (roaming is out of the scope of this paper due to space constraints). The access and mobility management function (AMF) is clearly separated from the session management function (SMF). The unified data management (UDM) of the home network and the Universal Subscriber Identity Module (USIM) of the UE contain the same long-term keys used for further key derivation during the authentication process. The authentication server function (AUSF) is located in the home network of the device and performs its authentication. It also provides high-level keys to the AMF that initiated the authentication session.

FIGURE 3: REFERENCE ARCHITECTURE IN THE 5G SYSTEM IN A NON-ROAMING CONTEXT – FOR DETAILED EXPLANATIONS AND MEANINGS OF ABBREVIATIONS, PLEASE REFER TO [2]



B. Novel 5G Security Features

In 5G standalone implementations, some new security features mitigate previously identified vulnerabilities. Contrary to previous versions of the 3rd Generation Partnership Project (3GPP) standards, the universal integrated circuit card (UICC) of the UE now contains an asymmetric key element, the public key of the home network

for use in elliptic curve algorithms. The (limited) use of asymmetric cryptographic algorithms allows the transmission of protected information to the core network without previous key negotiation with this network. This mechanism avoids IMSI catcher attacks that track mobile phones as the unprotected IMSI in the initial attach request has been replaced by an obfuscated subscription concealed identifier (SUCI) in the initial registration request. Further differences between 4G and 5G security features are summarized in Table II.

TABLE II: COMPARISON OF 4G AND 5G SECURITY FEATURES

Security feature	Applies to 4G	Applies to 5G
IMSI obfuscation on radio link	No	Yes, using ECIES scheme
User plane encryption on radio interface level	Yes (operator choice)	Yes (operator choice)
User plane integrity protection on radio interface level	No	Yes (operator choice)
RRC message integrity protection	Yes, EIA0 only allowed for emergency calls	Yes, NIA0 only allowed for emergency calls
RRC message encryption	Yes (operator choice)	Yes (operator choice)
NAS message integrity protection	Yes, EIA0 only allowed for emergency calls	Yes, NIA0 only allowed for emergency calls
NAS message encryption	Yes (operator choice)	Yes (operator choice)
Authentication of UE to serving network	Yes	Yes
Authentication of UE to home network even if using untrusted serving network	No	Yes
Network slicing to provide differentiated handling of service requirements for different applications	No	Yes

C. Protection Goals

We will now discuss the assets that need to be protected in the 5G ecosystem.

1) User Identity and Location

The first assets are user identity and location. The novel concept of transmitting a concealed SUCI instead of the IMSI in an initial registration/attach procedure provides some level of privacy protection. The visiting network is not supposed to be aware of the unconcealed subscription permanent identifier (SUPI) until the end of the authentication procedure. At this point in time, the home network has effectively authenticated the serving network to be trusted. Even when the SUPI is transmitted to the AMF of the visiting network, the identity is still not provided to the gNB.

However, in some cases (e.g., emergency procedures), the UE will still directly communicate its globally unique SUPI. Other temporally persistent identifiers are also still visible during the registration procedures, such as the global unique temporary identifier (GUTI). The core network can request the device's unique international mobile equipment identity (IMEI), which may allow the correlation of a connection with a specific user (particularly if the user connects to both 4G and 5G networks).

If an attacker is capable of correlating the 5G-GUTI with the SUPI or IMEI of a user, it is still possible to track the position of the UE. Indeed, all initial requests in the case of the change of the serving cell will still reveal the 5G-GUTI.

2) Service Availability
The impact of denying a device connectivity varies from small annoyance because a phone call cannot be placed to endangering human life if even emergency calls are no longer possible. For machine-to-machine communications, the systems are expected to be robust in the absence of reliable communications even though the consequences might be anything up to a "graceful" standby of the system.

3) Data Integrity
It is important that the data sink can trust that the incoming data stream is coming from an authentic source. If it is possible to also inject fake data, these pieces of data may not only result in wrong decisions on the receiving end, but the level of trust in any authentic data is also decreased. This can lead either to false-alarm-type situations or to a genuine alarm being disregarded by the system.

4) Data Confidentiality
In all communication contexts, the data transmitted over the radio link is the main asset of this link. Depending on the use case, the data may be sensitive, and its confidentiality has to be protected.

The keys involved in protecting the data both in terms of confidentiality and integrity are secondary assets that must be protected. Indeed, leakage of a device's keys allows an attacker to directly leverage this knowledge to decrypt confidential data and impersonate the device.

5) Network Performance
For safety-critical functions, the general availability of the network service might be insufficient but additionally requires a communications channel that fulfills certain boundary conditions. Such services rely, for example, on low latency or a minimum data rate (quality of service). If the network performance is downgraded below a

given threshold either in terms of latency or data rate, then for these devices, this situation can be equivalent to a complete denial of service condition.

3. THREAT SCENARIOS

In this chapter, we identify the threat scenarios for 5G. Table III lists the scenarios and their contexts according to the STRIDE methodology, which we discuss in detail in the following.

TABLE III: LIST OF POTENTIAL THREAT SCENARIOS

STRIDE	Threat scenarios	Context and potential security controls
STRIDE	TS 01: A disgruntled employee with access to the database of all device keys makes a copy of the keys and sells them to a criminal organization	The UDM manages all keys used inside the network. Security control: strict access control and use of a hardware security module (HSM) to protect the keys, update mechanism of keys stored in the operator's UICCs
STRIDE	TS 02: Key extraction through hardware attacks on the UICC element. First the attacker extracts the keys from the UICC of a valid device. The keys are then used to create clones and attack the network or, if an attack is invasive/ destructive to spy on communications of the legitimate user	Difficulty depends on the robustness of the UICC
STRIDE	TS 03: Malware on the mobile equipment (ME) with sufficient privilege dumps the current security context of a device. The dumped keys can then be used to impersonate the device to the network and to decrypt all previous communications of the device	Keys derived in the context of a registration procedure are held outside the UICC in the context of the ME security control: Regular renewal of the device security context by the network
D	TS 04: Physical or logical jamming of devices through fake gNB	- Impact per jammer limited to its coverage - Except for protocol-based jamming during the attach procedure of a device, the duration of impact is only as long as the jammer is active - Security control: Blacklist of fake gNB broadcast in nominal network
I	TS 05: Partial SUCI and permanent equipment identifier (PEI) catcher through interception of radio link	Security control: encryption of signaling messages both on radio and non-access stratum (NAS) level to protect PEI
D	TS 06: Physical or Logical jamming of gNB	- Impact per jammer limited to one gNB - Impact only as long as the jammer is active - Security control: beam forming networks (BFN) to eliminate the jammer's radio signal
TRIDE	TS 07: Exploit software vulnerability in a gNB (or malicious firmware update) to install backdoors to data buffers and extract signaling information in clear or might result in attacker-managed DoS	- Tampered gNB might share handled data - Might provide access to gNB level key-vulnerabilities might be built in unintentionally or by malicious supplier and actions triggered through radio interface - Security control: External audit of gNB code and secure coding rules Authentication of firmware

TRID	TS 08: Exploit software vulnerability in a network function (or malicious firmware update) can lead to misconfiguration of UEs, data leakage and bypass of security controls; in a virtualized network function this can include data leakage through side-channel attacks between virtual machines using the same physical resources	Tampered network function (e.g., AMF) might disclose current security context of a device or not implement all optional security features
TI	TS 09: Extraction of keys used to establish IPSec connection from link node memory. - If a link node (gNB, AMF, etc.) uses software implementation of IPSec, keys might be exposed through heartbleed-style attacks - In gNB, they might not be stored in secure storage and extracted through local physical access	- Software vulnerabilities - Software implementation of cryptographic suites - Security control: Use of robust hardware module for handling of root keys used for secure channel establishment
D	TS 10: Stealing or modifying the physical configuration of a gNB - Disrupting access to the backhaul - Removal of gNB or its antennas in insufficiently secured physical location	Mitigations: - Physical security for gNB access - Overlap in the cell coverage
D	TS 11: Overloading traffic in high priority slice at the cost of lower priority slices (or slices associated with another public land mobile network (PLMN) in the radio access network (RAN) sharing case)	Mitigations: - Proper implementation of service level agreements and resource management function in gNBs

A. TS 01: Operator UDM Database Theft

The keys contained in the UDM database are also stored in the UICC elements of the UEs. Having control of this database allows an attacker to fully impersonate the network. As it is difficult to update the long-term keys in the UICCs (in particular in embedded systems), it is very costly to respond to this attack and a root key update may be the better option.

The main mitigation is strict physical access control to the UDM. Using a hardware security module (HSM) also forces the attacker to make time-consuming attacks once in possession of the HSM to extract the data. This time window might be sufficient for the operator to be aware of the loss of the device and to deploy new keys in the UICCs in their network.

Threat agents: malicious/compromised employee with access to the UDM storage. Given the amount of confidential information being disclosed through one attack, the motivation for a criminal organization or hostile nation can be considered high.

B. TS 02: Device Long-Term Key Extraction Through Hardware Attacks on the UICC Element

The UICC contains the keys used at the root of the key derivation and agreement process between the UE and the network. If an attacker can extract the key material from a legitimate UICC, the attacker can generate clones of the device, eavesdrop on the communication and inject fake data. One attack vector would be to extract the keys before the initial use of the UICC in a UE, but tampering detection is also difficult later on in certain machine-to-machine contexts. As a mitigation, the network should only authorize one active security context at any given time, and thus the cloned (and legitimate) devices cannot function in parallel.

Threat agents: security researchers to check the robustness of products and test their technical capabilities, criminal organizations, and foreign government agencies.

C. TS 03: Non-permanent Key Extraction from Mobile Equipment

Most keys inside the UE are handled inside the ME and not the USIM. While the security requirements are clearly specified for the USIM [1], the requirements are less clear for the ME. While the baseband and application space inside normal UEs are often separate subsystems, both might be handled in the same processor, particularly for low-cost components.

This opens up the possibility that a malicious application running inside the ME has knowledge of the current security context and allows attackers to eavesdrop and inject messages nominally from the UE to the network. Unless the network triggers the renewal of the security context, these keys will remain valid. For a stationary IoT device, the network might want to limit the amount of exchanged data and thus only renew the security context within long intervals.

Depending on the security mechanisms used by the ME to protect against the installation of malware, this attack can be much easier to perform than TS 02, with a nearly comparable result. Even if more complex ME architectures are used, it is expected that the extraction of a security context from the ME is much less costly than extracting secrets from the UICC. The extraction of the security context can, however, only be achieved once the device is operational.

Threat agents: opportunistic hackers, criminals, and security researchers.

D. TS 04: Physical or Logical Jamming of Devices

The basic physical jamming of devices will only affect the UE if the jammer is active. Depending on the covered frequency bands and the beamforming capabilities of the device, the device might even be capable of blocking the angle of arrival of the

jammer. In the case of a logical jammer, however, equivalents to known 4G attacks [2] are possible, and their impact persists until the device has undergone a power cycle. Indeed, if a UE tries to switch to this rogue gNB following the cell selection and reselection mechanism described in [13], then the rogue gNB can trigger a new registration procedure followed by transmitting an unprotected REGISTRATION REJECT non-access stratum (NAS) message. As stated in [14, section 4.4.4.2], this message must be processed before a valid security context is established between the UE and the network.

In the case of stationary devices, the cell reselection criteria might be difficult to achieve by the rogue gNB as long as the current cell on which the device is camped remains powerful enough. For mobile devices, the rogue gNB only needs to provide a slightly better signal than other candidate cells in the attacked network. Given that some rejection causes require the device to either follow a power cycle or to have its USIM reinserted, this can have a near-permanent effect on some types of devices. For example, a drone being controlled through 5G would naturally either have to disregard the 5G specifications or go into a safe return mode, as there would be no means of a human manually triggering a power cycle while flying.

The cost of the rogue gNB can be estimated to be lower than a high-end physical jammer.

Threat agents: criminal and terrorist organizations.

E. TS 05: Location Tracking Through Standard Radio Link Interception

Depending on the choice of the network operator, signaling messages can only be integrity protected. Even though the SUPI will only be transmitted in its concealed form, an attacker can still gather the same amount of information through the home network identifier transmitted in the context of the authentication procedure, and the PEI transmitted inside the SECURITY MODE COMPLETE message.

If the network operator chooses to use encryption for signaling messages, an attacker can only capture the SUCI and the associated home network identifier. This may be of interest if the target user's home network is more uniquely identifiable (e.g., a visit of a foreign delegation).

If the attacker possesses a network of (potentially low-cost) radio sensors with sufficient density, it is possible to match and continuously track the location of a given set of UEs. Importantly, with knowledge of the target's location at the beginning of the tracking session, it might be possible to track the target without physically following it after this initial matching phase.

Threat agents: In the absence of the encryption of signaling data, location tracking might interest criminals, terrorist organizations, or foreign government agencies. If only the home network identifier could be intercepted, foreign government agencies might remain motivated to implement this attack. If the tracking is based on a sensor network, then it is likely that only government agencies have the resources to install this type of network.

F. TS 06: Jamming of a gNB

The effect of a physical jammer on a gNB will disappear as soon as it is no longer active. From a protocol point of view, it should, however, be quite easy to obtain a modified rogue UE that continuously jams the random access channels of a gNB. Such a logical jammer would deny new UEs from requesting access to the cell. The gNB would be severely impacted in its operations, and network performance for this cell would decrease drastically.

Suppose the gNB detects the presence of this logical jammer and is capable of locating its position. In that case, the gNB might configure its beam forming networks (BFN) to suppress the jammer signal's arrival direction. However, this suppression capability will depend on the size of its antenna array (and indirectly on cell center frequency). Standard external anti-jamming detection and mitigation by providers or authorities can also mitigate this attack.

This attack would only impact a single gNB.

Threat agents: criminals.

G. TS 07: Malware or Software Vulnerabilities on a gNB

The software stacks inside a gNB and the network functions of the 5G core are complex. The manufacturers of the equipment might also not be willing to share the code even with the network operators, as the scheduling function might contain highly proprietary optimizations. The software of a gNB is expected to be updatable.

Vulnerabilities may be present because of backdoors mandated by the government of the equipment manufacturer, due to coding errors, or after the replacement of the original firmware with malicious firmware. Consequences include threats to all availability, integrity and confidentiality. If the vulnerability is already present in the official firmware, it might be exploitable through the radio network. In this case, all gNBs with the same vulnerability would be at risk, and the result could be catastrophic for the infrastructure of a network operator or even a country.

The modified gNB could also be used as an entry point to attack core network functions through the existing link between the gNB and the core network (particularly the user plane function and the AMF). However, the feasibility of this attack depends on the absence of any load balancer in front of the 5G core.

Thanks to virtualization concepts, the non-time critical sections of the gNB central unit can be located in the cloud, which may handle more than one physical radio access network (RAN). In this case, a successful attack on the cloud instance (e.g., physical access to the data center hosting the cloud VM) directly impacts more than a single physical gNB instance.

Threat agents: disgruntled member of the development team for malicious inclusion of a backdoor in the firmware code base, member of the development team unintentionally inserting exploitable vulnerability into the firmware, security researcher analyzing the firmware and detecting a vulnerability, government agency mandating the inclusion of a backdoor in code provided to foreign operators that the mandating government agency can activate at will.

H. TS 08: Malware or Software Vulnerability in 5G Core Network Functions

Similar to the gNB, an attacker might be able to exploit a vulnerability in a network function such as the AMF. Given the key derivation schemes used in 5G, knowledge of lower-level keys does not provide knowledge of higher-level keys. However, this reasoning does not apply in the other direction. A misconfigured SMF could also instruct the gNB to configure the data bearers as not being confidentiality protected.

As the network functions do not require being distributed to cover the territory of the operator, they can be located in physically secure locations. This makes a local attack on network functions less likely.

If virtualization is used, they can also be operated from the cloud and thus be physically hosted in the data centers of cloud service providers. Besides the potential legal consequences, this may enable micro-architectural attacks or open up vulnerabilities in the hypervisor managing the virtual machines.

Threat agents: opportunistic hackers if the control interface of the network function is exposed on the public internet; criminal organizations for blackmailing the network operators; government agencies for espionage and control of foreign infrastructure.

I. TS 09: Stealing Keys Used for Link Protection Between Network Functions

If the network equipment is physically accessible, an attacker might also use physical attacks to extract the network keys. However, the network operator should not rely on physical security alone to protect the data in transit between different network functions. Alternatively, an attacker can extract the keys securing the link through a zero-day exploit against the software running inside the network function. It is also possible to attack cloud solutions via side-channel leakages [16] to other functions executed on the same hardware.

Threat agents: criminals, hackers, and security researchers.

J. TS 10: Theft or Physical Misconfiguration of a gNB

Depending on the type of gNB (stationary or mobile) and its location (e.g., a dedicated building or a shared space), physical access to its antenna may be difficult to protect. The connection between the gNB and backbone is likely even more difficult to protect. Given the skepticism related to 5G radio transmissions in parts of the population, it is possible to imagine that a small community of hacktivists disregards planning or court decisions and actively removes or destroys the antennas of 5G base stations whenever easily accessible.

Threat agents: hacktivists.

K. TS 11: Exploiting Bad Resource Management in Slice Resource Allocation

The sharing of the RAN between operators and, to a lesser extent, slice management by a single operator, opens up the issue of proper resource management under high loads. In the case of RAN sharing, the primary owner of the radio resource might privilege its radio resource requirements and no longer guarantee sufficient bandwidth to the sharing operator in the case of network overload. Apart from the generic network overload aspect, this attack will, however, heavily depend on implementation choices made by the network operator.

Threat agents: criminals, terrorists.

4. RISK ASSESSMENT

Figure 4 shows the summarized risk matrix for all identified threat scenarios, classified by impact and probability of occurrence. The dangerousness of the scenarios decreases from red to light green. The likelihood of an attack is related to various factors, such as a remote or local attack, logical or partial hardware attack, the time required to implement, the cost of equipment, and the expertise required for an attack.

FIGURE 4: RISKS MATRIX OF THREAT SCENARIOS

		Likelihood		
		Unlikely	**Probable**	**Very probable**
Impact	Catastrophic		TS 07	
	Critical	TS 01		
	Very high	TS 08		
	High	TS 02	TS 03	TS 09
	Moderate	TS 11	TS 04 TS 06	TS 05 TS 10
	Low			

5. MITIGATIONS AND SECURITY CONTROLS

Several threat scenarios are only possible due to the under-specification of the 5G standard. Indeed, if an operator implements all optional security and follows the recommendations inside the specifications, then some scenarios are impossible to exploit.

Other threat scenarios rely on an insufficient level of protection by the security features. Indeed, security is not achieved by merely activating a feature but by activating it in a robust manner that withstands attacks against its bypass or deactivation. Concerning TS 01 (lifting of the key database of the subscribers), if it is possible to update the keys in the UICCs used by the network operator and if the used HSM is sufficiently robust, then it might be possible to mitigate the attack before the attacker has been able to extract the keys of the lifted database. However, the robustness of the protection mechanism of the database in the UDM is highly dependent on its logical

and hardware implementation. It might even be possible that the operator is dependent on the physical security of their cloud service provider.

Extracting the keys of a single subscriber through an attack on the associated UICC (TS 02) might be made more difficult by using hardware elements with additional countermeasures against both passive and active attacks. External certification of the UICC might provide an increased level of confidence in its robustness.

Attacks that are based on potential vulnerabilities or non-compliances inside the ME of the user equipment (TS 03 and TS 06) can only be mitigated by the network operator inside the core network. Indeed, only the network operator has control over the UICC inside the terminal. Knowing that the trust in the security of the ME is limited, the network operator should force a renewal of the security context on a regular basis (TS 04) in order to limit the duration of a security breach and be able to suppress some directions of arrival to filter out logical and physical jammer signals (TS 06).

In the current version of the specifications, a compliant device has no means of mitigating logical jamming attacks of some REGISTER REJECT causes sent by the rogue network (TS 04). Indeed, this message can be sent before the establishment of a security context, and the network currently has no means of authenticating itself before the security context has been configured between the network and the device. A potential mitigation of this situation could be as follows: All current global reject causes should be limited to a single network. The network would identify itself by broadcasting a network pre-security context authentication public key (e.g., in one of the system information blocks) and signing the reject message using the associated private key. Therefore, an attacker without knowledge of the real network private key cannot fully impersonate this network.

A rogue gNB could naturally broadcast its own public key and reject the registration of any UE. However, the UE would still be authorized to try to re-register to another network broadcasting a different public key. Note that currently the impact of a fake gNB is potentially much higher for an IoT device (and particularly a moving IoT device) than for a normal mobile phone. In principle, an IoT device is more vulnerable than a mobile phone, since there are fewer optional security features implemented. If the real network is made aware of the presence of a rogue gNB in one of its cells, it can also blacklist this rogue gNB in the system information broadcast by the surrounding legitimate gNBs.

The disclosure of the PEI described in TS 05 is only possible if the network operator chooses not to apply NAS and radio-level encryption for control plane messages. The exploiting of vulnerabilities that allow the extraction of key material or tampering

with the firmware in the gNB or other network elements (TS 07 to TS 09) depends on the robustness of the authentication functions at boot time (but not only) and the presence of vulnerabilities inside the software. As these functions are essential for the correct operation of the network, the network operator should be aware of their importance and implement procedures that make it possible to increase trust in the correct and robust implementation of these functions. This applies to both the equipment manufacturer and other service providers (e.g., cloud operators). The operator should also evaluate the design features used to protect the authenticity of executed functions and the confidentiality of secrets.

Concerning threat scenario TS 10, increasing the acceptance of 5G systems by open discussions with the public should at least reduce the risk of the destruction of base stations by hacktivists. To avoid a network disruption by criminals or terrorists, the physical security of access to the base stations and redundancy in cell coverage are the only means to maintain network operations at all times and in all places.

For TS 11, appropriate resource management between slices taking into account their criticality and general QoS requirements should mitigate this threat.

6. DISCUSSION

Outside the context of UEs in a limited service state, exchanges with the gNB at radio resource control (RRC) level and with the 5GC at NAS level are expected to be integrity protected from a certain state onwards. However, it is unclear to which level UEs implement this part of the specifications and discard messages that are not protected using at least level NIA1. UEs that reply to unprotected Security Mode Commands will still expose their IMEI to a rogue network and thus indirectly disclose the identity of the subscriber. Verification of the adherence of a UE to the standard could be achieved by modifying a fully functional standalone Software-Defined Radio (SDR) implementation of a 5G network that allows deactivating the integrity protection for selected messages and using test SIM cards under the control of the researcher.

For data confidentiality, the activation of data encryption at the radio level and at NAS level is entirely under the network operator's control. To which extent operators activate RRC, NAS, and user plane encryption needs to be verified. Suppose in the control plane, an operator only relies on integrity protection. In that case, the IMEI/PEI and the associated 5G-GUTI of the device can still leak and allow tracking of the user even if the user plane data is encrypted. Using a fully instrumented test UE that

provides access to this level of information would verify the protection level used by operators in the field.

On the network side, it is unclear to which extent operators implement IPSec between all network functions. If an operator relies on the physical security of the network links, then this might allow interception of confidential data (including key material) between the network entities. Without physically forcing access to the operator's network, IPSec can only be verified by auditing the network operators.

In the latest 5G releases, 3GPP has added new services such as edge computing or proximity services with their related network functions that increase the complexity of the operator's networks. These new services and procedures may bring some additional risks or vulnerabilities that will have to be carefully analyzed and assessed. Furthermore, roaming architectures and procedures have been devised for 5G, not all of which have been fully specified by the GSM Association [15], and the use of these intermediate actors significantly increases the attack surface.

7. CONCLUSION

Our comprehensive analysis shows that 5G networks are still exposed to many threats previously identified in 4G implementations. This remains even more true in NSA deployments where the network is 5G in name only (or, to be more precise, only 5G for some aspects of the radio channels). Due to performance constraints in some 5G devices, the network operator might be tempted not to use all possible security controls (e.g., user plane encryption and integrity protection) for the communications of these device classes. The virtualization concepts create additional challenges for the operators, as they potentially create new trust relationships between the operator and third parties, such as cloud service providers.

ACKNOWLEDGMENTS

A special thank you goes to Max Duparc for the proofreading of this article, as well as contributors and reviewers from Kudelski SA for their insightful observations: Alain Paschoud, Nicolas Mutschler, and Benoît Gerhard.

REFERENCES

[1] "TS 33.501. Security architecture and procedures for 5G systems, V17.7.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169

[2] R. Piqueras Jover and V. Marojevic, "Security and protocol exploit analysis of the 5G specifications," *IEEE Access*, vol. 7, pp. 24956–24963, 2019, doi: 10.1109/ACCESS.2019.2899254.

[3] D. Basin, J. Dreier, L. Hirschi, S. Radomirovic, R. Sasse, and V. Stettler, "A formal analysis of 5G authentication," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, Oct. 2018, pp. 1383–1396. doi: 10.1145/3243734.3243846.

[4] "TS 23.501. System architecture for the 5G system (5GS), V17.6.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[5] J. Śliwa and M. Suchański, "Security threats and countermeasures in military 5G systems," in *2022 24th International Microwave and Radar Conference (MIKON)*, Sep. 2022, pp. 1–6. doi: 10.23919/MIKON54314.2022.9924818.

[6] E. Yocam, A. Gawanmeh, A. Alomari, and W. Mansoor, "5G mobile networks: reviewing security control correctness for mischievous activity," *SN Applied Sciences*, vol. 4, no. 11, p. 304, Oct. 2022, doi: 10.1007/s42452-022-05193-8.

[7] J. P. Mohan, N. Sugunaraj, and P. Ranganathan, "Cyber security threats for 5G networks," in *2022 IEEE International Conference on Electro Information Technology (eIT)*, May 2022, pp. 446–454. doi: 10.1109/eIT53891.2022.9813965.

[8] T. Yang et al., "Formal Analysis of 5G Authentication and Key Management for Applications (AKMA)," *Journal of System Architecture*, vol. 126, p. 102478, May 2022, doi: 10.1016/j.sysarc.2022.102478.

[9] V. Oeselg *et al*., "Research Report: Military Movement Risks From 5G Networks," CCDCOE, Tallinn, Estonia, 2022.

[10] B. Potter, "Microsoft SDL threat modelling tool," *Network Security*, vol. 2009, no. 1, pp. 15–18, Jan. 2009, doi: 10.1016/S1353-4858(09)70008-X.

[11] L. Kohnfelder and P. Garg, "The threats to our products," *Microsoft Interface, Microsoft Corporation*, vol. 33, 1999. https://adam.shostack.org/microsoft/The-Threats-To-Our-Products.docx

[12] "TS 37.340 Multi connectivity, overall description, stage-2, V17.1.0." 3GPP. Jul. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198

[13] "TS 38.304. User equipment (UE) procedures in idle mode and in RRC inactive state, V17.2.0." 3GPP. Oct. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3192

[14] "TS 24.501. Non-access-stratum (NAS) protocol for 5G system (5GS); stage 3, V17.8.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3370

[15] GSMA, "NG.132. Report 5G Mobile Roaming Revisited (5GMRR) Phase 1, Version 2.0," Apr. 2022.

[16] Y. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Cross-VM side channels and their use to extract private keys," in *Proceedings of the 2012 ACM Cconference on Computer and Communications Security*, Oct. 2012, pp. 305–316.