The Evolution of Security in 5G

A "Slice" of Mobile Threats

JULY 2019

CONTENTS

1. INTRODUCTION

5G is not only about "faster, bigger or better" networks. It is about enabling a diverse new set of services and use cases affecting nearly every aspect of our lives. But to live up to their potential, 5G-enabled applications must be delivered securely, and security issues must be dealt with at the network foundation from the very beginning to protect both the networks and customers.

With 5G, mobile takes that security focus to another level with a wide variety of new, advanced safeguards. This white paper describes those safeguards in depth, as well as the vulnerabilities and attack vectors that they are designed to mitigate. It also explores how 5G differs from 4G and 3G in terms of radio and core network architectures, and how those differences affect the security mechanisms available to mobile operators, their business partners and their customers.

> *"Looking at the bigger picture, we believe 5G security issues need to be addressed upfront. Making the right choices when deployment is beginning is much easier than trying to correct mistakes once network construction and operation is well underway. Moreover, decisions that impact 5G security need to be made with the long term in mind. Focusing too heavily on short-term considerations could result in choices that are penny-wise but pound foolish."*
>
> U.S. Federal Communications Commission Chairman Ajit Pai

Security has always been a top priority with all previous mobile generations. For example, Third Generation Partnership Project (3GPP) Release 8 added a variety of advanced security/authentication mechanisms[1] via nodes such as the services capability server. Release 11 provided additional capabilities to enable secure access to the core network. These and other 4G-era additions are noteworthy because LTE is the foundation for 5G architecture, including its security mechanisms. And Release 15 and beyond offer further specifications to deliver secure 5G mobile networks.

1.1 5G PROVIDES NEW CYBERSECURITY SAFEGUARDS TO PROTECT BOTH NETWORKS AND CUSTOMERS

5G is the first mobile architecture designed to support multiple, specific use cases, each with their own unique cybersecurity requirements. For example, 5G will enable Massive Internet of Things (MIoT) applications such as traffic sensors and Vehicle-to-Infrastructure (V2I) services that are the foundation for smart cities. It is critical that hackers cannot access that data, hijack IoT devices or disrupt the services with Distributed Denial of Service (DDoS) attacks.

The mobile wireless industry's longstanding emphasis on security has been a strong market differentiator against other wireless technologies—some of which have inherently more vulnerable network architectures. Even mobile's use of licensed spectrum provides a powerful additional layer of protection against eavesdropping on data, voice and video traffic. In the enterprise IT world, network segmentation is a common, proven way to mitigate security risks. Additionally, 5G introduces the concept of network slicing,

[1] *Wireless Technology Evolution Towards 5G*, 5G Americas Whitepaper. February 2017.

which provides mobile operators with segmentation capabilities that were not possible with previous generations.

1.1.1 NEW 5G CYBERSECURITY CONSIDERATIONS AND RESPONSES

5G is the first mobile technology designed to meet the unique requirements of connected cars, connected cities (smart cities), connected homes (smart homes), wearables, health care devices/applications, smart appliances and other IoT devices. In this section, key cybersecurity considerations and responses brought about by 5G are reviewed.

The 5G IoT market is an enormous business opportunity for mobile operators and their business partners. However, its devices and use cases also increase the potential for cyber threats. For example, many of the "things" that make up the IoT landscape have zero-day vulnerabilities such as security holes in software unknown to the vendors and vulnerable to exploitation by hackers. The 5G evolution means billions of these devices and use cases, collectively referred to as the Massive Internet of Things (MIoT), will be using the 5G Radio Access Network (RAN). Thus, MIoT could increase the risk of RAN resource overload by way of Distributed Denial of Service (DDoS) attacks.

Knowing this possibility, the industry needs to start looking at solutions. One strategy is to commission a project that will examine a standards-based solution to inherently and automatically detect and mitigate the risk. To assist with identifying such a solution, the MIoT DDoS scenario can be used to illustrate the threat:

- Hackers identify zero-day vulnerabilities and use them to create a botnet army by infecting many millions or billions of IoT devices with a "remote-reboot" malware
- Next, the hackers instruct the malware to reboot all devices in a specific or targeted 5G coverage area at the same time. This causes excessive, malicious "attach requests," creating a signaling storm that overloads the 5G RAN resources. This DDoS attack makes the RAN unavailable for legitimate use by subscribers.

The current lack of standardization of IoT devices and security features is a major concern, which is why the Internet Engineering Task Force (IETF) and other standards bodies are working to close these gaps. In the MIoT DDoS scenario, one potential solution is to develop malicious signaling storm detection and mitigation functions that would be added to the gNodeB's Central Unit – Control Plane (CU-CP), and Access and Mobility Management Function/Session Management Function (AMF/SMF) component functions.

In addition to the MIoT, 5G creates new cybersecurity considerations due to its use of cloud computing, edge computing, and the convergence of mobile and traditional IT networks by creating new attack vectors. This paper explores how 5G provides a new set of visibility and control elements to help operators protect their networks, business partners and customers.

One example of a visibility tool is the use of synthetically generated application-level probes that travel through the network to get a clear picture of how an application is behaving. Another visibility example is the Path Computation Element (PCE), which has a near-real-time database representing the network topology. This element is queried programmatically to determine the impact of a potential mitigation action on critical service classes for DDoS. Once all of the telemetry is gathered, a security controller and workflow will analyze it and determine suggested mitigation and controls to be applied based on policy.

The mobile industry itself provides layers of security. Operators, vendors, standards bodies, and associations form an iterative loop of continual learning regarding emerging threats and response options. Actions taken to mitigate an attack are considered the control aspect. Some controls are proactive while others are applied after an attack takes place. Typically, there are two types of attacks:

- Zero-day attacks are threats that do not already have either a fingerprint or previous history (signature). Typically, the security controller identifies deviations in known good behavior of the carrier cloud, as well as applications that request service and state. Action is then taken to mitigate the attack or to get additional visibility to properly identify the adversary
- Day-one attacks are threats that have a signature or fingerprint, and quite often, a mitigation strategy exists in advance to handle the attack. Controls take the form of modifications to the carrier cloud to apply quality of service changes in per-hop behavior to minimize the impact of an attack. Controls also take the form of physical and virtual security assets, and are applied as close to the source of the threat as possible in order to minimize collateral damage

Mobile operators have extensive information about the applications they deliver. To mitigate threats, the industry applies this information in a closed-loop iterative process. Innovation and visibility are two key enablers to security mitigation. That is where automation, orchestration and Network Function Virtualization (NFV) come together with cybersecurity technologies and techniques to prevent and contain present and future attacks. The three elements of the closed-loop iterative process are policy, analytics and the application delivery cloud, which is the entire transaction from the application to the servicing networks.

Operators can now correlate geo-location information to behavioral analytics, compare those against policy in the context of a threat to the carrier cloud, and ascertain the nature of that threat and how to address it with far greater clarity. Visibility and control properly applied to today's advanced threats provide the carrier cloud with a powerful level of protection.

In this context, segmentation is a key tool for stopping attacks and attackers from destructive outcomes against mobile networks. The role that network slicing plays in properly segmenting the 5G mobile network, security tools and best practices are key areas of focus in this report.

Network slicing is the ability for automatic configuration and concurrent operation of virtual/logical networks to support independent business operations (for example, vertical use case scenario) on a common physical infrastructure. Network slicing is a fundamental architecture component of the 5G. End-to-end (E2E) network slicing leverages the attributes of central virtualization technology in 5G to flexibly address a wide variety of use cases with different requirements. It also supports multi-vendor and multi-tenant network models over a shared infrastructure.

Service-Based Architecture (SBA) enables the creation of network slices that are optimized for specific services. SBA allows the 5G network to support applications with very different performance requirements simultaneously on the same infrastructure. Additionally, some of these services will have specific security requirements, such as applications where confidential enterprise data, or personal data may be transmitted. In these cases, an isolated network slice can be created to minimize the risk of data leaking outside the network. Another use of the network slicing concept is to create an isolated network slice to handle data streams where end-point trust has not been adequately proven. This approach complements the established process of detection of anomalous traffic patterns and steering traffic with dedicated resources for analysis, quarantining or cleaning. 5G networks will leverage Software Defined Networks (SDN) and NFV to create network slices with each slice tuned and engineered to meet the needs of specific vertical markets.

However, network slicing brings up a number of security issues – from slice isolation to concurrent multiple access to slices by a single user – that require addressing. 5G network slices must be appropriately secured for different use cases. As a result, service providers must place emphasis on measurable security management and assurance. This new architecture itself introduces new types of security threats and an increased attack surface. These issues are addressed in detail in section 4 of this white paper.

The highlights of 5G security considerations and responses discussed in this section were not intended as exhaustive coverage of this topic. 5G will enable complex ecosystems with a variety of new and evolving security needs. The industry must continue to evolve, grow and get smarter to keep networks safe and resilient as 5G begins to dominate the mobile landscape of the future.

1.2 OVERVIEW OF 5G USE CASES

LTE and its predecessors all include a variety of security mechanisms designed to protect networks and their voice, video and data traffic. 5G leverages not only those mechanisms, but also the mobile industry's collective, decades-long experience in analyzing and preventing attacks.

5G enables a wide scope and diversity of use-cases as illustrated in Figure 1.1, all of which create new cybersecurity considerations and requirements. The diagram illustrates the diversity of 5G use cases, along with the varied set of underlying network parameters necessary for a specific category of use cases. For example, the set of parameters important for Mobile Broadband (MBB) service is quite different from the set that defines the Virtual Reality (VR) use cases or Ultra Low Latency category for connected vehicle

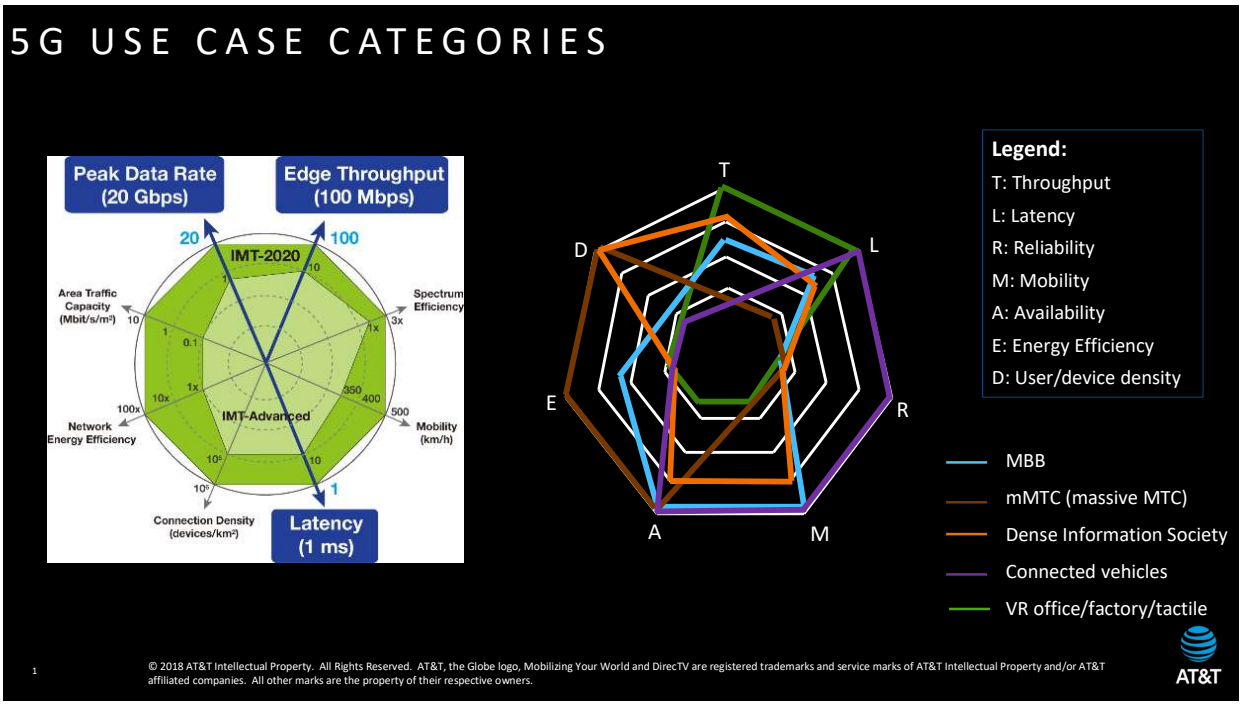

Figure 1.1. 5G Use Case Categories.

services. The difficulty of securing such a wide variety of access and service demands via a single integrated 5G network is readily understandable.

Clearly, for such a wide landscape of use cases, the security issues exposed will also be various. Hackers are continually developing new attack methods, so the mobile industry must also maintain an iterative loop of constant learning about emerging threats and response options. All of these insights, technologies and best practices are key for ensuring that 5G raises the bar for security and privacy similar to previous generations.

2. OVERVIEW OF 5G SECURITY ARCHITECTURE IN 3GPP

3GPP has completed many specifications for the requirements of network and IoT security. This section of the report identifies the new architecture and technology features from the standards designed to protect and secure our communications networks.

2.1 3GPP 5G SECURITY STANDARDS

3GPP unites seven telecommunications standard development organizations and provides their members with a stable environment to produce the reports and specifications that define 3GPP technologies. The project covers cellular telecommunications network technologies including radio access, the core transport network and service capabilities, in addition to work on codecs, security and quality of service. Thus, 3GPP provides complete system specifications, including hooks for non-radio access to the core network and for interworking with Wi-Fi networks.

3GPP technical work groups have specified and standardized mobile wireless industry security features and mechanisms for 3G, 4G and now 5G technologies. The SA3 Working Group (WG) is responsible for security and privacy in 3GPP systems, a role that includes determining the security and privacy requirements and specifying the security architectures and protocols. 3GPP also ensures the availability of cryptographic algorithms which need to be part of the specifications.

3GPP TS 33.501 V15.1.0 (2018-06) is the latest specification published by SA3 for 5G security. It defines the security architecture, features and mechanisms for the 5G system and the 5G core. In addition, it covers the security procedures performed within the 5G system, including the 5G core and the 5G New Radio (NR). Sections 2.1.1-2.2.21 explain the main features defined for 5G security by 3GPP.

2.1.1 INCREASED HOME CONTROL

Home control is used for authentication of the device location when the device is roaming. It allows the home network to verify if the device is actually in the serving network when the home network receives a request from a visited network.

Home control was added to address vulnerabilities found in 3G and 4G networks where networks could be spoofed: sending false signaling messages to the home network to request the International Mobile Subscriber Identity (IMSI) and location of a device. As a result, this information could be used to intercept voice calls and text messages.

2.1.2 UNIFIED AUTHENTICATION FRAMEWORK

In 5G networks, authentication will be access agnostic. The same authentication methods are used for both 3GPP and non-3GPP access networks (for example, 5G radio access and Wi-Fi access).

Native support of Extensible Authentication Protocol (EAP) allows for new plug-in authentication methods to be added in the future without impacting the serving networks.

2.1.3 SECURITY ANCHOR FUNCTION (SEAF)

5G introduces the concept of an anchor key, with the new function of the Security Anchor Function (SEAF). The SEAF allows for the re-authentication of the device when it moves between different access networks or serving networks without having to run the full authentication method (for example, Authentication and Key Agreement (AKA). This reduces the signaling load on the home network Home Subscriber Server (HSS) during various mobility services. The SEAF and the Access and Mobility Management Function (AMF) could be separated or co-located. In 3GPP Release 15, the SEAF functionality is co-located with the AMF.

2.1.4 SUBSCRIBER IDENTIFIER PRIVACY

In 5G, a globally unique Subscriber Permanent Identifier (SUPI) is allocated for each subscriber. Examples for SUPI formats include the IMSI and Network Access Identifier (NAI). The SUPI is never disclosed over the air in the clear when a mobile device is establishing a connection. This is different from 3G and 4G networks, where the IMSI is disclosed when a device is going through an attach procedure (and another vulnerability in 3G and 4G networks) before the device is even able to authenticate with the new network.

Instead of disclosing the SUPI, a Subscription Concealed Identifier (SUCI) is used until the device and network are authenticated. Only then does the home network disclose the SUPI to the serving network. This procedure has been defined to prevent IMSI catchers (also known as false base stations, or Stingrays) from retrieving the subscriber's identity. This is accomplished by forcing a device either to attach to the Rogue Base Station (RBS) or perform attachment process to operator's Base Station while sniffing the unencrypted traffic over the air.

2.1.5 3GPP 5G SECURITY ARCHITECTURE

3GPP defines the overall 5G security architecture, illustrated in Figure 2.1.

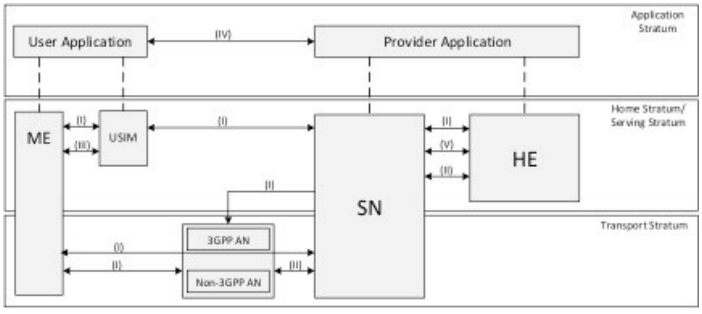

Figure 2.1. Overview of 5G Security Architecture.

This includes many network architectural elements and concepts such as:

- Network access security (I), which is the set of security features that enables user equipment (UE) to authenticate and access services via the network securely, including 3GPP access and non-3GPP access, and particularly to protect against attacks on the radio interfaces. In addition, it includes the security context delivery from SN to UE for the access security
- Network domain security (II), which is the set of security features that enables network nodes to securely exchange signalling data and user plane data
- User domain security (III), which is the set of security features that secures the user access to mobile equipment (ME)
- Application domain security (IV), which is the set of security features that enables applications in the user domain and in the provider domain to exchange messages securely
- SBA domain security (V), which is the set of security features regarding SBA. These include the network element registration, discovery and authorization security aspects, and also the protection for the service-based interfaces
  Visibility and configurability of security (VI), which is the set of features that enables the user to be informed whether a security feature is in operation

2.1.5.1 SECURITY EDGE PROTECTION PROXY (SEPP)

To protect messages that are sent over the N32 interface, the 5G system architecture implements Security Edge Protection Proxy (SEPP) at the perimeter of the Public Land Mobile Network (PLMN) network. SEPP receives all service layer messages from the Network Function (NF) and protects them before sending them out of the network on the N32 interface. Additionally, it receives all messages on the N32 interface and after verifying security where present, it forwards them to the appropriate network function.

The SEPP implements application layer security for all the layer information exchanged between two NFs across two different PLMNs. Figure 2.2 illustrates the SEPP's role.

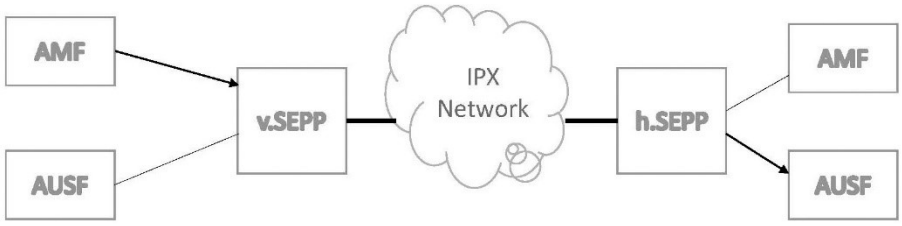

Figure 2.2. The Role of the SEPP in the Security Architecture.

2.1.5.2 ROLE OF THE SEPP IN THE SECURITY ARCHITECTURE

The application layer traffic comprises all the IEs in the HyperText Transfer Protocol (HTTP) message payload, sensitive information in HTTP message header and Request URI. Not all IEs get the same security treatment in SEPP. Some IEs require end-to-end (e2e) encryption, while others require only E2E integrity protection. Still, others may require E2E integrity protection but modifiable by an intermediate Internetwork Packet Exchange (IPX) provider while in-transit.

To enable the trusted intermediary IPX nodes to see and modify specific IEs in the HTTP message—while simultaneously protecting all sensitive information end-to-end between SEPPs—the SEPP implements application layer security in such a way that:

- Sensitive information such as authentication vectors are fully E2E, and confidentiality protected between two SEPPs. This ensures that no node in the IPX network shall be able to view such information while in-transit
- IEs that are subject to modification by intermediary IPX nodes are integrity protected and can only be modified in a verifiable way by authorized IPX nodes
- Receiving SEPP can detect modification by unauthorized IPX nodes

The SEPP shall support the following requirements:

- Act as a non-transparent proxy node
- Protect application layer control plane messages between two NFs belonging to different PLMNs That use the N32 interface to communicate with each other
- Perform mutual authentication and negotiation of cipher suites with the SEPP in the roaming network
- Handle key management aspects that involve setting up the required cryptographic keys needed for securing messages on the N32 interface between two SEPPs
- Perform topology hiding by limiting the internal topology information visible to external parties
- Provide a single point of access and control to internal NFs as a reverse proxy
- Verify whether the sending SEPP is authorized to use the PLMN ID in the received N32 message as the receiving SEPP
- Clearly differentiate between certificates used for authentication of peer SEPPs and certificates used for authentication of intermediates performing message modifications
- Discard malformed N32 signaling messages
- Implement rate-limiting functionalities to defend itself and subsequent NFs against excessive CP signaling; this includes SEPP-to-SEPP signaling messages
- Implement anti-spoofing mechanisms that enable cross-layer validation of source and destination address and identifiers (for example, FQDNs or PLMN IDs)

2.1.6 REQUIREMENTS FOR E2E CORE NETWORK INTERCONNECTION SECURITY

A solution for E2E core network interconnection security shall satisfy the following requirements:

- support application layer mechanisms for addition, deletion and modification of message elements by intermediate nodes except for specific message elements described in the present document. A typical example for such a case is IPX providers modifying messages for routing purposes
- provide confidentiality and/or integrity E2E between the source and destination networks for specific message elements identified in the present document. For this requirement to be fulfilled, the SEPP – cf [2], clause 6.2.17 shall be present at the edge of the source and destination networks dedicated to handling E2E Core Network Interconnection Security.[2] The confidentiality and/or integrity for the message elements is provided between two SEPPs of the source and destination PLMN

[2] 3GPP TS 23.501: *System Architecture for the 5G System*.

    - The destination network shall be able to determine the authenticity of the source network that sent the specific message elements protected; for this requirement to be fulfilled, it shall suffice that a SEPP in the destination network that is dedicated to handling E2E Core Network Interconnection Security can determine the authenticity of the source network
- have minimal impact and additions to 3GPP-defined network elements
- use standard security protocols
- cover interfaces used for roaming purposes
- account for considerations on performance and overhead
- prevent replay attacks
- cover algorithm negotiation and prevention of bidding down attacks
- account for operational aspects of key management

2.1.7 AUTHENTICATION FRAMEWORK

The purposes of the primary authentication and key agreement procedures are to enable mutual authentication between the UE and the network and to provide keying material that can be used between the UE and the serving network in subsequent security procedures. The keying material generated by the primary authentication and key agreement procedure results in an anchor key called the KSEAF, which is provided by the Authentication Server Function (AUSF) of the home network to the SEAF of the serving network.

Keys for more than one security context can be derived from the anchor key without the need of a new authentication run. A concrete example of this is an authentication run over a 3GPP access network that can also provide keys to establish security between the UE and a Non-3GPP Interworking Function (N3IWF) used in untrusted non-3GPP access.

The UE and the serving network shall support Extensible Authentication Protocol and Key Agreement (EAP-AKA) and 5G AKA authentication methods. The home network operator selects the authentication method to be used. The Universal Subscriber Identity Module (USIM) shall reside on a Universal Integrated Circuit Card (UICC). The UICC may be removable or non- removable.

For non-3GPP access networks, USIM applies in case of terminal with 3GPP access capabilities. If the terminal supports 3GPP access capabilities, the credentials used with EAP-AKA and 5G AKA for non-3GPP access networks shall reside on the UICC. EAP-AKA and 5G AKA are the only authentication methods that are supported in the UE and serving network.

2.1.8 GRANULARITY OF ANCHOR KEY BINDING TO SERVING NETWORK

The primary authentication and key agreement procedures shall bind the anchor key KSEAF to the serving network. The binding to the serving network prevents one serving network from claiming to be a different serving network, and thus provides implicit serving network authentication to the UE.

This implicit serving network authentication shall be provided to the UE regardless of the access network technology, so it applies to both 3GPP and non-3GPP access networks.

The anchor key binding shall be achieved by including a parameter called "serving network name" into the chain of key derivations that leads from the long-term subscriber key to the anchor key.

2.1.9 MITIGATION OF BIDDING DOWN ATTACKS

An attacker could attempt a bidding down attack by making the UE and the network entities believe that the other side does not support a security feature, even when both sides do support a security feature. A SEPP can help ensure that a bidding down attack, in the above sense, can be prevented.

2.1.10 SERVICE REQUIREMENTS

A UE shall support a man-machine interface setting for the user to disable use of one or more of the Mobile Equipment's (ME) radio technologies for RAN access, regardless of PLMNs. The radio technologies that can be individually disabled depends on the radio technology that the UE supports, such as the 3GPP standards -- GSM/EDGE, WCDMA, LTE and 5G New Radio (NR).

A UE shall support a man-machine interface setting enabling the user to re-enable use of one or more of the ME's radio technologies for RAN access, regardless of PLMNs. The user can only re-allow a radio technology that the user has previously disallowed. A UE shall support a secure mechanism for the home operator to disallow selection of one or more of the ME's radio technologies for RAN access, regardless of PLMNs. Radio technologies that individually can be disallowed are at least the 3GPP technology standards.

A UE shall support a secure mechanism for the home operator to re-allow selection of one or more of the ME's radio technologies for RAN access, regardless of PLMNs. Radio technologies that individually can be re-allowed are at least GSM/EDGE, WCDMA, LTE and 5G NR. The home operator can only re-allow a radio technology that the home operator has previously disallowed.

For a prioritized service (for example, emergency services, Multimedia Priority Service (MPS), mission-critical services), the UE shall support a mechanism to automatically override user- and network-disallowed Radio Access Technologies (RATs) when there are no PLMNs on the allowed radio technologies identified that the UE is able to access.

Upon power cycle or when the USIM is disabled, the UE configuration of enabled/disabled radio technologies configured by the user shall remain as it was before the USIM was disabled. In other words, the radio technologies disallowed by the HPLMN shall remain as they were before a power cycle. The radio technologies disallowed by the HPLMN shall be bound to the USIM.

2.1.11 5G IDENTIFIERS

Each subscriber in the 5G system shall be allocated one 5G Subscription Permanent Identifier (SUPI) for use within the 3GPP system. The Subscription Concealed Identifier (SUCI) is a privacy-preserving identifier containing the concealed SUPI.

The 5G system supports identification of subscriptions independently of identification of the UE. Each UE accessing the 5G system shall be assigned a Permanent Equipment Identifier (PEI). The 5G system supports allocation of a temporary identifier (5G-GUTI) in order to support user confidentiality protection.

2.1.12 SUBSCRIPTION PERMANENT IDENTIFIER (SUPI)

A globally unique 5G Subscription Permanent Identifier (SUPI) shall be allocated to each subscriber in the

5G system and provisioned in the Unified Data Management/User Data Repository (UDM/UDR). The SUPI is used only inside 3GPP system, and its privacy is specified in TS 33.501.

The following have been identified as valid SUPI types for this release:

- IMSI as defined in TS 23.003
- Network Access Identifier (NAI) using the NAI RFC 4282 based user identification as defined in TS 23.003 By using the NAI, it will be possible to also use non-IMSI-based SUPIs

It is possible for a representation of the IMSI to be contained within the NAI for the SUPI (for example, when used over a non-3GPP access technology).

In order to enable roaming scenarios, the SUPI shall contain the address of the home network (for example, the Mobile Country Code [MCC] and Mobile Network Code [MNC] in the case of an IMSI-based SUPI).

For interworking with the Evolved Packet Core (EPC), the SUPI allocated to the 3GPP UE shall always be based on an IMSI to enable the UE to present an IMSI to the EPC.

2.1.13 SUBSCRIPTION CONCEALED IDENTIFIER (SUCI)

When the SUCI uses the Null-Algorithm, it does not provide privacy protection. The UE shall generate a SUCI using a protection scheme with the raw public key that was securely provisioned in control of the home network.

The UE shall not conceal the home network identifier, such as the MCC or MNC.

The UE shall include a SUCI only to the following 5G Non-Access Stratum (NAS) messages:

- If the UE is sending a registration request message of type "initial registration" to a PLMN for which the UE does not already have a 5G- Globally Unique Temporary Identifier (GUTI), the UE shall include a SUCI to the Registration Request message
- If the UE includes a 5G 5G-GUTI when sending a registration request message of type "re-registration" to a PLMN and, in response, receives an identity request message, then the UE shall include a SUCI in the Identity Response message

The UE shall generate a SUCI using "null-scheme" only in the following cases:

- If the UE is making an unauthenticated emergency session and it does not have a 5G-GUTI to the chosen PLMN
- If the home network has configured "null-scheme" to be used
- If the home network has not provisioned the public key needed to generate a SUCI

2.1.14 SUBSCRIPTION IDENTIFICATION SECURITY

The subscriber identification mechanism is represented in Figure 2.3. This may be invoked by the serving network when the UE cannot be identified by means of a temporary identity (5G-GUTI). It should be used when the serving network cannot retrieve the SUPI based on the 5G-GUTI by which the subscriber identifies itself on the radio path.

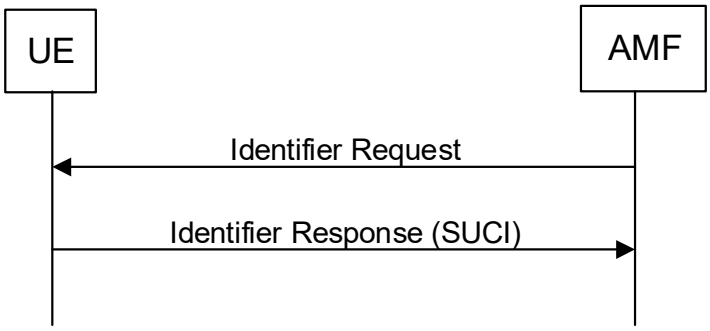

Figure 2.3. Subscriber Identification Mechanism.[3]

2.1.15 PERMANENT EQUIPMENT IDENTIFIER

Each UE accessing the 5G System shall be assigned a Permanent Equipment Identifier (PEI).

- The PEI shall be securely stored in the UE to ensure the integrity of the PEI
- The UE shall only send the PEI in the NAS protocol after NAS security context is established, unless during emergency registration when no NAS security context can be established

2.1.16 SUBSCRIPTION IDENTIFIER DE-CONCEALING FUNCTION

The Subscription Identifier De-Concealing Function (SIDF) is responsible for de-concealing the SUPI from the SUCI. The SIDF uses the private key part of the privacy-related home network public/private key pair that is securely stored in the home operator's network. The de-concealment shall take place at the UDM. Access rights to the SIDF shall be defined, such that only a network element of the home network is allowed to request SIDF.

2.1.17 5G GLOBALLY UNIQUE TEMPORARY IDENTIFIER

The AMF shall allocate a 5G Globally Unique Temporary Identifier (5G-GUTI) to the UE that is common to both 3GPP and non-3GPP access. It shall be possible to use the same 5G-GUTI for accessing 3GPP access and non-3GPP access security context within the AMF for the given UE. An AMF may re-assign a new 5G-GUTI to the UE at any time. The AMF may delay updating the UE with its new 5G-GUTI until the next NAS transaction.

The 5G Serving Temporary Mobile Subscriber Identity (S-TMSI) is the shortened form of the GUTI to enable more efficient radio signaling procedures, for example, during Paging and Service Request.

2.1.18 PROCEDURE FOR USING SUBSCRIPTION TEMPORARY IDENTIFIER

The procedure for using a subscription temporary identifier is an important element of 5G security as described:

[3] 3GPP TS 33.501.

- A new 5G-GUTI shall be sent to a UE only after a successful activation of NAS security. The 5G-GUTI is defined in the 3GPP TS 23.003
- Upon receiving registration request message of type "initial registration" or "mobility registration update" from a UE, the AMF shall send a new 5G-GUTI to the UE in a registration accept message
- Upon receiving registration request message of type "periodic registration update" from a UE, the AMF should send a new 5G-GUTI to the UE in a registration accept message
- Upon receiving a network-triggered service request message from the UE (therefore, a service request message sent by the UE in response to a paging message), the AMF shall use a UE Configuration Update procedure to send a new 5G-GUTI to the UE

This UE Configuration Update procedure shall be used before the current NAS signaling connection is released. Specifically, it does not need to be a part of the service request procedure because that would delay the service request procedure.

2.1.19 SUBSCRIBER PRIVACY

Subscriber privacy is an important element to the security aspects of the mobile network architecture as described in the process:

- The UE shall support 5G-GUTI
- The SUPI should not be transferred in clear text over 5G RAN except routing information, such as the MCC and MNC
- The ME shall support at least one non-null scheme
- The home network public key shall be stored on the tamper-resistant secure hardware component
- The UE shall support the null-scheme

If the home network has not provisioned the public key in the tamper-resistant secure hardware component, the SUPI protection in the initial registration procedure is not provided. In this case, the null-scheme shall be used by the ME.

Based on the operator's decision, indicated by the USIM, the calculation of the SUCI shall be performed either by the USIM or by the ME. If the indication is not present, the calculation is in the ME.

In case of an unauthenticated emergency call, privacy protection for SUPI is not required.

Provisioning, and updating the home network public key in the tamper-resistant hardware, shall be in the control of the home network operator. The provisioning and updating of the home network public key are out of the scope of the present document. It can be implemented using, for example, the over-the-air (OTA) mechanism.

Subscriber privacy enablement shall be under the control of the home network of the subscriber.

2.1.20 SECURE STEERING OF ROAMING

The 3GPP Release 15 standard for 5G added native support for a secure Steering of Roaming (SoR) solution. The 5G SoR solution enables the home network operator to steer its roaming customers to its preferred Visited Public Land Mobile Networks (VPLMN) to enhance roaming customers' experience and reduce roaming charges.

2.1.21 UE-ASSISTED NETWORK-BASED DETECTION OF FALSE BASE STATION

The UE in Radio Resource Control (RRC)_CONNECTED mode sends measurement reports to the network in accordance with the measurement configuration provided by the network. These measurement reports have security values in being useful for detection of false base stations or SUPI/5G-GUTI catchers.

3. 5G THREAT SURFACE

The 5G threat surface, is expansive and challenging for mobile operators. The good news is that the people, processes and tools have also evolved. This section covers some of the key areas of the 5G threat surface.

3.1 NETWORK THREATS IN 4G – BUILDING A SECURE PATH TO 5G

The security capabilities and baseline recommendations of 5G architecture and protocols are greatly enhanced compared to previous generations. Security functions provided by the 3GPP standard are based on proven 4G security mechanisms, and, as discussed in section 2, also include enhancements for encryption, mutual authentication, integrity protection, privacy and availability. As carriers transition to 5G, attackers may look to take advantage of existing as well as new 2G, 3G and 4G threats to target mobile devices and infrastructure. Threats can materialize from anywhere due to the volume of devices and the complexity of the existing and newly deployed operator infrastructures, and not all threats can be addressed at a standards level. Appropriate measures of resilience and mitigation should be in place to account for certain threats such as jamming, physical disruption or DDoS.

To securely transition to 5G, it is important to consider some of the primary network threats that concern the 4G networks operated today. Figure 3.1 highlights examples of threat vectors across Core, RAN, UE and other areas that network operators consider today, wherein they apply the concepts of visibility, segmentation and controls to mitigate or eliminate those threats. These are classic threat vectors, or specific entry or weakness points at various places in the network. This is not an exhaustive list, but it provides the foundation to look at what is new in 5G (for example, network slicing) and the incremental threat surface of 5G and the use cases that may create the threats.

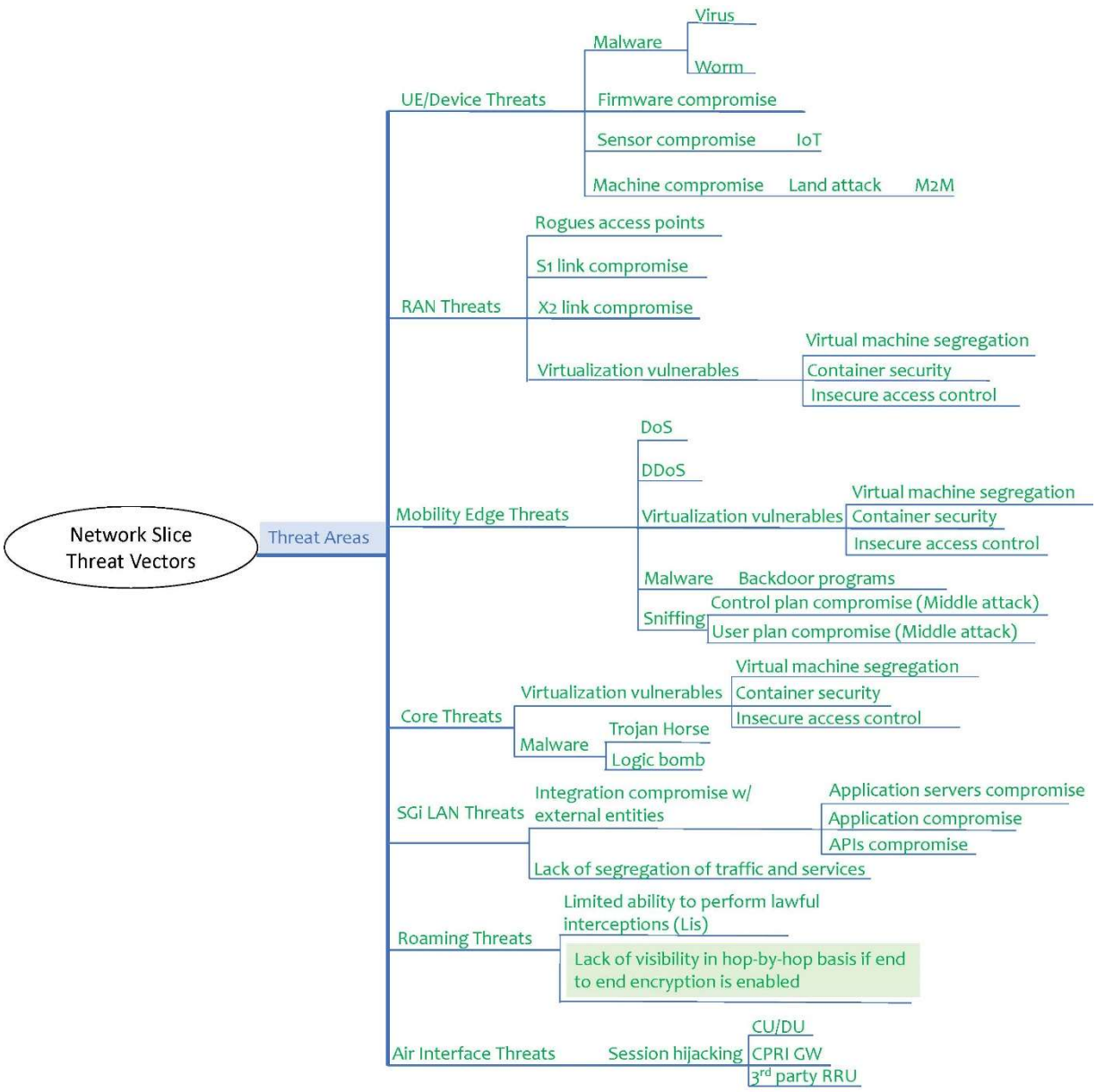

Figure 3.1. 5G Threat Surface Overview.

Some of the threats listed in Figure 3.1 are also applicable in 5G. For instance, in the UE/Devices threats vector, Machine-to-Machine (M2M) may be limited in power, processing and memory resources, rendering it a subpar candidate for implementing security. The M2M and other IoT use cases will require security application in the network but not necessarily on the sensor or UE. If not properly secured, under-protected M2M communication can disrupt critical infrastructure. 5G networks support a vast number of connected entities and enable a huge increase of bandwidth and/or total connections, which naturally create a threat landscape that can be more impactful on the network infrastructure.

Additionally, network slicing is an enabler for segmentation of mobile network functions, allowing the operator to be more agile in the application of security policy to each use case. As with any network structure, network slicing will also have its own threat surface and will require best common practices to

ensure that the network is segmented securely. Segmentation is especially important in 5G as ultra-low latency use cases (like those involving IoT and M2M) will require a widely distributed deployment of network and security functions out to the MEC edge. Subsequently, both the edge and the facilitating network underneath it must be properly secured.

The number of application scenarios where edge paradigms can be applied is huge (mobile edge threat vector). There is no agreed up on global perimeter to the edge. A core network function such as UPF can be installed at the edge but may not have the same level of security scrutiny as when it is installed at a centralized location. The open nature of mobile edge creates a scenario where malicious adversaries can deploy their own devices and applications. This threat produces the same outcome as the Man-in-the-Middle attack, which can create the ability to sniff traffic without authorization.

Threats can occur in roaming traffic (roaming threats vector) between visited and home networks. End-to-end encryption limits the ability to perform lawful interceptions on inbound roamers because the security parameter to decrypt and extract the service content is in the visited network. In addition, this can make it difficult to determine if sufficient security counter-measures are in place on hop-by-hop basis.

Threats culminating from serving different types of traffic and services at the SGi interface can be problematic (SGi threats vector). The SGi interface connects the PGW to an external network. The SGi interface could potentially be servicing different categories of devices. A hack into one category of devices can impact other device categories. These threats can target either or both the devices (for example, a UE compromise via possibly a botnet attack) and/or mobile network infrastructure (PGW User Plane DDoS, for instance) which render the possibility that an attack on one may impact the other.

Another way to look at the existing threat surface within 2G, 3G and 4G is to consider known attacks against classic tenets of security such as authentication, integrity, availability and privacy. This is further illustrated in Figures 3.2 to 3.5

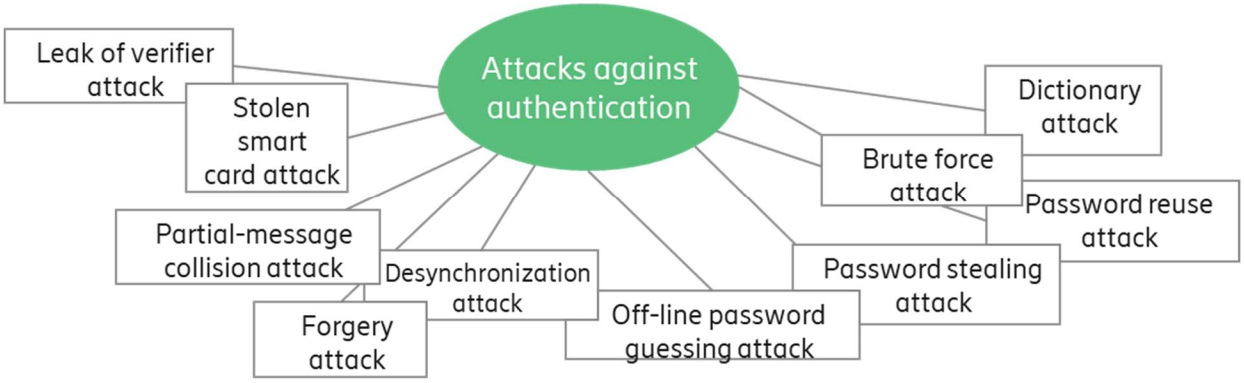

Figure 3.2. Attacks Against Authentication.

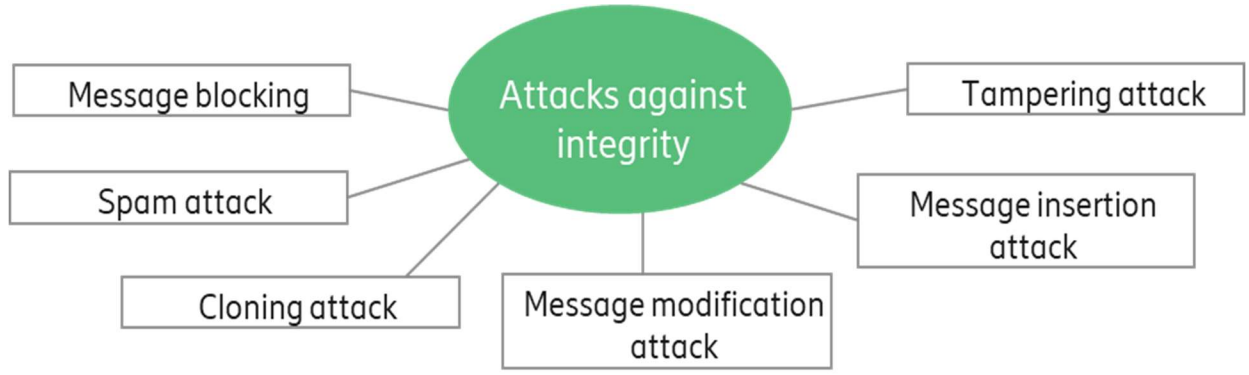

Figure 3.3. Attacks Against Integrity.

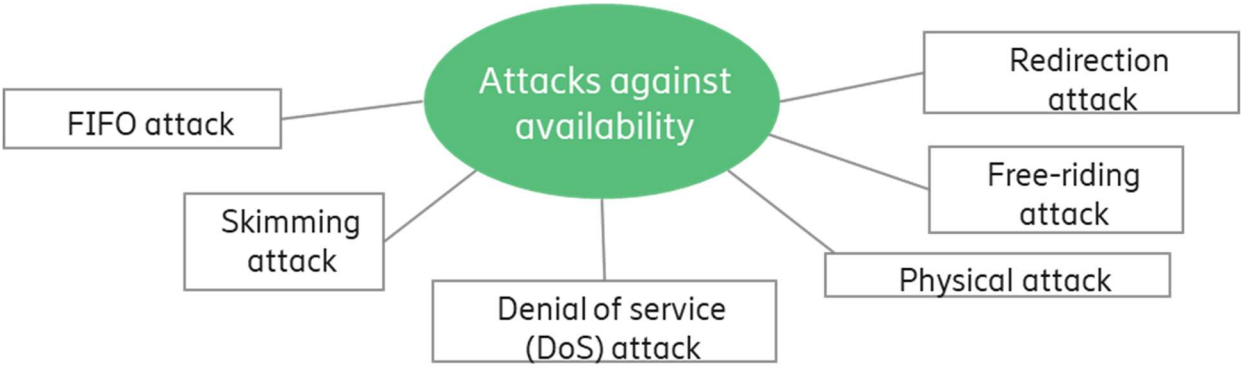

Figure 3.4. Attacks Against Availability.

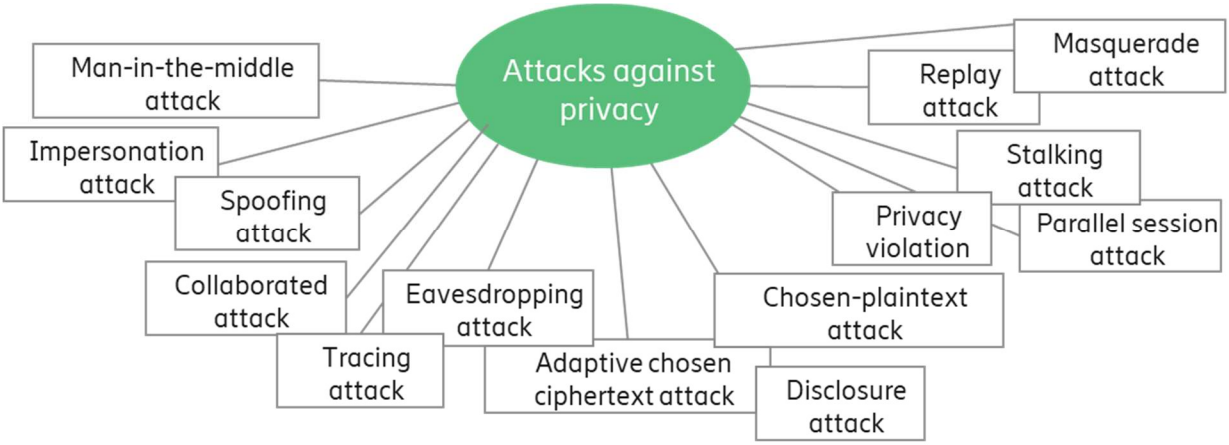

Figure 3.5. Attacks against Privacy.

Examples of continued 4G weaknesses are:

- IMSI that is sent unencrypted over the radio or with lack of variability (IMSI catchers)
- User plane on the radio interface is not integrity protected. This could lead to data injection or modification such as Man in the Middle (MitM)

- Fake base stations that are set up to track users or enable eavesdropping
- SS7/Diameter vulnerabilities could expose users to eavesdropping voice conversations, reading or transmitting of messages and tracking of phones

Known threats at the RAN, air interface and roaming areas, while theoretically possible in 4G, are mitigated or removed by improvements within the 3GPP current standard. Mutual authentication, greater confidentiality and integrity protection of the user plane, privacy enhancements to protect subscriber identity (IMSI encryption and the use of temporary, random identifiers) and inter-operator security should eliminate many of the threat vectors previously noted. These are explained in greater detail elsewhere in the paper.

While the 4G weaknesses have been dealt with due to multiple 5G improvements, other 4G threats within Edge, Core or SGi interface are largely mitigated via policy, architecture or functions. Thus, a vulnerable network design in 4G will likely still be vulnerable in 5G, so it is important to not rely on new standards to improve every aspect of security assurance. One caveat would be the use of network slicing to isolate and mitigate threats, which, although a new technology in 5G, is primarily a deployment/implementation consideration rather than within the scope of 3GPP standards. Thus, even good design choices within 4G could potentially be improved by taking advantage of new technological innovations across different areas of 5G.

The decentralization and virtualization of many areas of the 5G network will create new trust layers, domains and functional or exposed weak spots. However, in terms of secure 5G deployment considerations, new/enhanced approaches to handle threats within a service (vertical) or across a group of services (horizontal) protections are available. For horizontal, system-wide security, this would include:

- The strengthening of network resiliency
- Network-slicing and need-based security functions
- Application-level security that utilizes the trust stack of other domains
- Confidentiality and integrity protection across the radio network
- TLS between 5G Core functions, regardless of architecture
- Service Based Architecture (SBA) that allows for splitting of functional-level components, even at the radio unit. For increased vertical security across all the functional elements, hardening of the virtualized stack and the use of trusted layers within embedded systems is critically important. This may require the virtualized layers to utilize trusted components at the hardware level (via TPM, HSM or secure enclaves) and expose that to applications vertically

As service providers transition to 5G, the increasing and varying connectivity demands present an opportunity to offer new business models using different technologies and solutions. Network slicing creates multiple networks that share a common virtual and physical infrastructure. This enables service providers to dedicate a portion of their network to specific service or functionality and makes it easier to deploy various 5G applications. The 5G ecosystem can be delivered using a slew of technologies including, but not limited to physical boxes, virtual machines and containers. Although network slicing inheritably has the traits of security in the form of the isolation that it provides, it is important to note that network slicing is not guaranteed to provide security. The common virtual and physical NFV infrastructure where network slicing is hosted must also be built with security in mind.

Hence, it will be incomplete to omit the open security challenges that NFV brings to bear. One of the security challenges is to implement a complete and standardized ETSI NFV architecture to deploy virtual security functions to adapt and adjust to different threats in real time. Specifically, NFV security in the realm of 4G

has been more static than dynamic, and 5G's dynamic nature will not be successful without adaptable NFV security.

To illustrate this, Virtual Network Function (VNF) security will have to scale both horizontally and vertically to provide adequate security and performance to other VNFs. A perfect VNF security will not be useful if it cannot scale to cope with the velocity and variety of intensive 5G traffic. Therefore, VNFs will need to have support for orchestration modules which can be leveraged to communicate with the orchestrator and receive instructions which can be acted upon. Another challenge is to securely manage VNFs throughout their lifecycles.

Additionally, conducting the trust management amongst NFV hardware and software vendors is challenging. In particular, the maintainability of the trust chain can be problematic. Case in point, verifying the trust chain is still not completely ironed out, many attestation technologies only provide the boot time attestation, and there are usually no checks and balances that occur during run time. Run time attestation is still an open research area that needs to be explored further.

Finally, with the existing paradigm shift to containerization, vendors and operators alike have been experimenting with Container Network Functions (CNFs). Containers can be efficient, but they were not necessarily built with security in mind, which is usually a common theme for new technologies.

Typically, containers have loose access to kernel resources, rendering them vulnerable to tampering with the container's execution path. Although containers provide the convenience of micro-services creation and separation, that does not ensure the creation of security boundaries. In fact, they do not offer guaranteed security isolation, and can be considered a less secure deployment option. Case in point, network services instances (for example, containers) could break out of their running containers and gain control of other containers running on the host. This could be caused by unpatched vulnerabilities in the kernel, in the container infrastructure, and/or misconfiguration in the container or of the container host.

3.2 IOT THREAT SURFACE WITH 5G

A 2017 study[4] to investigate the impact of IoT security on IT and line-of-business (LoB) leaders revealed IT and LoB leaders' anxieties concerning IoT security because attacks can significantly affect critical business operations. One troubling fact involving IoT, revealed the majority of organizations cannot provide a complete account of all their network-connected devices. Each new device that comes online represents another expansion (another attack vector) of the overall threat surface. Even for identified IoT entities, the ownership, from a security point of view, frequently remains murky, further compounding the problem. Moreover, 90 percent of the companies expected an increase in the volume of connected devices.

In 2016, hackers launched some of the biggest cyberattacks in internet history. These DDoS attacks were executed by infecting multiple internet-connected devices (for example, surveillance cameras, DVRs, routers) and then using them to launch coordinated DDoS assaults on an array of targets such as web hosting service providers and journalists. This was named the Mirai virus. The disturbing fact about Mirai, which became clear when the source code was later revealed, was the relative lack of programming sophistication involved. Launching this botnet of things attack did not require a high degree of programming skill, as the basic tools were easily available and accessible to all on the internet. The Mirai event clearly highlighted key IoT security issues.

[4] *IoT and OT Security Research Exposes Hidden Business Challenges,* Forrester Consulting report commissioned by Forescout Technologies, Inc. 2017. https://www.forescout.com/iot_forrester_study/

The four broad principles worthy of note for securing IoT infrastructure are:

1. Securing IoT should not be an afterthought. IoT security needs to be addressed at the design phase, not added post deployment
2. Whether it is healthcare, automotive or energy, IoT intrinsically involves multiple layers of security: hardware, software, in-transit data, storage, network, application, and etcetera. The importance and interplay between these layers are highly contextual. Overall IoT security design must take this fact into account
3. IoT security can only be as strong as its weakest point. Significant attention is often paid towards securing a mobile phone without acknowledging what happens within the sprinkler control or car key applications that reside on it
4. Complex IoT devices (for example, industrial equipment, connected cars) are the most difficult IoT environments to secure. For example, the consequences of a hacked connected car can be substantially more detrimental compared to that of a connected electric meter or refrigerator

5G gives hackers an extended territory to penetrate networks, including, but not limited to mobile edge attacks. In addition, computing systems in home or enterprise settings can become a target for a focused attack—from IoT-enabled home devices to computers at the edge, and the data center or cloud. The large volume of traffic coming from sophisticated, combined attacks will make it harder to combat the attack without sophisticated security solutions.

This paper discusses the threat surface created by the introduction of IoT in the following sections. Comprehensive IoT security needs to consider security at many levels, as illustrated in Figure 3.6. The devices and network/transport may be the areas of primary focus today, but from a revenue standpoint, the platforms, applications and services will be key. While the scope of this paper is focused on IoT security in the context of 5G, it is worthwhile to take a brief look at the comprehensive landscape of IoT security.[5]

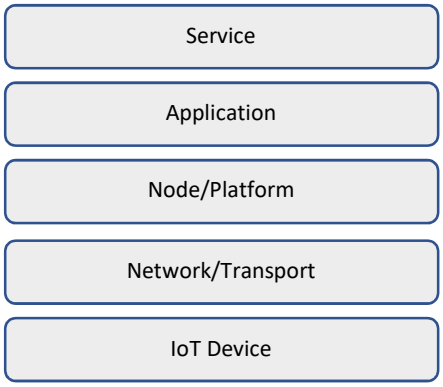

Figure 3.6. IoT Security Levels.

- **IoT Devices** - Many IoT devices will likely reside in exposed and vulnerable environments and Tampering may occur with device-resident sensitive data. Malicious updates of device firmware and OS pose a significant problem

[5] https://www.ericsson.com/en/white-papers/iot-security-protecting-the-networked-society.

- **Network/Transport** - Network connectivity enables secure interaction of devices or apps with serving network nodes. To secure this interaction, secure identification/authentication (credentials) and data transport are needed. IoT network connectivity must handle billions of devices, involving heterogeneous access technologies and capillary networks, cost effectively

- **Node/Platform -** IoT platforms must ensure the security of data and control commands. In addition, platforms are also responsible for ensuring isolation between devices, users, third-party apps, and platform-based services. Privacy concerns are one of the main inhibitors to adoption

- **Application** - Applications can be seen as a combination of micro services used to create a service. These applications can be statically located or dynamically migrated to the environment that is optimal for their realization. The security of the applications will be the result of the application code itself and the platform it is using. In cases where applications can migrate, it is important that migration between platforms happens securely

- **Service** - IoT enables a multitude of new services. A key new service in which IoT will play a significant role, and where ensuring security is of paramount importance, is connected cars. For large groups of connected vehicles traveling at high speeds, safety will remain a priority. If network connectivity is lost, either because of malfunction or jamming, backup mechanisms that the service can fall back on are necessary. There are many other sensor-based services with varying degrees of importance that can be enabled by IoT. The path to securing various IoT services will need to consider their uniqueness, as well as impact of the service itself.

3.3 5G THREAT SURFACE FOR MASSIVE IOT

MIoT spans a wide variety of new and exciting opportunities, such as autonomous vehicle communications, smart grids, highway and traffic sensors, drone communications, medical sensors and AR/VR. The MIoT market opportunity's unique requirements and cybersecurity considerations are directly influencing 5G architecture. Two examples are 5G's use of edge computing and its support of Ultra Reliable Low Latency Communications (URLLC).

An earlier section of this paper provided a high-level description of a scenario where hackers exploit zero-day vulnerabilities in MIoT devices to launch a DDoS attack on a 5G RAN. These hackers could be people simply looking to disrupt a mobile network, or they could be a nation-state attacking all of the mobile operators in another country. Figure 3.7 illustrates this scenario.

.

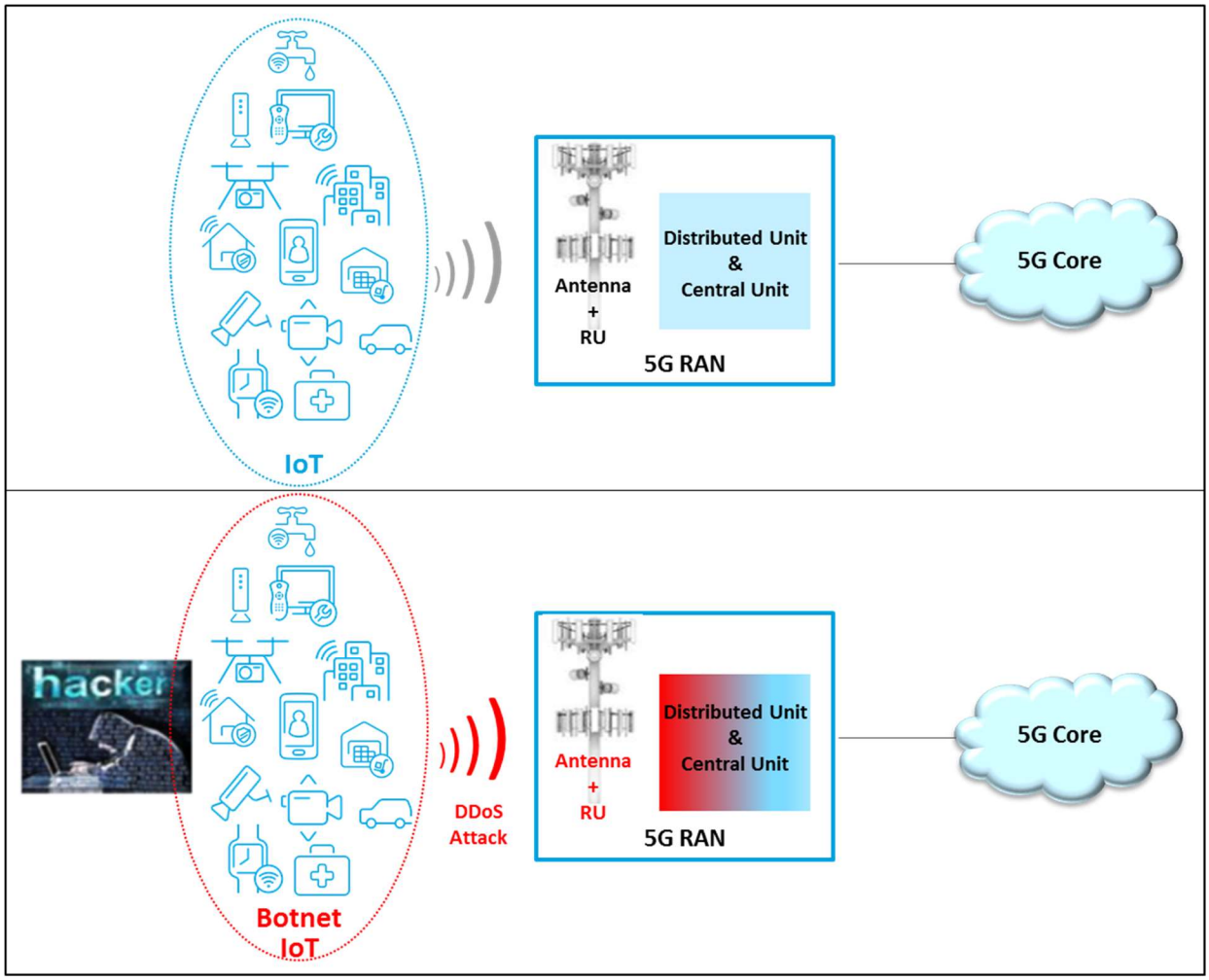

Figure 3.7. The Network vs. the Hacker.

Figure 3.8 is a high-level view of the 5G threat landscape. The different 5G entities and segments, such as UEs, the RAN, the core network and operator-hosted or third-party applications and services, could be targets from different threat actors. For example, hacktivists, organized crime, state-sponsored and insider-threat actors could launch cyber-attacks on 5G networks with the aims of theft of service, fraud, theft of customer identities and information, causing brand reputation damage, or making 5G NFs and services unavailable. This section describes the various threats and attacks that may target different 5G network elements and segments.

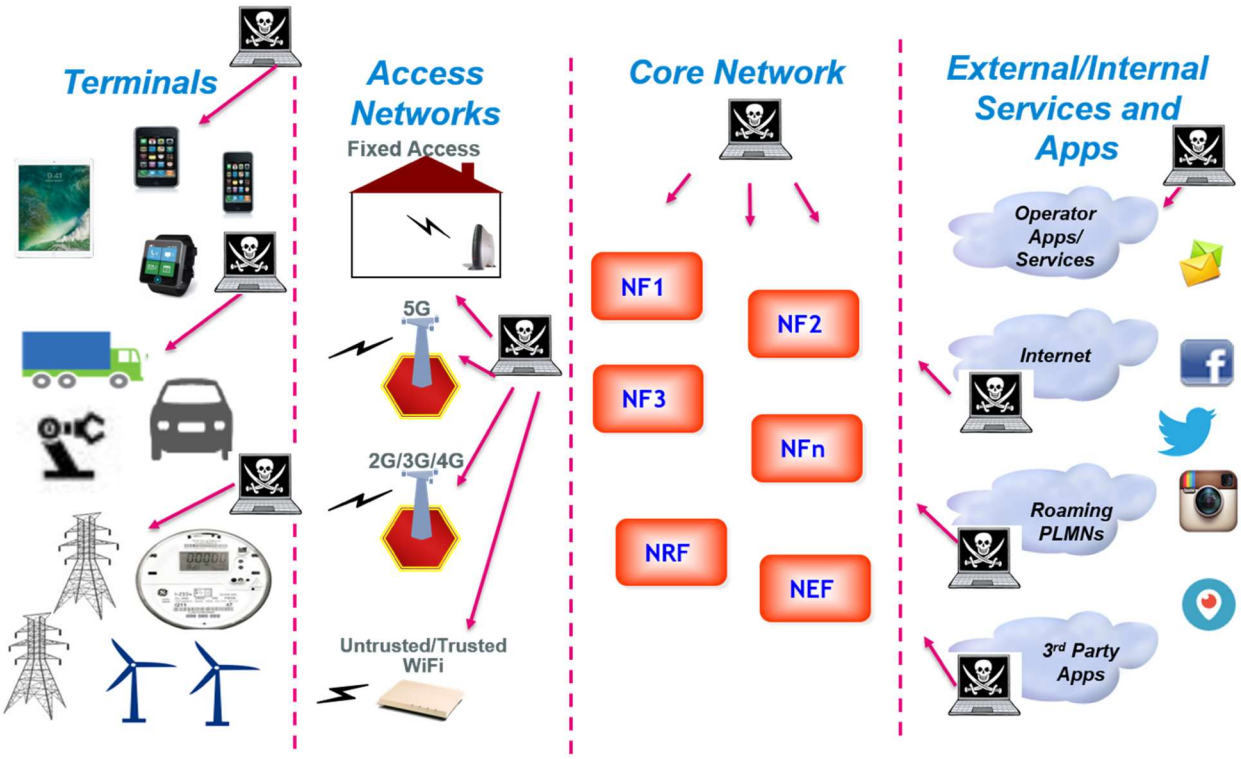

Figure 3.8. The 5G Threat Landscape.

3.4 UE THREATS

The widespread use of smartphones, diverse device form factors, increased data rate, wide variety of connectivity options (for example, Wi-Fi, Bluetooth, 2G/3G/4G) and the popularity of open source architecture, are all factors that make the UE a prime target for attacks in 5G networks. The different attacks targeting UE in 5G networks can be classified into four main categories:

1. **Mobile to Infrastructure**: A mobile botnet of a large number of infected devices controlled by an attacker's Command and Control (C&C) servers launch DDoS attacks on 5G infrastructure aiming to make 5G network functions and services unavailable
2. **Mobile to Internet**: A mobile botnet of a large number of infected devices controlled by C&C servers launch DDoS attacks on public websites through the 5G network
3. **Mobile to Mobile**: A number of infected devices launch attacks on other mobile customers with the aim of causing a denial of service or spreading of malware (for example, viruses, worms, rootkits)
4. **Internet to Mobile**: A malicious server on the internet targets each UE with malware embedded inside apps, games or video players from untrusted app stores. Once downloaded and installed, the malware enables the attacker to steal stored personal data on the device, further spread the malware to other devices, or control the device for launching attacks on other devices and networks

3.5 RAN THREATS

The fact that 5G will support many different access networks including 2G, 3G, 4G, and Wi-Fi means 5G will likely inherit the security challenges of those access networks. This section describes the main vulnerabilities and threats associated with the RAN.

In recent years, a large body of literature has revealed numerous security and privacy issues in 4G mobile networks. Most of the published attacks at the 4G RAN layer involve RBSs or IMSI catchers to target IMSIs during the UE's initial attachment procedure to the network, or paging attacks using the IMSI paging feature. In such attacks, the obtained information about particular IMSIs may be used later for other types of attacks. Fortunately, the 5G technology and standards are expected to address the known threats at this layer at all access types, including the licensed RAN and unlicensed Wi-Fi. For example, 5G will not transmit an unencrypted IMSI.

5G systems and networks will use Multiple -Input Multiple-Output (MIMO) antenna arrays and beamforming. In addition to other spectrum, many 5G systems will operate in millimeter wave (mmWave) spectrum. It is expected that mmWave will be as secure as other spectrum bands. The data and signaling transmitted and received at the radio layer are expected to be appropriately encrypted and integrity protected at higher layers, whenever possible.

3.5.1 ROGUE BASE STATION THREAT

One of the threats that face different mobile networks, potentially including 5G, is the Rogue Base Station (RBS) threat. The RBS masquerades as a legitimate base station to facilitate a Man-in-The-Middle (MiTM) attack between the mobile user equipment (UE) and the mobile network. An attacker can theoretically use an RBS to launch different attacks on mobile users and networks. An RBS attack typically involves one or more software-defined radios without access to, or membership within, an operator's core network, limiting its possible impact to early communication phases prior to UE authentication. These attacks might include stealing user information, tampering with transmitted information, tracking users, compromising user privacy or causing DoS for 5G services. However, practical avenues of attack remain controversial and possibly restricted, given the limited range/coverage of practical, portable software-defined radios, typical physical security constraints, and other factors. A detailed analysis of this threat is underway in other industry bodies, and would exceed the limited space available in this whitepaper.

The RBS threat has existed since GSM networks and continued to evolve and persist with the evolution of mobile networks. 5G networks are expected to introduce several security enhancements over 4G and legacy networks, as described in section 2. Despite these security enhancements, 5G networks could still be a target of RBS-based threats. The following threat vectors serve as examples:

- An attacker can exploit 5G/LTE interworking requirement to launch a downgrade attack
- A compromised 5G small cell can create an RBS threat to 5G networks and customers
- An attacker can exploit a lack of gNB authentication in an idle mode to force the user to camp on an RBS which could lead to a denial of services (such as public safety warnings, incoming emergency calls, real-time application server push services, and etcetera)

3.6 SUBSCRIBER PRIVACY THREATS

Subscriber privacy has always been a top concern for the mobile industry. As the 5G era begins, subscriber privacy is an even higher priority because of growing attention from media and regulators. For example, there have been several news stories related to allegations of mass surveillance. Reports have also emerged regarding unknown RBSs tracking users in major cities and performing suspicious activities.

Figure 3.9 shows different potential exposure points for compromising subscriber privacy in 5G networks, using protocol attacks, malware attacks on 5G network functions (NF) and insider threats.

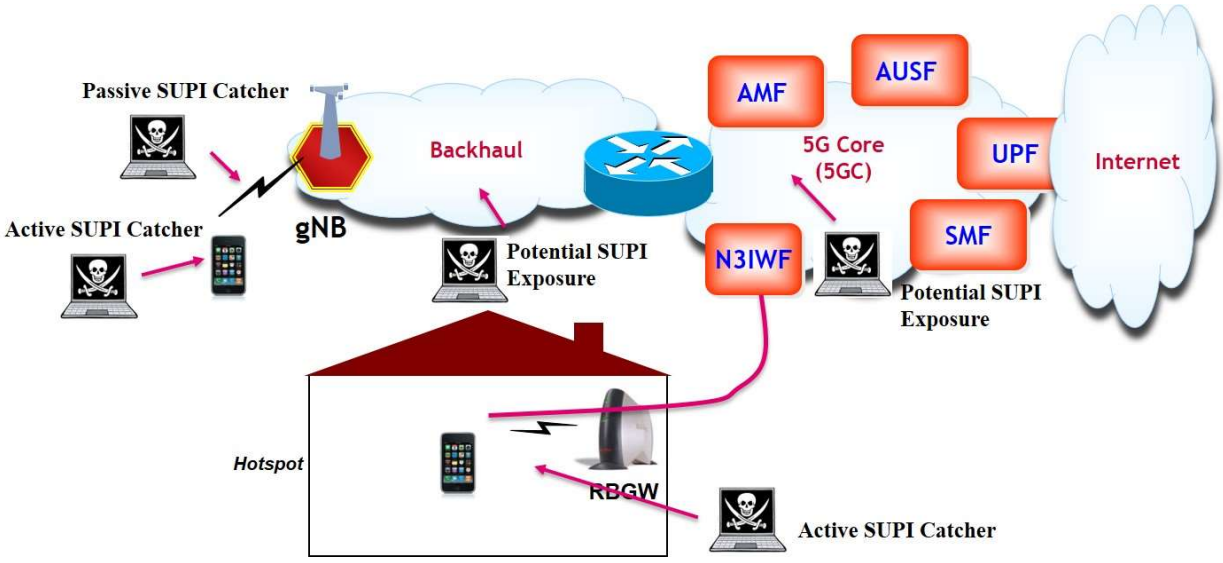

Figure 3.9. Exposure Points for Compromising Subscriber Privacy.

Note that there have been recent press reports about unknown individuals and groups operating IMSI catchers. Here the attacker takes advantage of an oversight in the original 3GPP mobile standards. These standards require a device to authenticate to the network, but do not require networks to authenticate to devices. This allows IMSI catchers to impersonate base stations and capture IMSIs. Such devices can also prevent UEs from using encryption during calls or force UEs to use easily breakable encryption, allowing eavesdropping. 5G standards mitigate these vulnerabilities through the use of SUPI and SUCI (as described previously). SUPI is encrypted using the network operator's public key, which allows UE to authenticate the network to which it is connecting. However, advanced attackers may be able to force UEs to communicate in non-5G mode (for example, 3G), thus nullifying these mitigations.

The attacks on user privacy could lead to exposure of a user permanent identifier (for example, SUPI) to enable unauthorized tracking of user movements and activities. With the introduction of vertical applications (for example, SmartX, eHealth, and etcetera) in 5G, compromising user privacy can lead to significant damages and losses to both operators and users.

3.7 CORE NETWORK THREATS

Due to their IP-based service architecture, 5G networks could be vulnerable to IP attacks common over the internet, including DDoS attacks. Also, a large number of infected mobile devices, controlled by malicious

Command and Control (C&C) servers, can launch both user plane and signaling plane attacks on 5G core network functions. As a result, critical services are degraded and rendered unavailable for legitimate users.

The Access and Mobility Management Function (AMF), the Authentication Server Function (AUSF) and Unified Data Management (UDM) are the main network functions in 5G. The AMF provides UE authentication, authorization and mobility management services. The AUSF stores data for authentication of UEs, and the UDM stores UE subscription data. These functions are critical in 5G; a DDoS attack against these functions, from the internet or a mobile botnet, can potentially reduce the availability of 5G services significantly or even cause network outages.

3GPP recommends using Internet Protocol Security (IPSec) encryption for non-3GPP access. An attacker can exploit the massive number of IPSec tunnel establishment requests with a large number of infected mobile devices simultaneously launching DDoS on 5G core network functions.

3.8 NFV AND SDN THREATS

New networking paradigms—such as NFV and SDN—must be adopted to efficiently to support the new levels of performance and flexibility required for 5G networks. However, these new techniques present new threats. For example, when applying NFV, the integrity of Virtual NFs (VNFs) and the confidentiality of their data may depend on the isolation properties of a hypervisor. More generally, they will also depend on the whole cloud software stack. Vulnerabilities in such software components have often surfaced in the past. In fact, it remains a major challenge to provide a fully dependable, secure NFV environment.

Also, the 5G cloud data centers are expected to be connected through enhanced transport networks and improved networking concepts, such as SDN. SDN bears the threat of control applications potentially creating chaos on a large scale by erroneously or maliciously interacting with a central network controller. SDN introduces a separation of forwarding and control and thus introduces an interface between the SDN controller and SDN switch. This interface makes the overall system more vulnerable to attack. It could allow attacks on the integrity and confidentiality of the controller-switch communication or DoS attacks.

In addition, it could permit attacks aimed at gaining some control over switches and controllers by exploiting vulnerabilities in the protocol software or the interface configuration. However, securing such an interface is a well-known task and suitable means are readily available, for example, the use of IPsec or TLS to cryptographically protect the legitimate communication and exclude communication by malicious third parties.

3.9 INTERWORKING AND ROAMING THREATS

Roaming in 5G applies some new protocols delivering new flexibility as well as new threats compared to 4G. The following roaming considerations in the 5G architecture pertain to embedded security at the 5G roaming links.

- 5G architecture has introduced the Security Edge Protection Proxy (SEPP) node as the entity to terminate signaling messages between PLMNs through inter-exchange/roaming links
- The interconnection model will be equivalent to the SS7 or DIAMETER interconnect that exists in today's 3G and 4G networks. However, the application layer protocol (for example, HTTP/2) will support encryption on the inter-exchange/roaming links
- The embedded application layer encryption at the SEPP will provide protection against the known inter-exchange/roaming vulnerabilities that exist in SS7 and DIAMETER protocols

4. NETWORK SLICING SECURITY

Network slicing is quickly becoming one of the defining features of 5G networks today. Briefly, network slicing is the ability of the network to automatically configure and run multiple logical networks as virtually independent business operations on a common physical infrastructure. Slicing is expected to be a fundamental architecture component of the 5G network, fulfilling the majority of the 5G use cases.

Chapter 4 describes network slicing in further detail and provides deeper insights into network slicing threats and mitigation.

4.1 INTRODUCTION TO NETWORK SLICING CONCEPT AND RESULTING SECURITY THREATS

According to 3GPP specifications, a network slice is a complete logical network (providing Telecommunication Services and Network Capabilities) including Access Network (AN) and Core Network (CN). Network slicing enables the management of multiple logical networks as virtually independent business operations on a common physical infrastructure. In practice, this corresponds to the idea that the mobile network could be partitioned into a set of resources that might be virtual. Each one is called a "slice" that can be allocated for different purposes. For example, a slice can be allocated to a mobile virtual network operator (MVNO), an enterprise customer, an IoT domain, or some other convenient set of services (for example, mobility as a service). A network slice extends the access point name (APN) concept used in the mobile network today.

One of the key new aspects of the 5G architecture is segmentation through network slicing. The concept of segmentation of a carrier network, and application of policy with that segmentation as a foundation, is standard in mobile networks. However, in 5G, it is improved even further. New trust boundaries are created both in the packet core and in places where the packet core interacts with businesses and governments served by the 5G network. Following is an overview of network slicing and some of the security aspects.

In 3GPP, network slicing is defined in TS 23.501. A network slice is defined within a Public Land Mobile Network (PLMN) and includes the Core Network Control Plane and User Plane Network Functions, as well as the 5G Access Network (AN). The 5G AN may be:

- Next Generation (NG) RAN described in 3GPP TS 38.300
- Non-3GPP AN where the terminal may use any non-3GPP access to reach the 5G core network via a secured IPSec/IKE tunnel terminated on a N3IWF

TS 23.501 further defines Network Function (NF), Network Slice and Network Slice Instance (NSI) as follows:

- NF: A 3GPP-adopted or 3GPP-defined processing function in a network, which has defined functional behavior and 3GPP defined interfaces. A NF can be implemented either as a network element on a dedicated hardware, as a software instance running on a dedicated hardware or as a virtualized function instantiated on an appropriate platform, such as on a cloud infrastructure
- Network Slice: A logical network that provides specific network capabilities and network characteristics.

- Network Slice instance: A set of NF instances and the required resources (for example, compute, storage and networking resources) that form a deployed Network Slice
- NSI ID: an identifier for a Network Slice instance

Based on the current 3GPP specs TS 23.501 Release 15, the 5G core supports the following architecture for virtualized deployments:

- A NF instance can be deployed as fully distributed, fully redundant, stateless and fully scalable NF instance that provides the services from several locations and several execution instances in each location. It implies that for a typical cellular services network, different NFs deployed using the network slicing may be fully geo-redundant
- A NF instance can also be deployed such that several NF instances are present within a NF set provide fully distributed, fully redundant, stateless and scalability together as a set of NF instances. With this approach, for a small cellular network, network resiliency can be obtained at a single location using local redundancy with replicated virtualized NFs within a NF set

Figure 4.1 shows a simplified example of 5G network slicing concept.[1]

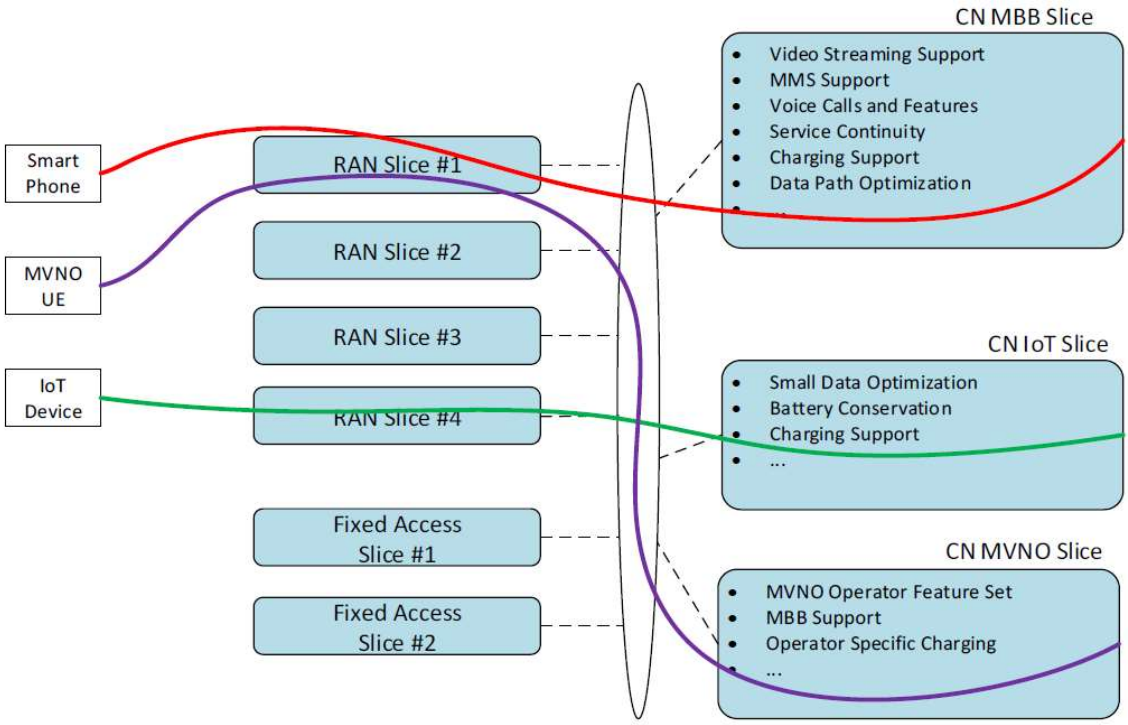

Figure 4.1. Network Slicing Concept.

As shown in Figure 4.1, different slices offer different services. Each service may have distinct performance constraints, and/or different security requirements. The fact that certain aspects of security are emphasized in one slice does not mean similar security policies are replicated for all slices. Consider the following examples:

- One slice may require very low latency. Applicable security protocols (key derivation, key management) need to be aligned with the primary slice service requirements (for example, low latency)
- A slice intended to serve applications where privacy is paramount may require frequent reallocation of temporary identities
- A slice servicing remote device that must minimize battery consumption may need to cut down frequency of re-authentication to a minimum

Figure 4.2 illustrates some of the key architectural elements of 4G and 5G. A 4G network might migrate to 5G using a model called Non-Stand Alone or NSA. This allows some of the 4G control capabilities to be deployed with 5G user plane. Sometimes there is a clean transition to 5G, commonly referred to as 5G Stand Alone.

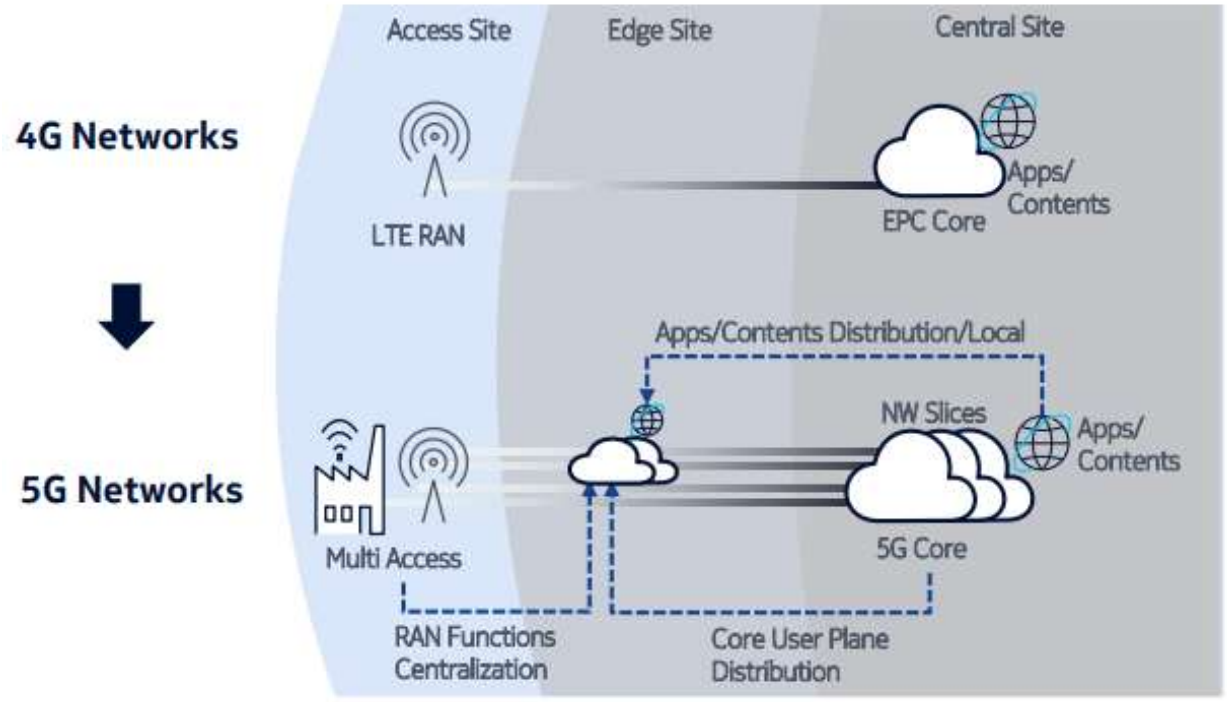

Figure 4.2. 4G --> 5G.[6]

With 5G core architecture, the plan is to deliver the whole network as a service. The 5G core network is re-designed based on a service-oriented architecture by parsing everything into detailed functions and sub-functions. For example, the Mobility Management Entity (MME) functionality has been redistributed into precise families of mobility and session management network functions. Functionalities offered by 4G MME such as registration, reachability, mobility management and connection management services are offered by a new 5G general network function called Access and Mobility Management Function (AMF).

Additionally, session establishment and session management, also formerly part of the MME, are now services offered by a new network function called the Session Management Function (SMF). Furthermore,

[6] Source – Nokia.

packet routing and forwarding function (currently performed by the SGW and PGW in 4G) are now realized as services rendered through a new network function called the User Plane Function (UPF). This is achieved with the support from 5G core technologies such SDN and NFV, which are software-based solutions. With this granular approach, more resilient networks may be realized.

Figure 4.3 illustrates the concept of network slicing, where a single physical network can be partitioned into multiple virtual networks. This architecture enables operators to offer optimal support for different types of services for different types of customer segments. The key benefit of network slicing technology is it enables operators to provide networks on an as-a-service basis, which enhances operational efficiency and resilient network services.

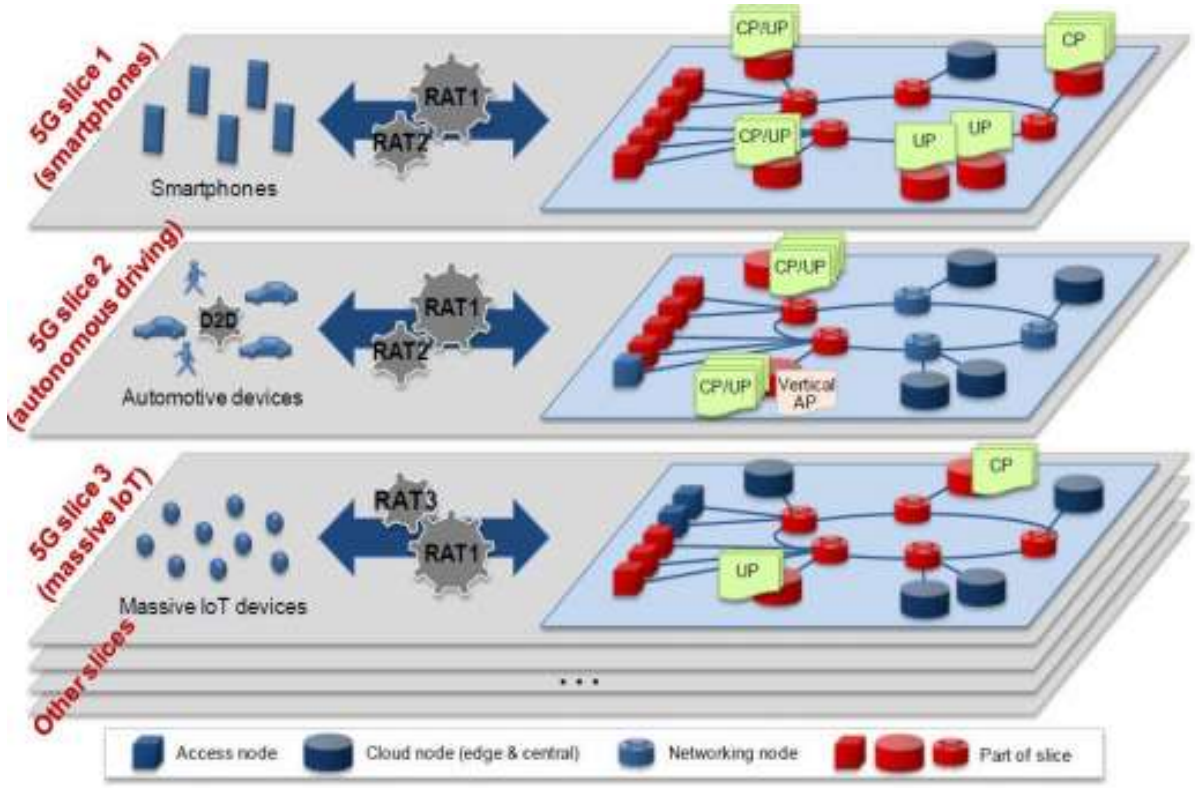

Figure 4.3. Network Slicing in 5G.[7]

5G leverages a service-based architecture; NFs, while they are being set up, can register to the network. This functionality is controlled by the Network Repository Function (NRF). Such a solution allows many improvements over the current 3G/4G network functions. Service selection was limited and major integration was required to make NFs visible to other peer nodes before any services could be provided by the network.

[7] Source-Nokia

By running network slices in this service-based architecture, operators can select NFs using multiple different criteria, such as geographical proximity for low-latency services, or required capacity/load. There could be also other non-technical selection criteria, such as the cost of the service.

This makes 5G networks very flexible. They can provide exactly what is required because NFs can be established and removed on a per-need basis and used simultaneously by multiple different slices. Also, network Operations, Administration and Management (OAM) can be simplified and made more flexible. Service providers can utilize automated tools to provide the network services with the predefined redundancy, capacity and other capabilities. Additionally, they can multiply NFs to the same or multiple locations as needed. Automation can also optimize the unnecessary need for extra hardware and use it for other purposes (like analytics or data mining) while the regular network load is low. Generally, some of these tools and capabilities have been available in the network prior to 5G.

Figure 4.4 further elaborates the concept of 5G network slicing functions with three slices, namely IoT Slice, eMBB Slice & Vehicle-to-Everything (V2X) slice.

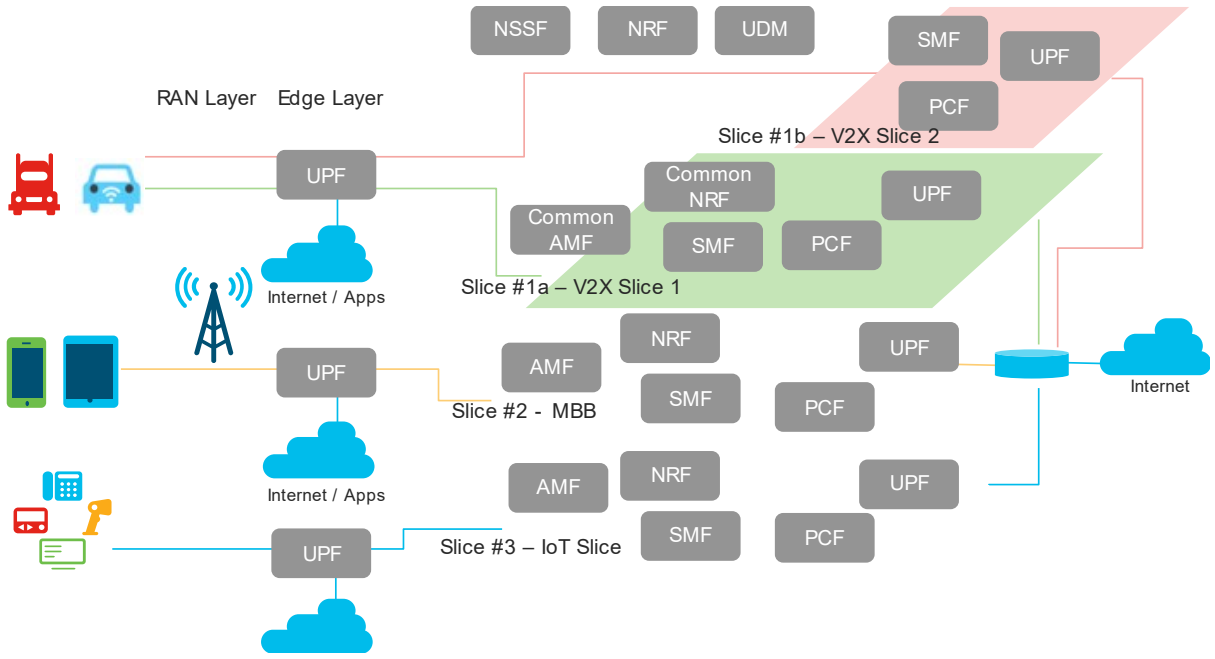

Figure 4.4. 5G Segmentation by Service.

In Figure 4.4, there are common resources and components shared between all the slices, such as the NSSF (Network Slice Selection Function). The rest of the slices may have individual resources assigned, such as AMF, SMF, PCF layers that are dedicated functions catering to specific slices.

4.1.1 THREATS IN NETWORK SLICING

One of the primary factors leading to threats in network slicing is improper isolation between the network slices (Inter-Slice Isolation) and improper isolation between the components of the same slice (Intra-Slice Isolation).

As shown in Figure 4.5, a threat could be migrated between the slices if one of the devices in the IoT slice gets infected by a malware using a vulnerability in the IoT device. As a result, critical slices (such as V2X slice) will also be impacted.

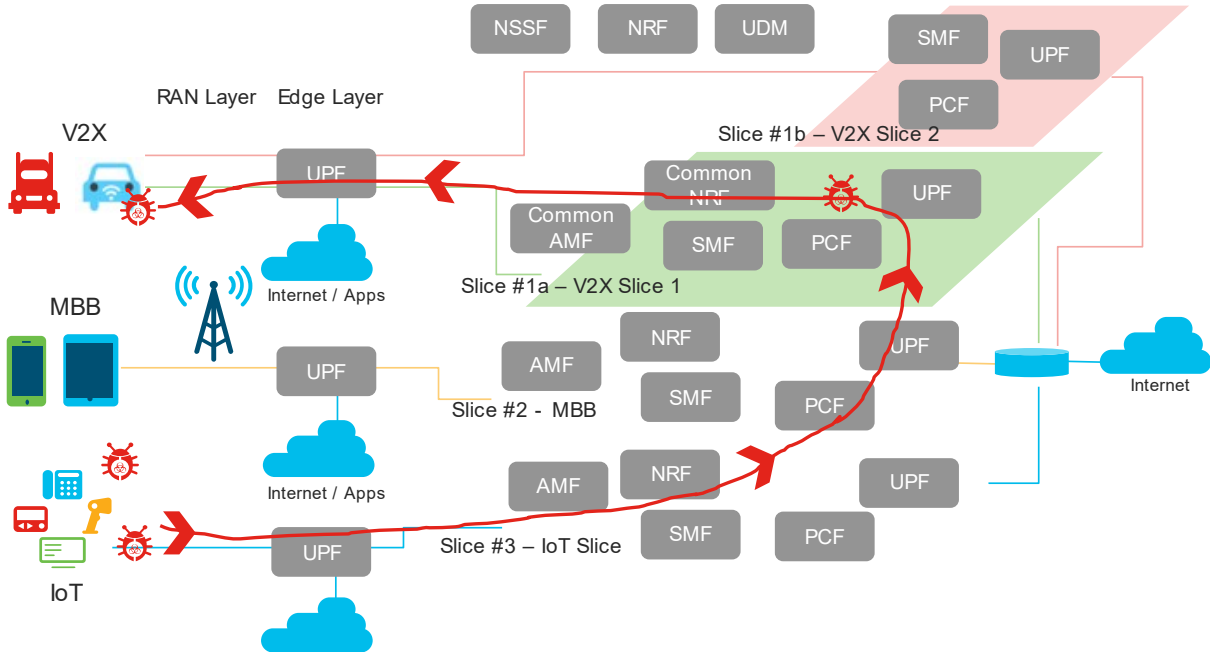

Figure 4.5. 5G Segmentation Threat in Motion – Part 1.

In the scenario from Figure 4.5, the attack could be multi-factored by allowing the malware to have the ability to deplete the resources of the slice, therefore causing DoS (Denial of Service) to the actual subscriber. An attacker may also exhaust resources common to multiple slices, causing denial of service or service degradation in other slices as well. This leads to severe degradation in the offered network services.

As a cloud native architecture, 5GC (5G Core) has all the functions virtualized that provide the added flexibility required for network slicing. However, this leads to another threat vector. Side channel attacks, coupled with improper isolation between network slices, leads to data exfiltration. This is critical in sensitive parts of the mobile network such as billing, charging and subscriber authentication layers.

As shown in Figure 4.6, if the slices and the components within the slice are not adequately isolated, the attacker could access other slice components using the infected device or endpoint in another slice. Figure

4.6 shows the infected device allowing the attacker access to the slice resources. Ultimately, the other slices are exposed and data exfiltration proceeds to an external server (a C&C center, for example). Once the attacker gathers all the network's information behind the firewall, they could launch an attack on subscribers based on the leaked information. Furthermore, the attacker could use the information for fraudulent financial gains.

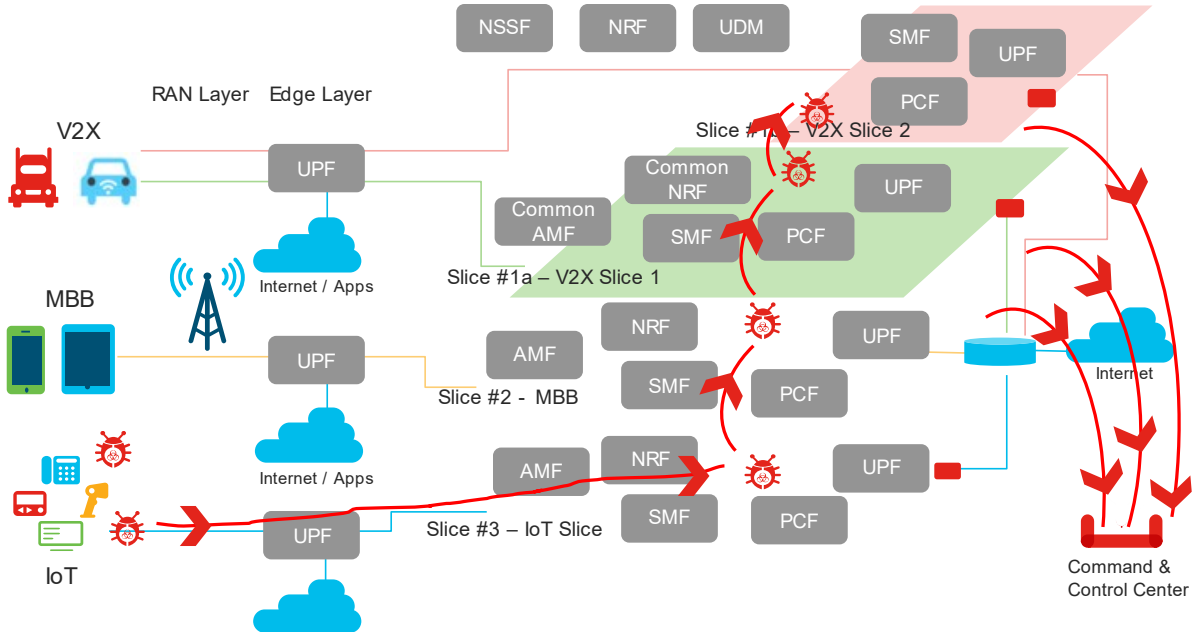

Figure 4.6. 5G Segmentation Threat in Motion – Part 2.

Network slicing allows operators to offer customized services to customers. It is possible for 5G systems, based on operators' policies, to provide standardized APIs to create, modify, delete, monitor, and update the services of network slices. Slice management also contains critical threat vectors if not secured. Additionally, as per 3GPP standards specifications, the management interface between the Network Slice Management Function (NSMF) and the Communication Service Management Function (CSMF) or between Communication Service Provider (CSP) and Communication Service Customer (CSC) is specified. Furthermore, interfaces are also specified for the operation phase of management aspects of a Network Slice Instance (NSI), supervision, and performance reporting.

Without securing these network slice management interfaces, attackers may gain access to the management interface. That access allows attackers to create network slice instances requiring significant network resources or a large number of network slice instances. As a result, the network resources are exhausted, leading to Denial of Service (DoS) attacks. Attackers could incite fraudulent activities, like false charging, by replaying management messages. An attacker may also eavesdrop on the transmission of supervision and reporting data and extract sensitive information to execute attacks of running network slice instances.

4.1.2 THE MITIGATING THREATS IN NETWORK SLICING

Network slicing architecture that enables running multiple, logical networks as virtually independent business operations on a common, physical infrastructure requires high security layers. Specifically, layers like isolation between slices and isolation within the components of the slice. This prevents vulnerabilities from spreading to other components within the slice and between slices in case of any malicious attacks.

In Network Slicing, we can distinguish two isolation aspects: resource isolation and security isolation. The former refers to the fact that resources assigned to a network slice (such as computing, storage and networking resources) cannot be "hijacked" by another slice. Resource isolation ensures that the necessary resources of a slice remain available even in situations where other slices try to scale out and acquire additional resources, such as when a slice is a target of in a Denial of Service (DoS) attack. Security isolation refers to the property that information in one slice cannot be accessed or modified by other slices sharing the same infrastructure.

Isolation is a basic property provided by virtualization layers, such as a hypervisor. Thus, sound NFV security—including robust implementation of the virtualization layer and the overall cloud stack—can ensure both types of isolation within NFV environments. For the transport between different hardware platforms, including even distributed cloud deployments, dedicated virtual networks can be created per slice to ensure isolation where SDN allows for highly dynamic and flexible control of different virtual transport networks sharing the same transport infrastructure.

Slices may also share non-virtualized equipment. For example, base station equipment handling the lower protocol layers. Here, isolation needs to be assured by respective equipment-specific mechanisms. For instance, when several slices share a single cell, a common radio scheduler may be configurable to implement scheduling policies. This ensures each slice gets radio resources in that cell according to the service-level agreement (SLA) valid for the slice.

Each slice's role as an isolated logical network or network part clearly needs state-of-the-art network security measures. This may include measures such as: perimeter security and network zoning by means of firewalls (typically not firewall appliances, but virtual firewalls), separation of different traffic types, intrusion and anomaly detection, the use of cryptographic traffic protection, etcetera. Slices allow security measures to adapt to the individual needs of the application(s) supported by that slice. For example, an IoT slice may provide a different choice of authentication and/or encryption mechanisms than an eMBB slice.

A quarantine slice should also be enabled to ensure intra- and inter-slice security. The implementation of the quarantine slice is made possible in transport by segment routing and in the data center by segmentation technology. This architecture joins these two segmentation methods and delivers an end-to-end, segmented network delivering visibility and agility in threat detection and management.

Together, segment routing and network wide policy enforcement deliver Software Defined Segmentation. Software defined segmentation makes it possible to enforce the access policies for users, applications and devices, which can apply to IoT devices, M2M based devices and enterprise network devices as well.

In the data center part of the 5G architecture, software defined segmentation leverages segment routing. When software defined segmentation is used, the security group tags can be defined and managed by an identity and segmentation policy controller. The controller can also share group information with other group-based policy schemes, allowing segmentation of the traffic and restriction of communication between

defined network interfaces. This security approach shifts the network security away from reliance on long lists of IP address to a flexible, automated model which is better managed and more effective against new and expanding threat vectors

Slice Isolation can be provided using namespaces ACL, TAGs and the software defined segmentation controller. Additionally, visibility on the slices can use applications which gather telemetry and deliver secure outcomes. These secure outcomes include the following: 1) secure the core VNF components using tokens & Istio; 2) secure the API interfaces using WAF (Web Application Firewall), and; 3) enforce application policies using a white listing application behavioral approach.

Understanding the normal behavior of an IoT device helps determine when that device is acting inappropriately. Systems can identify areas of concern by detecting anomalies or changes in the behavior of IoT devices. The anomaly could be the result of anything from a software update to a malicious activity.

Finally, another challenge of monitoring IoT devices involves encrypted traffic. Vendor-specific implementations for solving the security issues around encrypted traffic exist, but the industry needs an evolved approach from today's man-in-the-middle techniques. Enabling malware detection within encrypted communications is critical to understanding the behavior of traffic for deeper analysis.

Attack prevention is one of the most important aspects of security and today's solutions can assist by delivering security from the cloud, blocking malicious destinations before a connection can be made. An approach leveraging DNS-based security identifies internet communications from every device on the network based on name resolution, blocks malicious domains, and breaks C&C callbacks.

4.2 SECURITY ISSUES FOR NETWORK SLICING – A DEEPER DIVE

The aim of 5G is to provide an end-to-end infrastructure capable of delivering consistent services that support multiplicity of use cases for a variety of users with widely different characteristics and requirements. To achieve this, 5G must be capable of supporting, for example, very high data rates, vehicular speeds, ultra-low latencies and high device densities across an array of radio access technologies. Designing a 5G network that can support such demanding performance requirements based on a single set of standardized network functions would normally be prohibitively complex and expensive.

In 5G architecture, the solution to this problem is provided by network slicing: a technology that provides simultaneous, diverse user experiences with future-proof scalability and flexibility on top of a common physical infrastructure.

Network slicing would allow a 5G network to be partitioned logically into multiple virtual networks (slices) where each slice can be optimized for a set of requirements to serve a specific vertical application. This results in a high degree of flexibility, enabling multiple use cases to be concurrently active across the same physical infrastructure. Network elements/functions are configured and reused in each slice to meet a specific set of requirements. Slicing results in end-to-end virtual networks, comprised of both the core, RAN and UE components. Each slice has specific architecture, engineering, provisioning, security policy and O&M aspects depending on user needs and/or SLAs.

Let's examine some issues in further detail:

4.2.1 ISSUE 1

5G systems are fundamentally service-oriented. Operators could, for example, offer a network slice for each enterprise customer, not unlike the per APN offer for an enterprise today. End-to-end virtual network slices cater, by design, to different industry/business verticals with correspondingly different security needs. For instance, remote health care requires resilient security while IoT may only require lightweight security. Different E2E security capabilities may include strength of security algorithms, ways to derive and negotiate secret keys, and mechanisms for protecting confidentiality and integrity, etcetera. It is reasonable to expect differentiated security for different services. In addition, within a virtual network slice, security capabilities could be distributed among multiple infrastructure vendors involved in instantiating an E2E service.

4.2.2 ISSUE 2

Differentiated services running over individual network slices over a common infrastructure need isolation. For virtual network slices—each handling a different category of application that involve flexible resource orchestration—there is a clear need to isolate slices from each other. Resources (CPU, memory, storage, and etcetera) in use by one slice should not be accessed by infrastructure components serving other slices. For instance, patients in a health care slice would prefer to only allow their doctors to access their health data. Such data should never be accessed by users in other slices. Note that the isolation criteria is equally applicable to virtual network slices with the same category of application. For instance, two healthcare enterprises A and B may be served by the same virtual network slice (or same category) but would require complete isolation of data from each other. Only isolation guarantees that users would be willing to store and access private data on cloud servers, without concern about privacy/security risks.

An attacker may adversely impact overall availability of resources common to multiple slices by exhausting specific resources in one slice, even with isolation. This would amount to a DoS attack with effects that permeate beyond the specific slice under direct attack.

4.2.3 ISSUE 3

Slice Managers are responsible for dynamically creating and destroying instances of network slices. In addition, Slice Managers instantiate them on available physical host platforms. These hosts may be deployed across disparate geographic locations spanning the operator's network domain. It is possible that the manager or host (or both) may have been compromised by hackers impersonating the authentic entity.

4.2.4 ISSUE 4

Flexibility of slice security architecture is another key requirement. For each industry vertical, the speed and extent of change in individual business environments are likely unique. This could potentially require corresponding adjustments (for example, rapid response) in the overall security profiles of the serving network slice. In other words, slice-specific E2E security capabilities would need to rapidly align with evolving business needs. This adds to the complexity of the overall slice security architecture requiring early attention in the overall design process.

4.2.5 ISSUE 5

5G architecture, especially 5G network slicing, is fundamentally designed to allow a wide variety of apps to thrive and grow. Apps served over E2E network slices would exchange a plethora of user device and network-specific information (user IDs, device IDs, user preferences, user locations, sensitive user information, financial/billing data, and etcetera). Potential erosion of privacy due to leakage of sensitive information between slices that coexist on the same infrastructure must be adequately addressed.

4.2.6 ISSUE 6

Services provided over 5G network slices are likely to be governed by new trust models. User authentication may not be the sole responsibility of the network in 5G. It is possible for user authentication to be a cooperative activity between the 5G network and the application service-provider (for example, Netflix). Unlike 4G, this model requires that 'trust' is first established between the service-provider and the network independently of users. Different choices for user authentication may be appropriate based on the category of vertical being served and/or operative SLA governing a specific slice. With network-only user authentication, service providers can task networks for service authentication. Here users can access multiple services after completing a single authentication. The concept service IDs, recognized by the network, may come into play in this context. Service-provider-only relies on authentication by vertical industries to exempt devices from radio network access authentication. Joint/hybrid networks control network access, service-providers control service access. The trust model is illustrated in Figure 4.7.

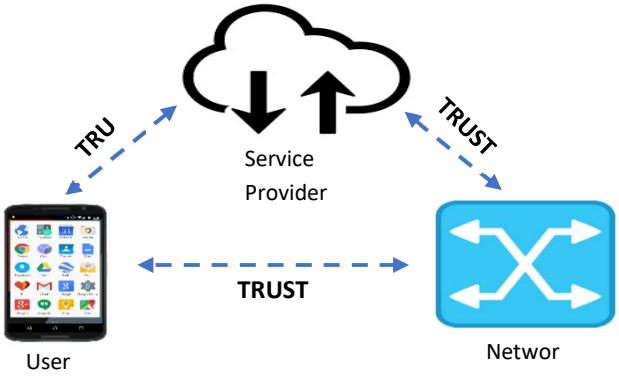

Figure 4.7. 5G Trust Triad.

4.2.7 ISSUE 7

Both software and hardware infrastructures are expected to run in a multi-vendor environment. Mitigation of unauthorized access to network resources will require special attention to identity management. In addition, designing E2E security chains could reduce reliance on individual link (hop-by-hop) security and simplify overall security management.

4.2.8 ISSUE 8

Another issue in the context of identity management is diversified identities. In 5G, user-based identities would coexist with device-based identities. This would allow users to determine which device can access which service by combining device identities (globally unique physical identity) with service identities. For example, such user-based control would allow devices belonging to the same user to share user's bandwidth quota among each other dynamically.

4.2.9 ISSUE 9

aCertain delay-sensitive applications (for example, V2X, remote surgery) require high-security and high reliability service with Quality of Service (QoS) guarantees in addition to being ultra-low-latency. These conditions are significantly more demanding than anything commonly encountered today. To address such new challenges, 5G security, including slice security, may need to be optimized for efficient lightweight operation compatible with mobility management mechanisms capable of serving vehicular speeds.

4.2.10 ISSUE 10

Side channel attacks occur when an attacker is capable of gathering actionable information about cryptographic secrets by observing the implementation of a platform (for example, power consumption, run time, etcetera) and use that information to induce faults or modify the cache. Side channel attacks do not necessarily require detailed information on the platform/system being attacked. Many side channel attacks rely on statistical analysis of platform metadata that is typically exchanged (or available) in the clear. Network slices are susceptible to side channel attacks just like regular (non-virtual) platforms.

4.2.11 ISSUE 11

It is obvious that various network slices will share common physical and virtual infrastructure, such as servers, memory, network, and storage—particularly with NFV in the underlying 5G infrastructure. In such cases, the resource reservation and its isolation for each slice could be variable in nature, possibly sharing a set of common pool of resources. In such cases, DoS/DDoS type attacks on one or more slices can impact other slices indirectly in terms of compute, network and storage aspects. The issue could be more severe in the event the operator employs automated MANO models for auto scaling based on perceived resource load (for example, CPU usage) on the VNFs, quickly aggravating the problem. Hence, the infrastructure should provide adequate resource isolation between slices when common resources are shared across various slices.

5. 5G THREAT MITIGATION CONTROLS: IOT, DDOS ATTACKS AND NETWORK SLICING

Network slicing is the ability of the network to automatically configure and run multiple logical networks as virtually independent business operations on a common physical infrastructure. Slicing, although sparsely used today for Enterprise use cases, is expected to be a fundamental architecture component of the 5G network, fulfilling the majority of the 5G use cases.

As per the 3GPP specifications, the network slice is a complete logical network (providing Telecommunication Services and Network Capabilities) including Access Network (AN) and Core Network

(CN). Network slicing enables the management of multiple logical networks as virtually independent business operations on a common physical infrastructure. In practice, this corresponds to the idea that the mobile network could be partitioned into a set of resources that might be virtual. Each one is called a "slice" that can be allocated for different purposes. For example, a slice can be allocated to a mobile virtual network operator (MVNO), an enterprise customer, an IoT domain, or some other convenient set of services (for example, mobility as a service). A network slice extends the access point name (APN) concept used in the mobile network today.

5.1 5G NETWORK THREAT MITIGATION

This paper's introduction describes how the 5G threat surface has the widest scope and the most complex consequences due to a number of factors. This includes a widely distributed network of mobile edge compute nodes. Smaller data centers push function closer to the "edge" to serve many use cases covered in this paper, specifically, ultra-low-latency IoT use cases. One way to address threat mitigation for a network with a threat surface as expansive as 5G is to separate it into parts and examine the threats and mitigations for that part of the network. The first part, shown in Figure 5.1 and Figure 5.2, describes the threat surface of the "edge" and mitigation techniques applied at specific points in the network to solve for threats.

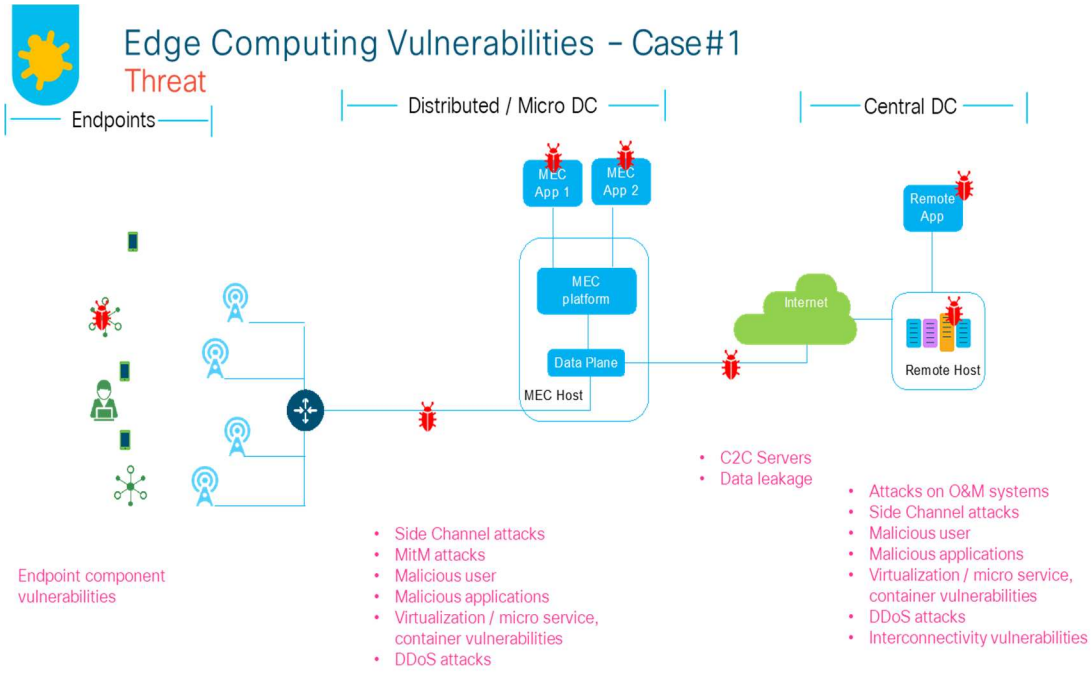

Figure 5.1. Edge Computing Vulnerabilities.

Figure 5.1 shows the need for endpoint protection (anti-malware, day 0 and day 1 protection on the endpoint). This not only protects the UE (for example, phone, iPad) but also the RAN by preventing the creation of a botnet that would attack the RAN. Commonly used techniques at the DNS protection level thwart attacks at the first step of the malware kill chain by stopping C&C communications with known "bad actors". This is just one example, but operators will have their own use cases that build on top of this foundation.

Visibility is essential to all security. This paper's introduction described the concepts of visibility and controls as the foundational elements of securing 5G. Visibility provides a continually updated picture of how the network is behaving. Threat feeds make that picture operational against known and unknown threats. Policy and segmentation are used to ensure that we know what is abnormal or an anomaly. The network is then segmented to avoid threats from spreading and compromising other functions or workloads. As illustrated in Figure 5.2, various controls are utilized at this place in the network to mitigate DDoS threats (volumetric and application based), web application threats via a web application firewall, API protection (commonly referred to as a cloud service access broker type function) and protection against malware. These controls provide for protection of the "edge."

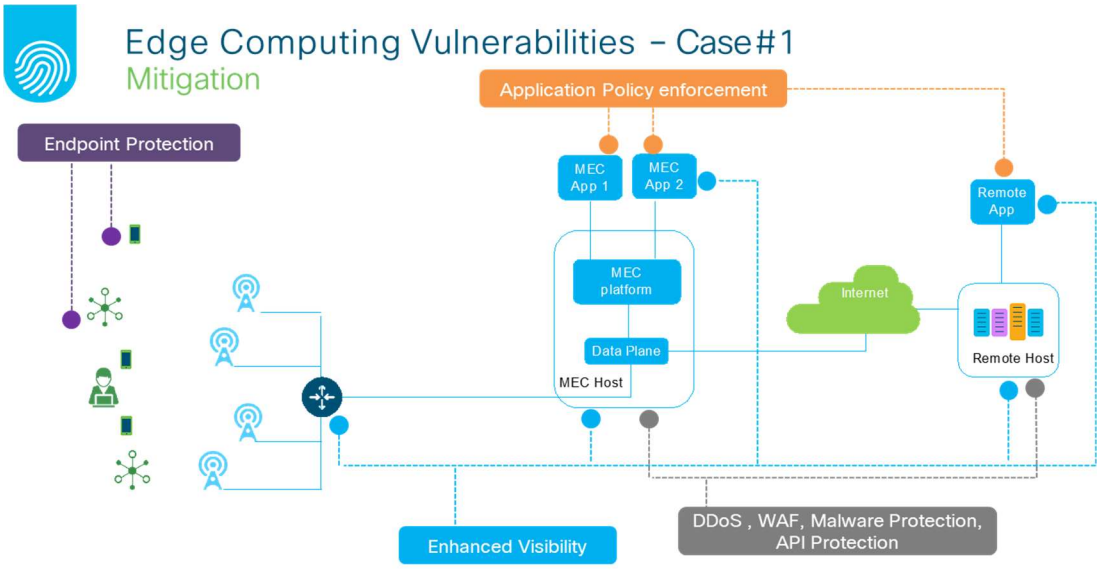

Figure 5.2. Edge Computing Vulnerabilities – Mitigation.

Figure 5.2 and Figure 5.3 describe the distributed 5G Core and the associated threat surface. 5G presents layers of orchestration, NFV, containers, micro-services and virtualized implementation of key evolved packet core functions.

Slicing is also present as part of Mobile Edge Computing (MEC) solutions. For example, there can be multiple tenants for V2X services (MEC App1, MEC App2, and etcetera) at the edge, each running on a different slice. Over time these tenants may need to communicate with each other, as well as adjacent MEC instances.

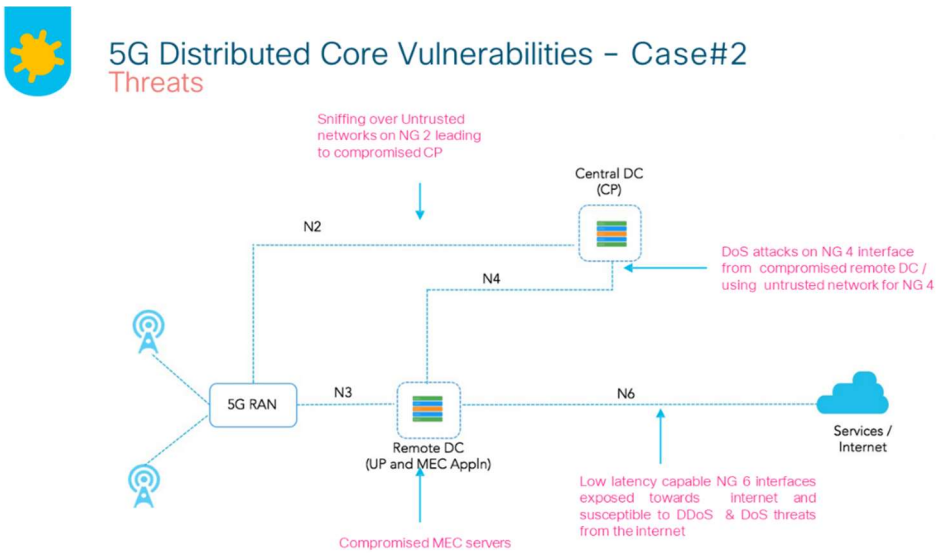

Figure 5.3. Distributed 5G Core Vulnerabilities.

Interface names in 5G change compared to 4G, with specific mapping in certain cases. One such case in 4G is the "Gi/SGi", more commonly referred to as the "N6" in 5G. Threats against the distributed core by interface include, but are not limited to:

- N2: Sniffing over untrusted networks on NG2 leading to compromised control plane
- N4 (between centralized and remote data centers): DoS attacks on NG 4 interface from compromised remote data centers using untrusted network for NG 4
- In the remote data centers: Compromised MEC servers
- N6 (facing the internet): Low latency capable NG 6 interfaces exposed towards the internet and susceptible to DDoS and DoS threats from the internet

Figure 5.4 shows the visibility points and the mitigation controls. Many of these controls are familiar to operators today. 5G presents challenges of distributed deployment, orchestration, and scale-up/scale-out with automation to keep up with threats on a distributed core architecture.

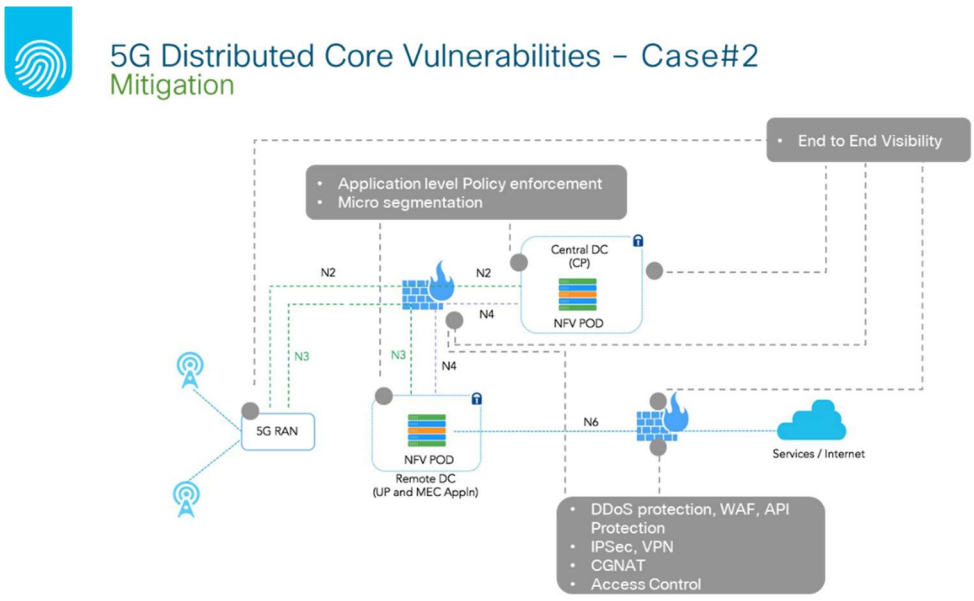

Figure 5.4. Distributed 5G Core Vulnerabilities – Mitigation.

Mitigating threats at the virtualization layer must also be addressed regarding 5G architecture. Networks built today are highly virtualized in key NFs. 5G takes virtualization to a higher level. The operator's backbone network connects a number of widely distributed smaller data centers to a few larger data centers. This infrastructure requires visibility of application dependencies and of traffic patterns feeding that information into the broader analytics function. This is further described in Figure 5.4. An orchestrated NFVi layer exists on top of that infrastructure. 5G catalyzes a move to highly virtualized workloads and in certain cases, movement of certain key parts of the network to the cloud (CUPS model for Control and User Plane Separation). CUPS is not a strictly 5G feature, but is another aspect of the new trust boundaries and threat surface of the 5G network deployments. Slicing must also be accounted for in the distributed core mitigation and also in the increased number of application policy enforcement instances. Virtualized workloads bring a new set of threats that include, but are not limited to:

- Trust and compromised VNFs
- VM hopping and sprawl
- Compromised micro-services
- Container image vulnerabilities

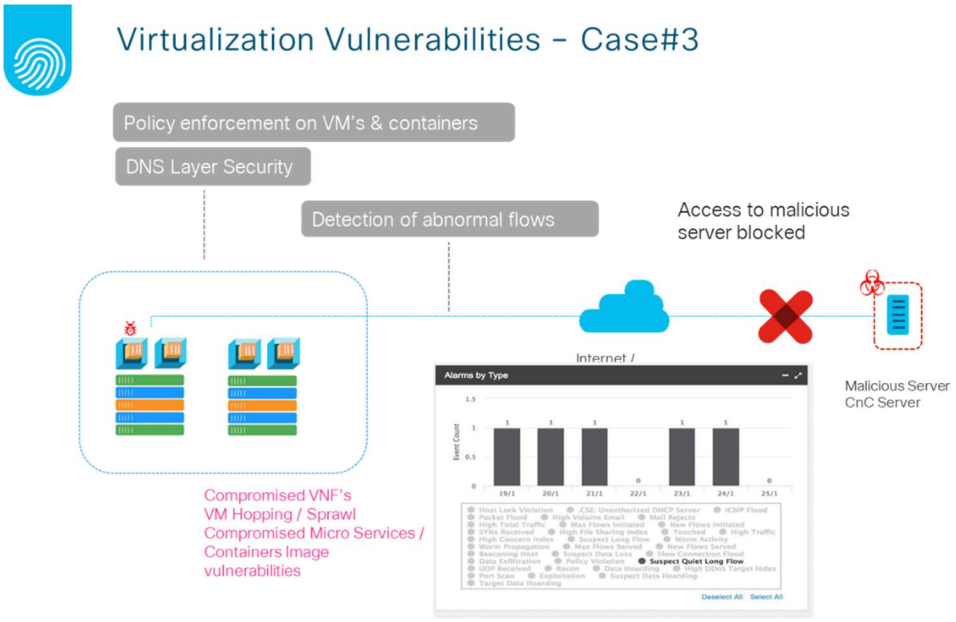

Figure 5.5. Virtualization Vulnerabilities.

Proper visibility, segmentation, DNS level security (for example, known bad talkers, bad domains) and detection of abnormal flows all deliver a foundational layer of security for the virtualization of 5G architecture.

Figure 5.5 shows how proper visibility (flow analysis, ledger of flows and traffic, threat feed integration updated in real time, application dependencies all on a foundation of proper network segmentation) and behavior analysis allows the operator to detect threats impacting the 5G network core.

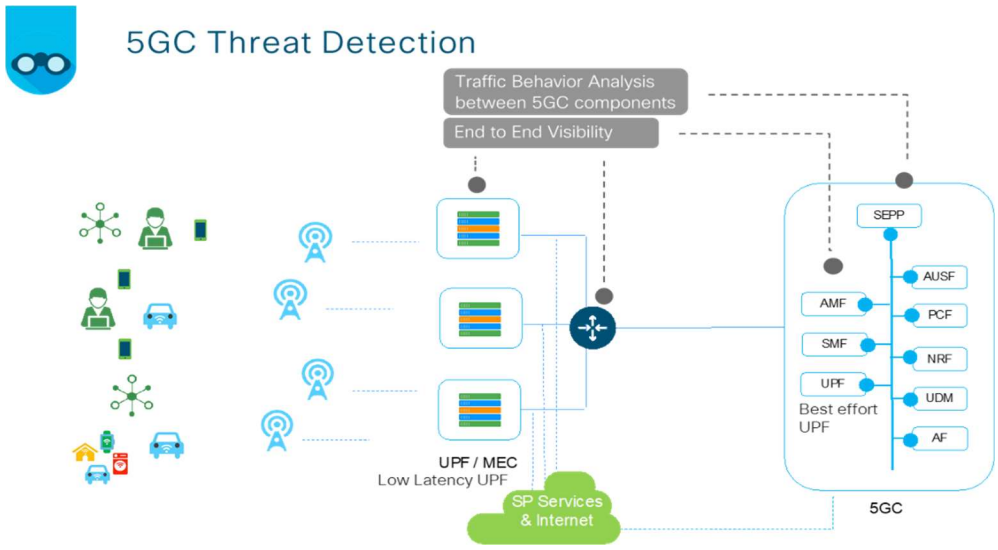

Figure 5.6. 5G Threat Detection.

Mitigation controls, segmentation tools, and visibility tools that provide the foundation for an operator to secure the 5G network and its services are examined throughout this paper. The discussion of mitigation of IoT threats and DDoS threats will now be addressed.

5.2 IOT & DDOS THREAT MITIGATION

There will be several ways to mitigate IoT threats and threat surfaces with 5G technology and these methods are addressed in this section.

5.2.1 IOT DEVICE

Sensitive data in non-secure physical device locations needs to be encrypted and its integrity protected. Devices must cryptographically verify firmware and software packages at system boot or update, as well as maintain the ability to receive remote firmware updates even in case of malware infection. Sufficient storage must be provided for automatic rollback in the event of an update failure. However, malicious rollback to older software/firmware versions that reintroduce old vulnerabilities must be prevented.

The need for security isolation between device-resident applications is critical. One option is to provide hardware-based isolation between applications. This involves a 'root-of-trust' approach to prevent compromised OS, depicted in Figure 5.7. Although this functionality has typically been provided by dedicated hardware, it can also be realized with a Trusted Execution Environments (TEE). The TEE is isolated from the client-side execution environment and referred to as Rich Execution Environment (REE) in common processors. For low-cost devices, the use of TEE is preferred. The TEE specification set is publicly available from Global Platform.

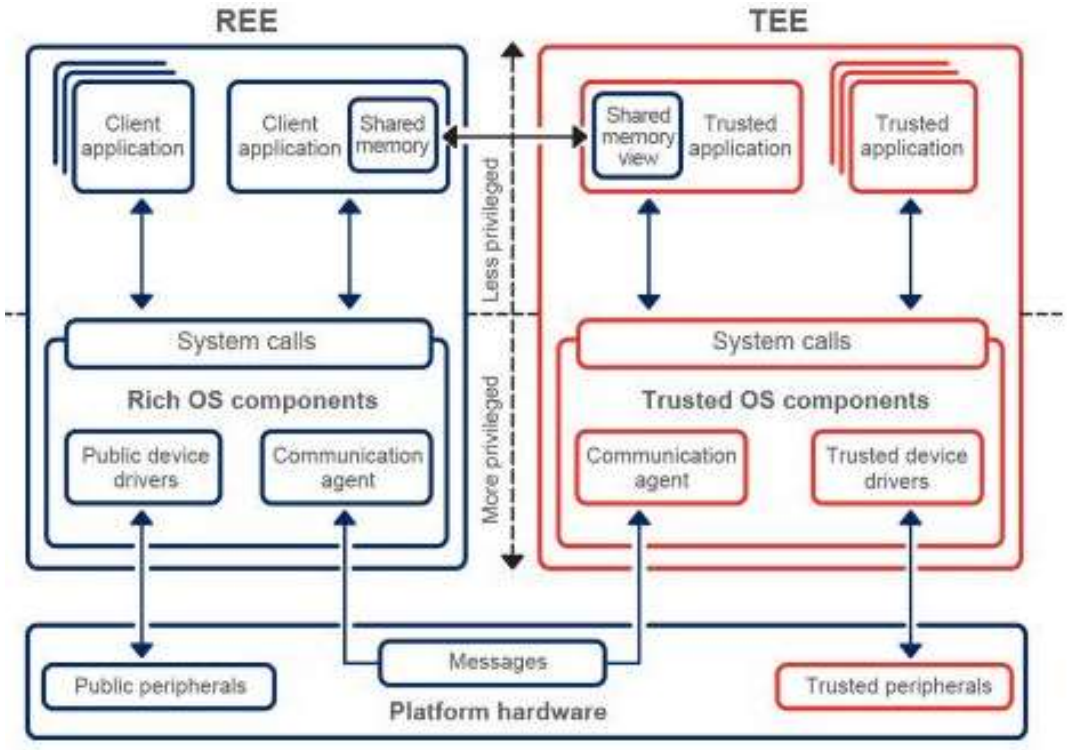

Figure 5.7. Root-of-Trust Approach.

Today's cryptographic algorithms, even asymmetric algorithms, are significantly faster than legacy algorithms and better suited for IoT. Lightweight cryptography may be appropriate for at least some

scenarios. For IoT devices that reside in exposed environments, protection against side-channel attacks is essential to prevent leakage of keying material through timing information, electromagnetic signatures, power consumption, and etcetera.

5.2.2 NETWORK/TRANSPORT

Mobile operators can leverage their unique position in the IoT space as both connectivity and platform providers. Technologies such as Long-Term Evolution Category M-1 (LTE-M) and NarrowBand-Internet of Things (NB-IoT) are superior solutions designed to provide more robust global connectivity compared to unlicensed access. Mobile networks can enhance IoT security by providing device management, secure bootstrapping, and by verifying device location or platform trustworthiness.

Typically, device credentials are pre-provisioned on removable UICCs. An embedded UICC (eUICC) enables remote provisioning and credential management. The risk of security breaches can be reduced by actually generating credentials on the device. A logical next step is to use a TEE that is already integrated in the baseband processor. This combination offers advantages like reduced hardware cost and power consumption, improved speed, as well as the flexibility of secure modification of credentials and network configuration.

IoT covers a wide variety of ecosystems. The flexibility for securely bootstrapping connectivity credentials—from device credentials, and/or application credentials from connectivity credentials—can be very important for certain use cases. For instance, a customer seeking a single service layer agreement with a single connectivity aggregator. The slices that a device can use are also an important aspect of this provisioning and management that could benefit from this flexibility.

5.2.3 NODE/PLATFORM

IoT platforms can and should bear the responsibility of managing the lifecycle of IoT devices from installation to decommission, ideally with minimal need for manual intervention.

During the device installation step, an IoT device will typically automatically bootstrap itself into active service using pre-configured credentials (keys identifiers) stored in a secure hardware module or baseband processor. The corresponding IoT platform will perform initial configuration steps including firmware update, application configuration and provisioning of credentials for application layer services.

During device operation, the platform should enforce security policies such as authorization and access control, as well as any required delta updates in software, credentials, storage, and etcetera. At decommission, it is important that the platform be able to remotely delete all sensitive information stored on the device.

5.2.4 APPLICATION

IoT applications should be placed on secure platforms by using roots of trust in a cloud infrastructure. The exchange of data between IoT applications, or between applications and devices, can be secured via lightweight IETF security protocols, such as an authorization framework based on OAuth (IETF) suitable for constrained environments.

To protect against intermediaries, sole reliance on IPsec and TLS may not be sufficient. These protocols only support trust models that can guarantee fully trusted endpoints. Authorization to access information

should only be allowed on a need-to-know basis. To accomplish this goal, end-to-end security needs to be at the application layer. The preferred solution for protecting message exchanges is the use of information containers at the application level, rather than at lower layers in the protocol stack. These containers are capable of confidentiality, integrity and origin authentication.

5.2.5 SERVICE

To illustrate service level security, the modern connected vehicle scenario mentioned previously is used as an example. Connected vehicles contain a complex system of thousands of sensors, actuators and a code base distributed across multitudes of embedded processors. Here isolation, both logical and physical, is critical. For example, a breach in the entertainment system must not be allowed to impact the steering system. Firmware updates must ensure compatibility between related subsystems. Vehicle-to-vehicle communication has the potential to prevent almost all accidents. Malfunctioning machines accidents may never be completely eliminated, but ensuring secure communication has the potential for realizing a significantly safer transportation system.

There are many other scenarios where IoT can enhance public safety. For example, vehicles could be made aware of pedestrians in advance by integrating sensors and cameras into traffic lights. Emergency response is another area where IoT can make a significant positive impact. Free traffic lanes could automatically be created for emergency vehicles, missing children can become easier to find or track, and natural/man-made disasters could be better monitored and contained. The critical nature of these scenarios implies that service-wide security is essential for preventing misuse or even the suspicion of such misuse. Of course, public safety needs will always need to be balanced against privacy needs (for example, the right to be forgotten). A secure IoT service infrastructure can be tuned to achieve that balance.

5.2.6 SECURITY REQUIREMENTS FOR 5G NETWORK MASSIVE IOT THREATS

Deliberate security requirements for the 5G network are needed to prevent 5G service disruption caused by MIoT botnets used for DDoS RAN attacks, and to ensure 5G service resiliency. The fundamentals of these security requirements are detection and mitigation of DDoS attacks against the 5G RAN, also classified as 5G RAN overload functions. Realization of these security requirements will involve collaboration between the 5G standards community, 5G operators and the 5G RAN vendors. Although each operator's unique 5G network implementation may provide some limited protection against this type of attack, it will not fully suffice. 5G RAN components will need to play a significant role in truly and effectively detecting and mitigating these types of attacks in real time. This is where the 5G standards community and the 5G RAN vendors will play a key role.

5.2.7 DETECTION OF DDOS ATTACKS AGAINST THE 5G RAN

To detect a DDoS attack against an operator's 5G RAN caused by MIoT botnets, detailed aspects of the attack must be examined. The previously described attack scenario states the following: malicious hackers instruct their MIoT botnet army to reboot all the devices in a specific or targeted 5G coverage area at the same time. This will cause excessive malicious attach requests, creating a malicious signaling storm. Using these details, the detection requirements can be formulated.

The 5G RAN components immediately impacted by this type of attack will be the most effective elements to play an instrumental role in the detection process, given the required real-time response. The related 5G

RAN NR or gNodeB components are: The Radio Unit (RU), the Distributed Unit (DU), and the Centralized Unit (CU). Given the functions of these components, the ideal component to leverage for the detection of this type of attack will be the Central Unit Control Plane (CU-CP).

Because the CU-CP is instrumental in managing the Radio Resource Control (RRC) connections, it would be most the efficient location for embedding detection functions. The key software elements of the detection functions that need to be embedded in the CU-CP are: an adjustable threshold for all aspects of RRC connection requests, analytics algorithms to determine if it is a DDoS event (based on threshold), volumetric anomaly, timing, Radio Network Temporary Identifiers, etcetera. The adjustable threshold function and analytics function should also be able to get updates from an external Machine Learning (ML) and Artificial Intelligence (AI) platform by means of open interfaces.

5.2.8 MITIGATION OF DDOS ATTACKS AGAINST THE 5G RAN

The same attack scenario will be considered for the mitigation of a DDoS attack against an operator's 5G RAN. Once the DDoS attack is detected natively by the CU-CP, some type of mitigation action is needed. The CU-CP would also be the most effective 5G RAN component to mitigate this type of attack. This is because the CU-CP is instrumental in managing the RRC connections, making it ideal to block the excessive malicious Attach Requests. The described combined actions of detecting and mitigating this attack will demonstrate inherent closed loop automation.

5.2.9 PROTECTING 5G NETWORKS AGAINST DDOS AND ZERO DAY ATTACKS

5G networks are vulnerable to attacks on both the control and data planes. Threats (regarding the control and data planes) and strategies for mitigation are detailed below.

The first example concerns the control plane. Before the UE has an established connection (for example, to make calls), a series of messages must be exchanged between the evolved NodeB (eNB), next generation NodeB (gNB), and finally the MME. If an attacker is able to take control of several devices and cause them to reconnect (for example, by restarting them), this could cause a signaling storm. In the 5G era, there can be 100x more devices and 1000x more bandwidth per unit area compared to LTE networks.

Another example can be an attacker using legitimate devices on an operator's network to target either the operator itself or a third party to produce a denial of service attack. Such attacks create large amounts of traffic at the level of the data plane.

Although these attacks occur on the data and control planes, in principal, they are not very different. In both cases, abnormal amounts of traffic (of varying kinds) are produced by network devices, and the traffic is characterized by sharing some common, albeit complicated, attributes.

There are many ways to detect these attacks. Supervised models have excellent performance in network intrusion detection when they are given good training data. For example, simple deep neural networks (DNN) perform extremely well to detect attacks on the KDD99 dataset, which is widely used in machine learning research and intrusion detection systems. This leaves the problem of generating good labels, which can be done with an unsupervised pipeline.

A combination of these two approaches is recommended. The first is to calculate statistics from 24 hour sliding windows and feed this as input to an anomaly-detection algorithm. There are many viable

approaches here. Isolation forests[8] work well, as do approaches based on the Mahalanobis distance function[9] and auto-encoders.

This approach will produce many false positives. Recognizing that DoS service attacks produce connections with shared commonalities will reduce false positives. Simple vertical features (counting the number of anomalous connections per gNB, or a given User Agent string or Type Allocation Code) can be used to build basic rules to reduce false positives at this stage of the pipeline. One approach is to identify clusters automatically with a clustering technique such as K-Nearest Neighbors. A more robust approach is to produce a view of the data which can be fed into a Convolutional Neural Network (CNN) for anomaly detection.

5.3 NETWORK SLICING SECURITY THREAT MITIGATION

There are different techniques for achieving security isolation that provide different benefits and drawbacks. Figure 5.8 is a proposed structure for classifying existing techniques to better address the problem of isolation in future systems.[9]

The problem of secure isolation can be framed in two ways depending on the threat model. In one approach, isolation may involve executing untrusted programs within a security perimeter.

For another approach, hardening a system will protect execution of trusted but vulnerable programs that have an increased attack surface. For example, Internet-facing programs (Web Servers, Email Servers and DNS) are trusted, but require protection to limit exploitation of vulnerabilities.

Figure 5.8's hierarchical model has been suggested to better understand security isolation.[10]

[8] *Isolation forest*, Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. Eighth IEEE International Conference on Data Mining, ICDM '08. 2008.
[9] A novel anomaly detection scheme based on principal component classifier In IEEE Foundations and New Directions of Data Mining Workshop, in conjunction with ICDM'03 (2003), pp. 171-179 by M-L Shyu, S-C Chen, K. Sarinnapakorn, L. Chang.
[10] *A Study of Security Isolation Techniques*, Rui Shu, Peipei Wang, Signund A. Gorski III, Benjamin Andow, Adwait Nadkarni, Luke Deshotels, Jason Gionta, William Enck and Xiaohui Gu. ACM Computing Surveys, Vol. 49, No. 3. October 2016.

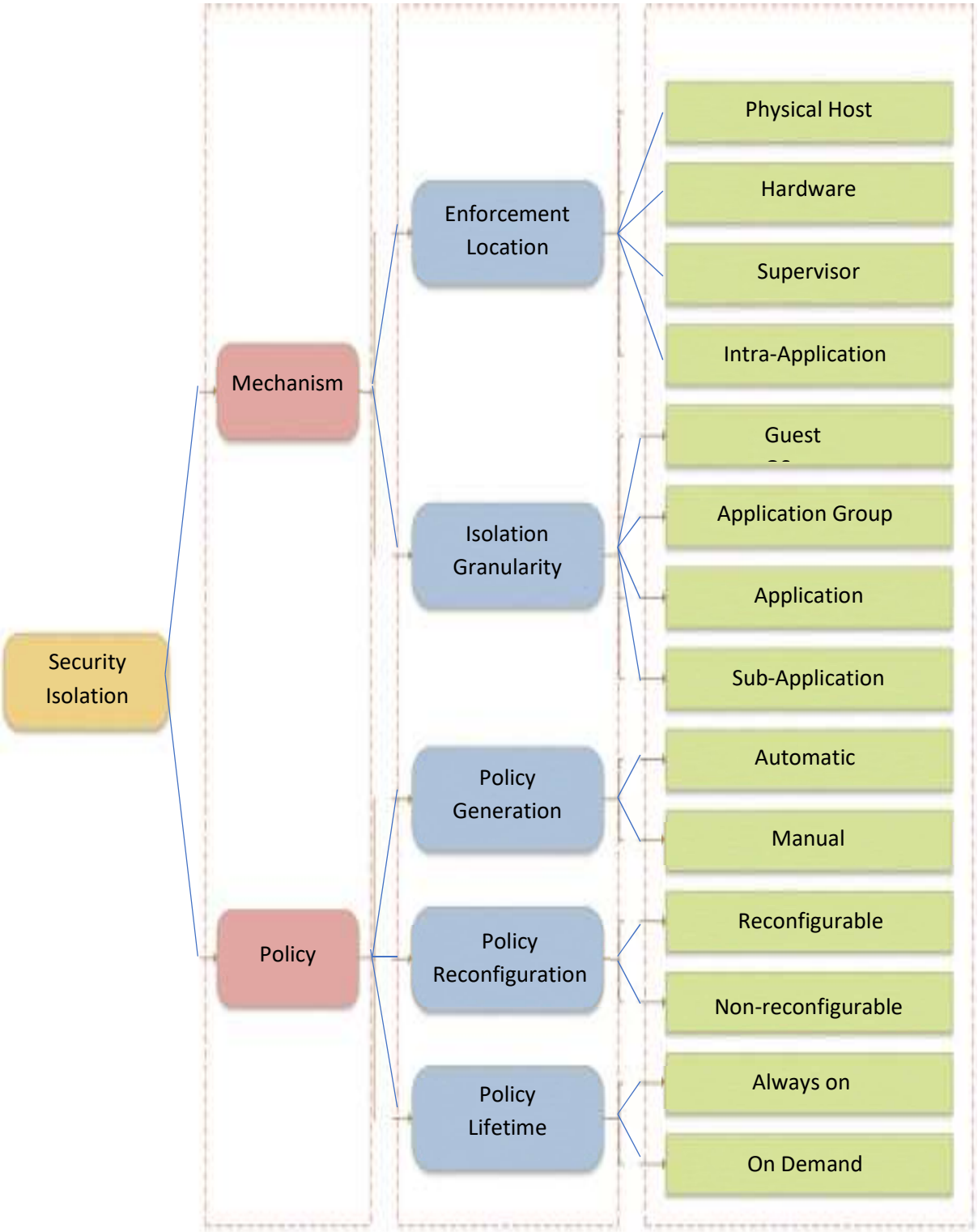

Figure 5.8. 5G Security Isolation.

The classification hierarchy for security isolation can be broadly partitioned into two main design categories: Mechanism and Policy. A set of relevant sub-categories can be considered for each category. For example, the design sub-categories considered for Mechanism are Enforcement Location and Isolation Granularity. Both are broken down into further sub-categories reflecting specific design choices. A similar hierarchical schema applies to the Policy design aspect.

A detailed exploration of this model of isolation (for example, tradeoffs based on performance overhead, code requirements, security assurance levels, and etcetera) is outside the scope of this white paper. The suggested classification is a way to frame the problem to identify the best isolation approaches for real-life network slicing scenarios. It is introduced here as a possible topic for investigation in a future version of this white paper. A few concepts relevant for mitigation of security threats in the context of Network Slicing are noted below.

- One security impact of network slicing architecture is the potential expansion of the attack surface through which malware can be introduced. As mentioned earlier, isolation between network slices is a key requirement. In addition to network slice isolation, multi-layer isolation could also reduce the attack surface and lessen the impact. Examples of multi-layer isolation include: NFVI boundary isolation, isolation of MANO system, security domain isolation, service instance isolation, VNF isolation, etcetera. Various technologies and software/hardware cryptography would need to be adapted to the desired isolation levels. The technologies include various software, hardware, and cryptographic mechanisms. The actual implementations may cover a range of options including managed containers, hypervisor-managed virtual machines, and VPN

- Both Network Slice Managers and host platforms should support mutual authentication. Network Slice Managers should authenticate the hosts before activating a slice instance. Host platforms should also authenticate the Slice Manager before allowing a slice instance to be loaded and run on the physical hardware. In addition, mutual authentication between slice managers should be a requirement where multiple slice managers are involved for instantiating an end-to-end network slice

- While no combinations/options are entirely ruled out, network slicing architecture leans in favor of end-to-end security over hop-by-hop security. This is a consequence of shared infrastructure that is likely owned and operated as multiple independent segments. E2E security implementation is less reliant on segment-level security. However, E2E security also highlights the situation where the endpoint user/device needs to access multiple slices and each provides differing levels of security. If the endpoint can access multiple network slices, it needs to be authenticated (per TS 33.501) to access the 5G system before accessing any slice. In addition, the endpoint should be authorized and/or authenticated for accessing each network slice, especially in case of concurrent use of multiple slices. A common authentication framework (for example, EAP) could be used for implementing slice-specific authentication

- An estimate of resources (CPU, Memory, Storage, and etcetera) needed for individual slices has to be considered for overall network slice architecture design. It is recommended that the design process caps either resources up to a prescribed maximum or ring-fence resources for individual slices to ensure a guaranteed a minimum level. The nature of slice usage could be used to determine the best approach

- If possible, slices that have very different characteristics (for example, levels of sensitivity, levels of vulnerability, and etcetera) should not be co-hosted on the same hardware platform to avoid side-channel attacks

- In terms of mission criticality, highly sensitive slices with similar characteristics may warrant separate hardware as well. These decisions need to be managed on a case by case basis

6. CONCLUSION

5G may be seen as evolutionary in the context of cellular technology generations. Key functions and frameworks specific to previous generations (3G, 4G) continue to work within the overall 5G umbrella. For example, the 5G Radio (NR) can be "plugged" into a 4G core, a backward compatibility feature that did not exist for either 3G or 4G radios, as well as coexist with 4G radios as part of the overall network. In addition, 5G allows for a proliferation of access technologies of all types with data speeds from Gbps to Kbps, licensed and unlicensed, that are based on wide swaths of spectrum bands and include technologies specified by standards bodies other than 3GPP. Viewed from this angle, 5G appears to be a continuous upgrade that incorporates previous generations of cellular/wireless technologies. However, when viewed from a broader perspective, 5G is nothing short of transformational.

One aspect that cannot be overlooked in the "journey" to a secure 5G is that the core tenets of the security architecture are an evolution of best common practices, people, processes and tools that the mobile wireless industry uses to secure our networks today. This paper highlighted a number of new components of the threat surface. Many of them, such as NFV, are not new; they are just now more prevalently deployed in the virtualization of the 5G packet core workloads. The innovation applied to how to secure the networks we operate today in visibility, segmentation and mitigation controls builds on previous success, making the daunting threat surface of 5G a bit more manageable by applying techniques such as automation, orchestration, distributed network build and operation, policy, analytics and much more.

Security is, and always has been, critical to the mobile networks that operators build and manage. The importance and critical nature of security to the mobile wireless industry will remain into the foreseeable future. The connected healthcare IoT service might be powering a pacemaker or insulin delivery unit that someone's life depends upon-- all empowered by the secure 5G networks.

Key aspects of the impact on security for 4G to 5G evolution are summarized below.

The 5G networks are both an evolution and innovative revolution of the 4G mobile networks. Accordingly, 5G security has been designed to build upon the top of, and further enhance, the current 4G strong security controls. The main security enhancements in 5G as defined by 3GPP include the following:

- Secure communications and state of the art encryption and integrity protection mechanisms are utilized in 5G to protect the user plane, control plane and management traffic
- Unified authentication framework for the various 5G access technologies and devices. This would enable seamless mobility across different access technologies and support of concurrent connections
- User privacy protection for the information that can be used by unauthorized parties to identify and track subscribers (for example, protecting permanent identifiers such as SUPI, IMSI, and IMEI)
- Secure Service-Based Architecture and slice isolation that enable different services and applications to implement optimized security mechanisms and prevent attacks from spreading to other slices
- RBS detection and mitigation techniques, utilizing UE-assisted RBS-detection mechanisms and radio-reporting analytics

- In the roaming scenarios, the home and the visited networks are connected through SEPP to address the security vulnerabilities that were found in the legacy roaming networks that use SS7 and Diameter vulnerable protocols. Also, 5G added native support for a secure steering of roaming (SoR). The 5G SoR solution enables the home network operator to steer its customers while roaming to its preferred visited partner networks to enhance roaming customers' experience, reduce roaming charges and prevent roaming fraud

Several features characterize 5G as a revolutionary step in the annals of mobile technology evolution. From the concept of network slicing to support for highly constrained IoT devices, from NFVI to cloudification, from ultra-low latencies to orders of magnitude enhancement of data rates, 5G brings in concepts and features that mark a significant discontinuity with the past. A full discussion of the 5G architecture is outside the scope of this paper. Instead, this paper focused on a review of the security aspects of 5G, some of which are attributable to the uniqueness of 5G architecture. It is worthwhile to note a few characteristics that distinguish 5G security from that of previous generations of cellular technologies.

- In the context of IoT, DDoS attacks coming from 5G RAN originated via botnet-controlled compromised devices were explained in the paper. However, such threats go well beyond IoT. While RAN-based threats are not new, for future full-function 5G devices, capable of data rates that are orders of magnitude higher than what is possible today, the DDOS threat may be significantly magnified, requiring any mitigation approaches to scale accordingly. The criticality of the speed with which such attacks are detected is likely to be enhanced. Automated defenses, to ensure the quickest possible response in the event of an attack, may become indispensable

- 5G is unique in its focus on services that go beyond just monetary/economic values. For the first time in cellular history, 5G incorporates, as part of its core support areas, services that directly pertain to users' wellbeing and livelihood, including such services as automotive and health. Of course, the cost of a security breach for such services also goes well beyond monetary losses. Consequently, the scope of security compliance may also need to go beyond conventional IT security metrics into the realm of stringent government regulations. While the scope of this category of security requirements remains largely undefined at this time, we are certain that, with increasing adoption of 5G for these sectors, 5G will need to contend with unique security requirements in future. To complicate matters there may be multiple authorities (nations, states, other authoritative bodies) imposing a diverse set of security/privacy requirements across the globe. A global mobility standard such as 5G will need to account for a diverse and complex regulatory environment

- 5G leads to a future where software rules. Hardware components do exist, but primarily as "white box" commodities. The software-centric 5G picture has two important consequences. First, a convergence of all communication modes, mobile/fixed/wireless/wireline, becomes a reality with 5G. The security solutions cannot be limited to addressing specific communication modes serving only their niche ecosystems as they do today. Security needs to be both comprehensive and embedded into the design, not appended as a separate mechanism. Second, the move to virtualization will accelerate with time. Today's NFV implementations largely mimic a software version of the hardware being virtualized. Such implementations frequently replicate existing security mechanisms. For a fully automated and cloud-based NFV infrastructure, existing security solutions are likely to fall short. The market will continue to include service providers with only limited/partial 5G implementations for some time. However, the sooner security solutions can address a fully virtualized 5G end state that includes orchestration, dynamic network management

and cloud-based infrastructure, the better prepared the overall industry will be against threats that may yet to be fully envisioned

- Key IoT security threats such as DDoS are addressed in this paper. Privacy is intimately tied with security, and for many, is of equal or greater concern for IoT. A plethora of information strewn around both clouds and multitudes of IoT devices heightens the privacy risk. While individual fragments of information may not reveal much, the collective magnitude of data could be very revealing through use of big data analytics. Seemingly harmless data related to electricity consumption or room temperature settings, for example, may reveal too much about an individual. With billions of sensors everywhere, IoT drastically increases the amount of potentially sensitive information generated. Compounding the problem, people may be unaware of the sensors around them or how combined data from various sources can be misused. Even if IoT traffic is encrypted, significant and meaningful patterns containing confidential information could be exposed through analysis. Finally, many IoT devices remain in exposed unguarded locations for long periods further increasing the risk. Beyond individual exposure, industrial espionage is another significant concern related IoT privacy

- The depth and breadth of the 5G ecosystem guarantees a level of complexity for 5G that goes well beyond previous generations of cellular technologies. For example, an important pillar of 5G is dynamic network slicing. The intent is to provide customers with not just guaranteed access to the network, but also network resources that are customized to satisfy customer needs dynamically. In the context of such dynamic and tailored scenarios, providing security for individual slices for individual customers, while also assuring security for all other customers, promises to be one of the biggest security challenges for 5G. The complexity of multiple simultaneous network slices, each operating under a different set of service and security requirements, may need a completely new paradigm for how the problem of network security is approached. Adding to 5G complexity, will be multiple radio access technologies, ultra-low latency services and IoT devices

- Network slicing is a new aspect of 5G used to segment the network for service delivery to meet the stringent demands of 5G and even more 5G services with IoT services on top that require ultra-low latency. The 5G network slice provides an additional threat surface that is addressed in this paper as are the mitigations for those threats. That said, there is a lot of operational enhancement to how we find threats faster, fix them faster and operationalize those solutions. This is the promise of security in 5G. Innovation, service agility and solutions crafted and enabled for 5G demand this new level of operational agility in security

5G is very early in deployment at large scale. Today, the 3GPP standards for 5G along with the use case based early deployments have collectively led to some early best common practices. Some of those best practices include, but are certainly not limited to the following as it relates to network slicing:

- Slice tenants have different needs for certain features or customizations. It is a good idea to group tenants according to their requirements; tenants with similar needs should be put on the same deployment

- Identify ‘most asked for features’ and build it into the core platform to avoid customizations at the tenant level as much as possible

- Close monitoring of each tenant's activities for exercising timely control over any particular tenant's actions that adversely impacts other tenants

- Consideration of the use of 'role-based' fine-grained access controls to limit a tenant's access across the entire stack. Determine who can access individual data items and what actions that can be performed on them

Finally, for the level of complexity introduced by 5G, canned (i.e. preconfigured) security mechanisms may need to be supplemented with dynamic security measures where the defense mechanisms are instantiated and deployed by AI-based systems as responses to a new generation of multi-pronged zero-day attacks. Early and integrated threat detection is key. Detection needs to go beyond signature-based tools to spot the attacks designed to evade basic filters. Behavior-based checks on endpoints are important. Combinations of packet capture, big data and ML can be used to identify threats not spotted by basic filters. When detection is 'embedded' into switches and routers network, nodes themselves becomes 5G security sensors, enhancing the effectiveness of overall defenses. These defenses are made more effective by properly segmenting the network to ensure that the operator can contain a threat if the network is compromised. Network slicing is one of the enablers tor segmentation and will continue to evolve both as a catalyst to accelerate development of use cases and proper partitioning of the network to make sure those use cases can properly be delivered. Integrated AI-based defense mechanisms are likely to remain in the realm of research for few more years to come.

A. APPENDIX

Acronym	Description
2G, 3G, 4G & 5G	2nd, 3rd, 4th & 5th Generation mobile architecture
3GPP	The 3rd Generation Partnership Project (3GPP) unites seven telecommunications standard development organizations and provides their members with a stable environment to produce the Reports and Specifications that define 3GPP technologies.
AI	Artificial Intelligence
AKA	Authentication and Key Agreement
AMF	Access and Mobility Management Function
AUSF	Authentication Server Function
C&C	Control and Command
CNN	Convolutional Neural Network
CU	Centralized Unit
CU-CP	Central Unit – Control Plane
CUPS	Control and User Plane Separation
DDoS	Distributed Denial of Service
DNN	Deep Neural Network
DU	Distributed Unit
E2E	End-to-end
EAP	Extensible Authentication Protocol
eMBMS	Evolved Multimedia Broadcast Multicast Services, also known as LTE Broadcast
eNB	Evolved NodeB
eUICC	Embedded UICC
FQDN	Fully Qualified Domain Name
gNB	Next Generation NodeB
GUTI	Globally Unique Temporary ID
HSS	Home Subscriber Server
HTTP	HyperText Transfer Protocol
IETF	Internet Engineering Task Force
IMSI	International Mobile Subscriber Identity
IPSec	Internet Protocol Security
IPX	Internetwork Packet Exchange
LTE-M	Long Term Evolution CategoryM1, or LTE for Machine-Type Communication
MCC	Mobile Country Code
ME	Mobile Equipment
MEC	Mobile Edge Computing
MiTM	Man-in-the Middle
MIMO	Multiple-Input Multiple Output
MIoT	Massive Internet of Things
ML	Machine Learning

Acronym	Description
MME	Mobility Management Entity
MNC	Mobile Network Code
MPS	Multimedia Priority Service
NAI	Network Access Identifier
NAS	Non-Access Stratum
NF	Network Function
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NR	New Radio
OAM	Operations, Administration, and Management
OS	Operating System
PCE	Path Computation Element
PEI	Permanent Equipment Identifier
RAN	Radio Access Network
RAT	Radio Access Technology
RBS	Rogue Base Station
REE	Rich Execution Environment
RRC	Radio Resource Control
RU	Radio Unit
S-TMSI	Serving Temporary Mobile Subscriber Identity
SA3	SA Working Group 3 is responsible for security and privacy in 3GPP systems
SEAF	Security Anchor Function
SEPP	Security Edge Protection Proxy
SDN	Software Defined Network
SDIF	Subscription Identifier De-Concealing Function
SLA	Service Level Agreement
SMF	Session Management Function
SN	Serving Network/Serving Node
SOR	Steering of Roaming
SUCI	Subscription Concealed Identifier
SUPI	Subscription Permanent Identifier
TEE	Trusted Execution Environment
UDM	Unified Data Management
UDR	User Data Repository
UE	User Equipment
UICC	Universal Integrated Circuit Card, a type of smart card technology
URI	Uniform Resource Identifier
URLLC	Ultra-Reliable Low-Latency Communications
USIM	Universal Subscriber Identity Module
V2I	Vehicle-to-Infrastructure

Acronym	Description
V2X	Vehicle-to-Everything
VPLMN	Visited Public Land Mobile Network
VNF	Virtual Network Function
VR	Virtual Reality
WG	Working Group

ACKNOWLEDGEMENTS

The mission of 5G Americas is to advocate for and foster the advancement of 5G and the transformation of LTE networks throughout the Americas region. 5G Americas is invested in developing a connected wireless community for the many economic and social benefits this will bring to all those living in the region.

5G Americas' Board of Governors members include AT&T, Cable & Wireless, Cisco, Ciena, CommScope, Ericsson, Intel, Kathrein, Mavenir, Nokia, Qualcomm Incorporated, Samsung, Shaw Communications Inc., Sprint, T-Mobile USA, Inc., Telefónica and WOM.

5G Americas would like to recognize the significant project leadership and important contributions of project leaders Sankar Ray from AT&T and Mike Geller from Cisco and notably representatives from member companies on 5G Americas' Board of Governors who participated in the development of this white paper.