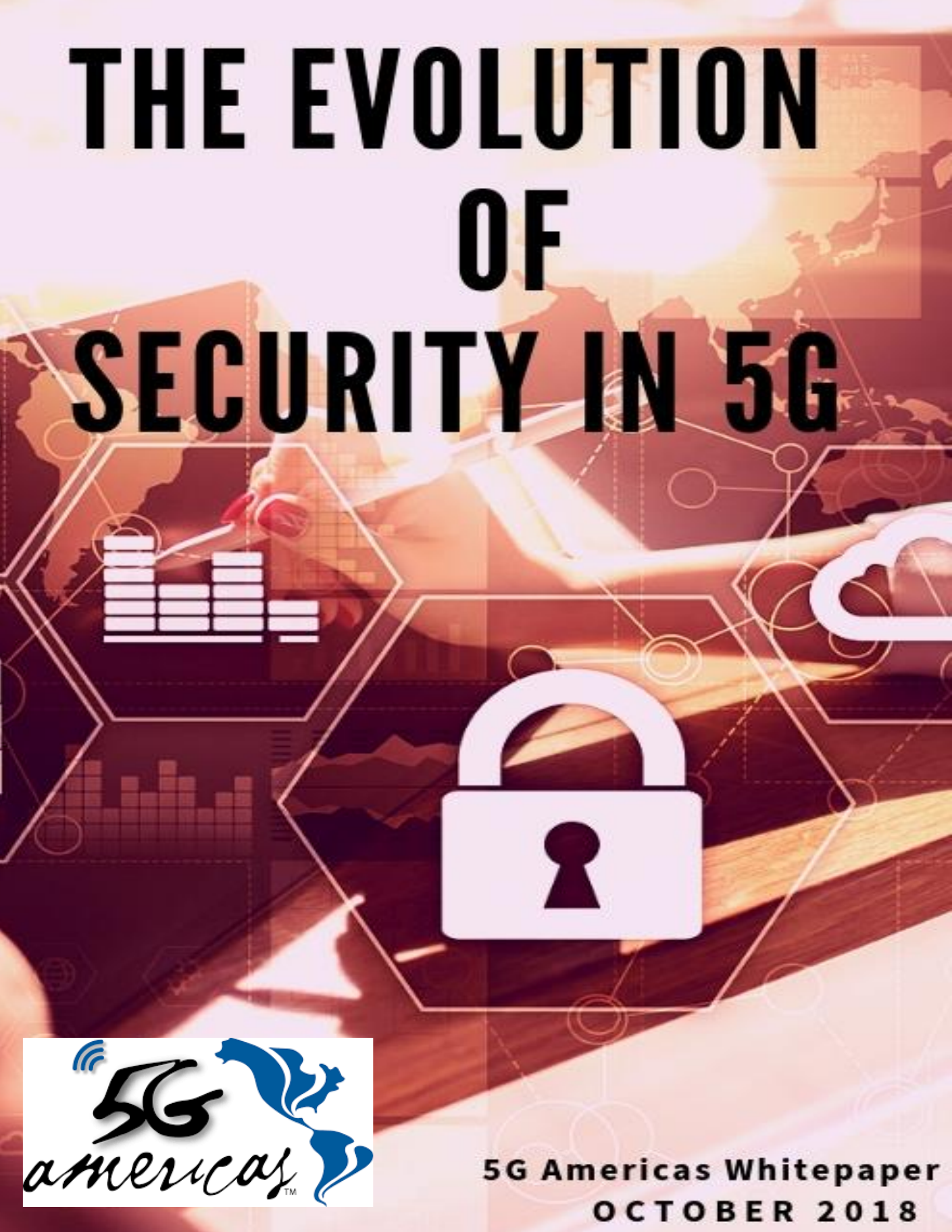
THE EVOLUTION OF SECURITY IN 5G



5G americas™

5G Americas Whitepaper

OCTOBER 2018

TABLE OF CONTENTS

1. INTRODUCTION

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

5G is not just about faster, bigger or better. It's 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.

For example, 5G will enable Massive Internet of Things (MIoT) applications such as the traffic sensors and Vehicle-to-Infrastructure (V2I) services that are the foundation for smart cities. It's critical that hackers can't access that data, hijack IoT devices or disrupt the services with Distributed Denial of Service (DDoS) attacks.

Fortunately, security has been a top architectural 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, while 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, including its security mechanisms.

The mobile wireless industries longstanding emphasis on security has been a strong market differentiator against many other wireless technologies which have network architectures that are inherently more vulnerable. Even mobile's use of licensed spectrum provides a powerful additional layer of protection against eavesdropping on data, voice and video traffic.

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're 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.

For example, 5G is the first mobile architecture designed to support multiple, specific use cases, each with their own unique cybersecurity requirements. In the enterprise IT world, network segmentation is a common, proven way to mitigate security risks. 5G introduces the concept of network slicing, which provides mobile operators with segmentation capabilities that weren't possible with previous generations.

NEW CYBERSECURITY CONSIDERATIONS AND RESPONSES

In addition to the new opportunities and capabilities, 5G creates new cybersecurity considerations. Its use of the cloud and edge computing, and convergence of mobile and traditional IT networks, create 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 visibility example is the use of application-level probes that are synthetically generated and 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

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

on critical service classes for DDoS. Once all of the telemetry is gathered, a security controller and workflow will analyze it and determine, based on policy, suggested mitigation and controls to be applied.

Additionally, the mobile industry itself provides an additional layer of security. Operators, vendors, standards bodies and associations form an iterative loop of constant learning about emerging threats and response options. This illustrates the control aspect, which is the actions taken to mitigate an attack.

Some controls are proactive, while others are applied after an attack takes place. There are also two types of attacks:

- Zero-day attacks are threats that don't already have either a fingerprint or previous history (signature). Typically, deviations in known good behavior of the carrier cloud, and applications that request service and state from it, are identified by the security controller. Some action is then taken to mitigate the attack or to get additional visibility, an action sometimes taken 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 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. Innovation is the way that the industry applies this information, in a closed-loop iterative process to mitigate threats. Thus, 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 today's and tomorrow's attacks. The three elements of the closed-loop iterative process are policy, analytics and the application delivery cloud, which is the whole transaction from the application through to the servicing networks.

Operators can now apply innovative methods to 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 what to do about 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. Even so, the industry must continue to evolve, grow and get smarter to keep networks safe and resilient.

OVERVIEW OF 5G USE CASES

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

However, as *Figure 1* illustrates, 5G introduces new network architectures and use cases, 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 uses case. 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 services. The difficulty of securing such a wide variety of access and service demands via a single integrated 5G network is readily understandable.

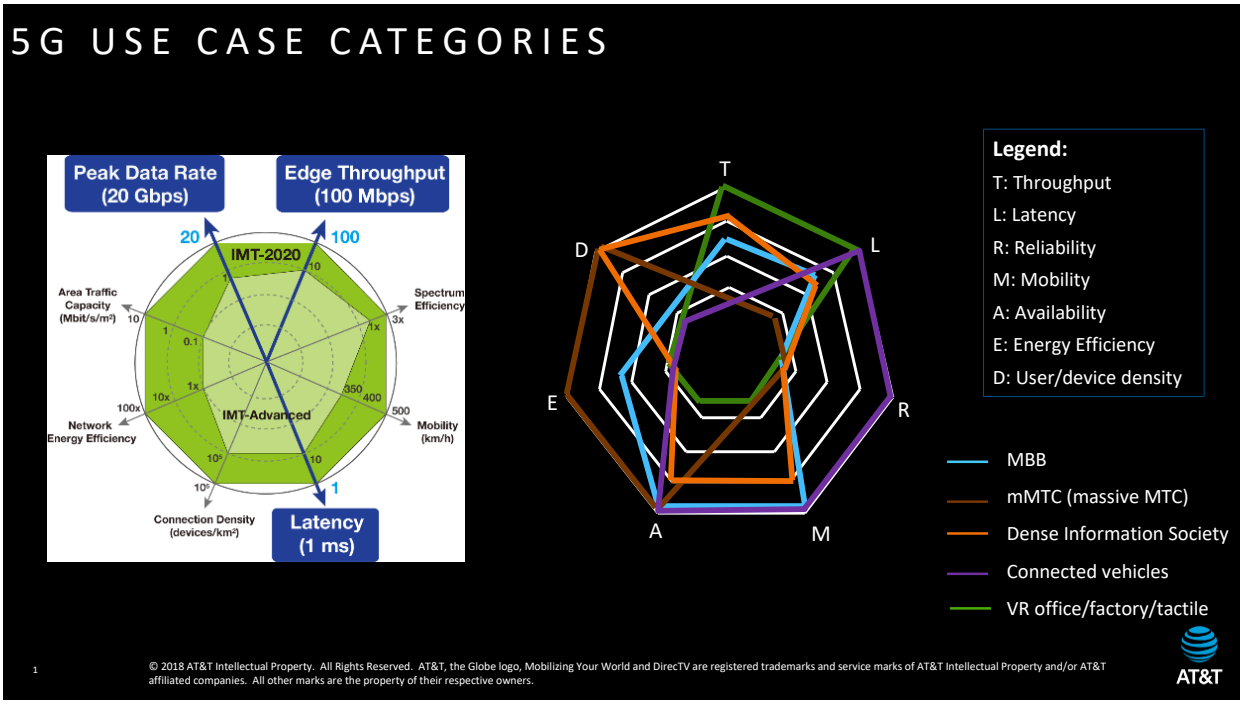


Figure 1. 5G Use Case Categories.

Meanwhile, hackers are continually developing new attack methods, so the mobile industry also maintains the 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.

SECURITY FUNCTIONS FOR 5G-DDOS

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. The IoT market is an enormous business opportunity for mobile operators and their business partners, but its devices and use cases also increases the potential cyber threats.

For example, many of the “things” that make up the IoT landscape have zero-day vulnerabilities, which are security holes in software discovered by the hackers, of which the vendors are ‘as yet’ unaware. The 5G evolution means billions of these things, collectively referred to as MIoT, will be using the 5G Radio Access Network (RAN). Thus, MIoT could increase the risk of RAN resource overload by way of 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, this paper uses the aforementioned MIoT DDoS scenario:

- Hackers identify zero-day vulnerabilities and use them to create a botnet army by infecting many (millions/billions) 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 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, which 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.

2. OVERVIEW OF 5G SECURITY ARCHITECTURE IN 3GPP

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, as well as work on codecs, security and quality of service. Thus, 3GPP provides complete system specifications, which also includes 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: the security features and the security mechanisms for the 5G system and the 5G core; and the security procedures performed within the 5G system, including the 5G core and the 5G New Radio (NR). Following are the main features defined in 33.501.

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 the device is actually in the serving network when the home network receives a request from a visited network.

This was added to address the vulnerabilities found in 3G and 4G networks where networks could be spoofed and send false signaling messages to the home network in an effort to request the International Mobile Subscriber Identity (IMSI) and location of a device. This information could then be used to intercept voice calls and text messages.

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.

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 even serving networks without having to run the full authentication method (for example, AKA authentication). This reduces the signaling load on the home network HSS during various mobility services. The SEAF and the AMF could be separated or co-located. In 3GPP Release 15, the SEAF functionality is co-located with the AMF.

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) is 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 being able to retrieve the subscriber identity 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.

3GPP 5G SECURITY ARCHITECTURE

3GPP defines the overall 5G security architecture, illustrated in *Figure 2.*

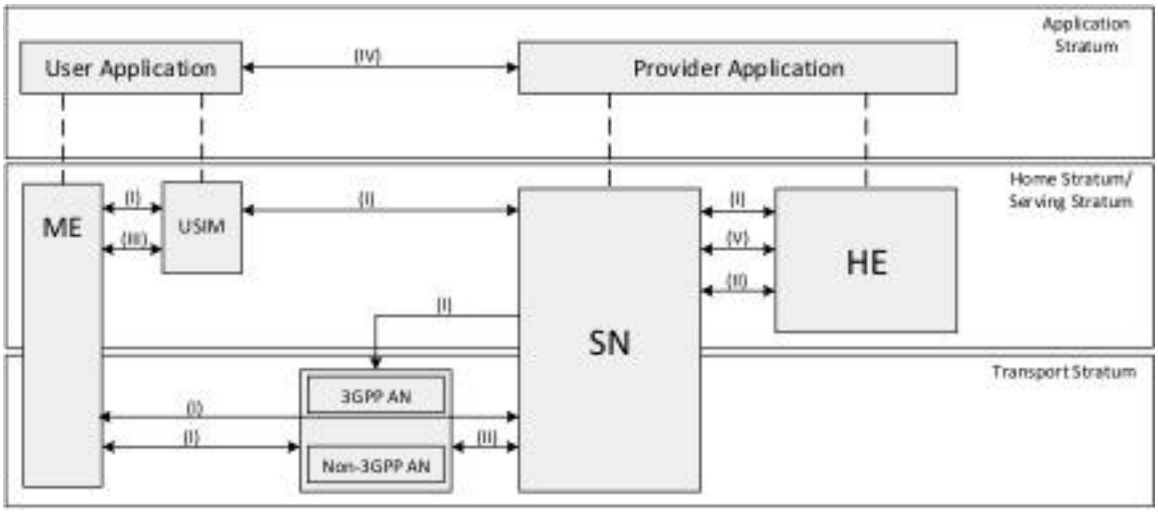


Figure 2. Overview of 5G Security Architecture.

This includes many network architectural elements and concepts such as:

- Network access security (I), which is the set of security features that enables a UE to authenticate and access services via the network securely, including the 3GPP access and non-3GPP access, and in particular 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, user plane data
- User domain security (III), which is the set of security features that secures the user access to mobile equipment
- 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 about the SBA security. 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

Security Edge Protection Proxy (SEPP)

To protect messages that are sent over the N32 interface, the 5G system architecture introduces Security Edge Protection Proxy (SEPP) as the entity sitting at the perimeter of the Public Land Mobile Network (PLMN) network that:

- Receives all service layer messages from the Network Function and protects them before sending them out of the network on the N32 interface
- Receives all messages on the N32 interface and forwards them to the appropriate network function after verifying security, where present

The SEPP implements application layer security for all the layer information exchanged between two NFs across two different PLMNs. *Figure 3* illustrates the SEPP's role.

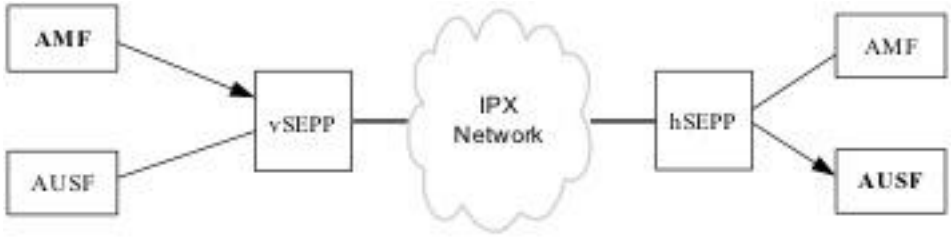


Figure 3. The Role of the SEPP in the Security Architecture.

ROLE OF THE SEPP IN THE SECURITY ARCHITECTURE

The application layer traffic comprises all the IEs in the 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 and still others may require e2e integrity protection but modifiable by intermediate IPX provider while in-transit.

To enable the trusted intermediary IPX nodes to see and possibly modify specific IEs in the HTTP message, while completely 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 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:

- The SEPP shall act as a non-transparent proxy node
- The SEPP shall protect application layer control plane messages between two NFs belonging to different PLMNs that use the N32 interface to communicate with each other
- The SEPP shall perform mutual authentication and negotiation of cipher suites with the SEPP in the roaming network
- The SEPP shall handle key management aspects that involve setting up the required cryptographic keys needed for securing messages on the N32 interface between two SEPPs
- The SEPP shall perform topology hiding by limiting the internal topology information visible to external parties
- As a reverse proxy, the SEPP shall provide a single point of access and control to internal NFs
- The receiving SEPP shall be able to verify whether the sending SEPP is authorized to use the PLMN ID in the received N32 message
- The SEPP shall be able to clearly differentiate between certificates used for authentication of peer SEPPs and certificates used for authentication of intermediates performing message modifications
- The SEPP shall discard malformed N32 signaling messages
- The SEPP shall implement rate-limiting functionalities to defend itself and subsequent NFs against excessive CP signaling. This includes SEPP-to-SEPP signaling messages
- The SEPP shall implement anti-spoofing mechanisms that enable cross-layer validation of source and destination address and identifiers (e.g. FQDNs or PLMN IDs)

REQUIREMENTS FOR E2E CORE NETWORK INTERCONNECTION SECURITY

A solution for e2e core network interconnection security shall satisfy the following requirements:

- The solution shall 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
- The solution shall 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
- The destination network shall be able to determine the authenticity of the source network that sent the specific message elements protected according to the preceding bullet. 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
- The solution should have minimal impact and additions to 3GPP-defined network elements
- The solution should be using standard security protocols
- The solution shall cover interfaces used for roaming purposes
- The solution should account for considerations on performance and overhead
- The solution shall cover prevention of replay attacks
- The solution shall cover algorithm negotiation and prevention of bidding down attacks
- The solution should account for operational aspects of key management

AUTHENTICATION FRAMEWORK

The purpose of the primary authentication and key agreement procedures is to enable mutual authentication between the UE and the network and 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 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 that an authentication run over a 3GPP access network can also provide keys to establish security between the User Equipment (UE) and a N3IWF used in untrusted non-3GPP access.

The UE and the serving network shall support EAP-AKA and 5G AKA authentication methods. The home network operator selects the authentication method to be used. The USIM shall reside on a 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] 3GPP TS 23.501: "*System Architecture for the 5G System*"

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 irrespective 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.

MITIGATION OF BIDDING DOWN ATTACKS

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

SERVICE REQUIREMENTS

A UE shall support a man-machine interface setting for the user to disable use of one or more of the ME's 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 GSM/EDGE, WCDMA, LTE and 5G 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 GSM/EDGE, WCDMA, LTE and 5G NR.

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, MPS, mission-critical services), the UE shall support a mechanism to automatically override user- and network-disallowed 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 such events happened. 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.

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.

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 UDM/UDR. The SUPI is used only inside 3GPP system, and its privacy is specified in TS 33.501 [xx].

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

- IMSI as defined in TS 23.003 [xx]
- Network Access Identifier (NAI) using the NAI RFC 4282 [xx] based user identification as defined in TS 23.003 [xx]. 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 contains the address of the home network (for example, the MCC and MNC in the case of an IMSI-based SUPI).

For interworking with the 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.

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 Mobile Country Code (MCC) or Mobile Network Code (MNC).

The UE shall include a SUCI only to the following 5G 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-GUTI, the UE shall include a SUCI to the Registration Request message, or
- If the UE includes a 5G Globally Unique Temporary Identifier (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, or
- If the home network has configured "null-scheme" to be used, or
- If the home network has not provisioned the public key needed to generate a SUCI

SUBSCRIPTION IDENTIFICATION SECURITY

The subscriber identification mechanism in represented in *Figure 4*. This may be invoked by the serving network when the UE cannot be identified by means of a temporary identity (5G-GUTI). In particular, 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.

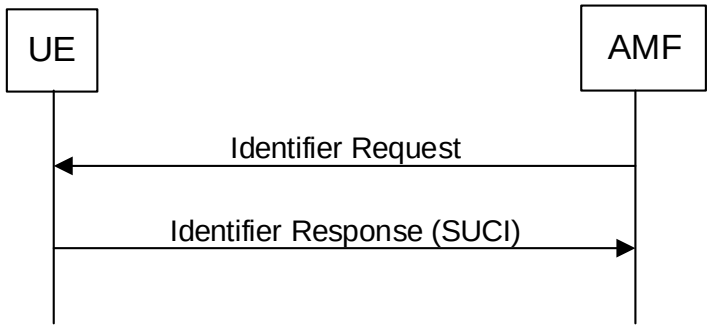


Figure 4. Subscriber Identification Mechanism.[3]

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

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.

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

[3] 3GPP TS 33.501.

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-S-TMSI is the shortened form of the GUTI to enable more efficient radio signaling procedures (for example, during Paging and Service Request).

PROCEDURE FOR USING SUBSCRIPTION TEMPORARY IDENTIFIER

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

- 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 need not be a part of the service request procedure because that would delay the service request procedure.

SUBSCRIBER PRIVACY

Subscriber privacy is an important element to the security aspects of the mobile network architecture. This subscriber privacy is described in the process below:

- 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 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.

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 VPLMN networks to enhance roaming customers' experience and reduce roaming charges.

UE-ASSISTED NETWORK-BASED DETECTION OF FALSE BASE STATION

The UE in 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.

NETWORK REDUNDANCY IN 5G CORE AND NETWORK SLICING

One of the key new aspects of the 5G architecture is segmentation through a concept called network slicing. The concept of segmentation of a carrier network, and application of policy with that segmentation as a foundation, is not new in mobile networks. However, in 5G, it's taken to the next level. New trust boundaries are created both in the packet core and in places where the packet core touches 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 being 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, or
- 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 Non-3GPP Interworking Function (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 5 illustrates some of the key architectural elements of 4G and 5G. A 4G network might migrate to 5G using a model called Nmjon-Stand Alone or NSA. This allows some of the 4G control capabilities to be deployed with 5G user plane. Sometimes, there is a clean cut to 5G. This is commonly referred to as 5G Stand Alone.

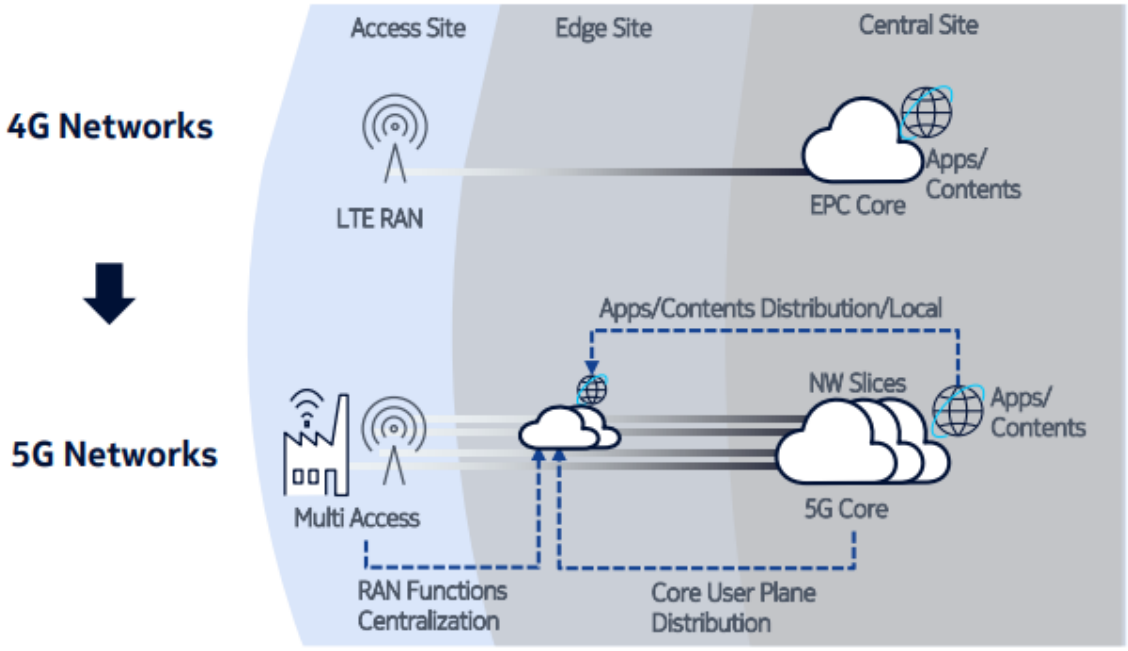


Figure 5. 4G --> 5G.[4]

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 breaking everything down into detailed functions and sub-functions. For example, the 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). Session establishment and session management, also formerly part of the MME, are new services offered by a new network function called the Session Management Function (SMF). Furthermore, packet routing and forwarding functions, 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.

[4] *Source – Nokia.*

Figure 6 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.

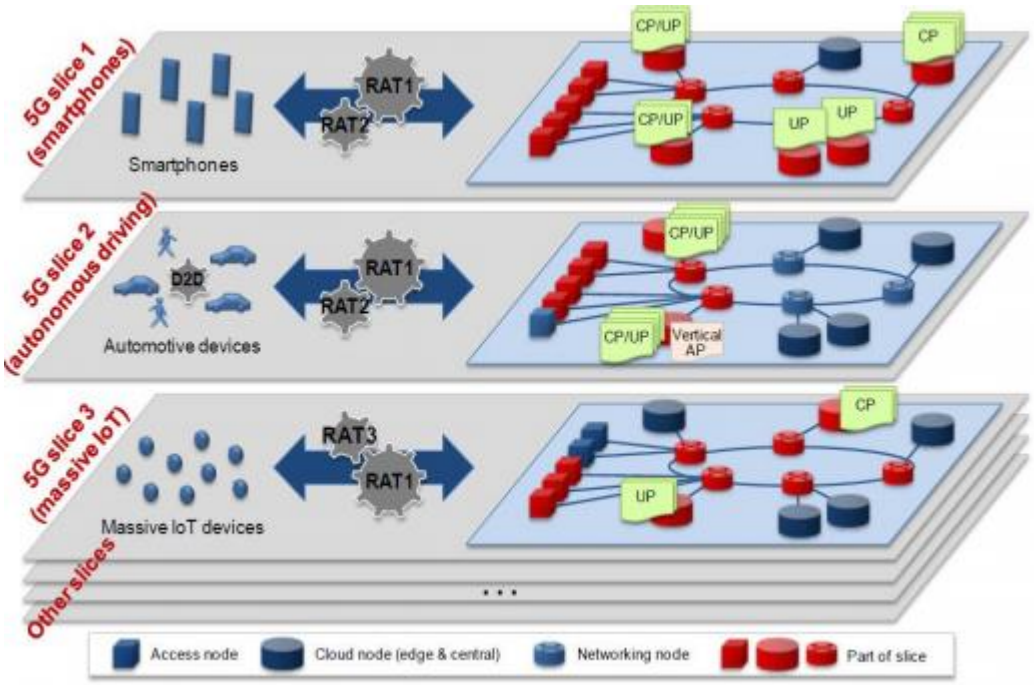


Figure 6. Network Slicing in 5G.[5]

5G leverages a service-based architecture; while they are being set up, NFs 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, where 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.

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's required as NFs can be established and removed on a per-need basis and used by multiple different slices at the same time. Also, network OAM can be made simpler and more flexible as service providers can utilize automated tools to provide the network services with the predefined redundancy, capacity and other capabilities, as well as multiply NFs to the same or multiple locations as needed. Automation can also optimize the unnecessary need for extra hardware and perhaps use it for other purposes, such as analytics or data mining, while the regular network load is low. Of course, some of these tools and capabilities have been available in the network prior to 5G.

[5] Source-Nokia

3. 5G THREAT SURFACE

The 5G threat surface, as described so far in this paper, 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, starting with IoT as it pertains to 5G.

IOT THREAT SURFACE WITH 5G

A 2017 study[6] to investigate the impact of IoT security on IT and line-of-business (LoB) leaders revealed that IT and LoB leaders are anxious about IoT security because attacks can significantly affect critical business operations. One troubling fact revealed was that when it comes to IoT, the majority of organizations cannot provide a complete accounting of all their network-connected devices even as 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. At the same time 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 used to launch coordinated DDoS assaults on an array of targets, including 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 skills. The basic tools are easily available and accessible to all on the internet. The main focus of the Mirai event was that it highlighted key IoT security issues.

The four broad principles that are 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, 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 link. Significant attention is often paid towards securing a mobile phone while ignoring 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. Also, the consequences of hacked connected car, for example, can be substantially more serious compared to that of a connected electric meter or refrigerator.

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 *Figure 7* illustrates. The devices and network/transport may be the areas of primary focus today but from a revenue standpoint, the

[6] *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/

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 IoT security landscape.[7]

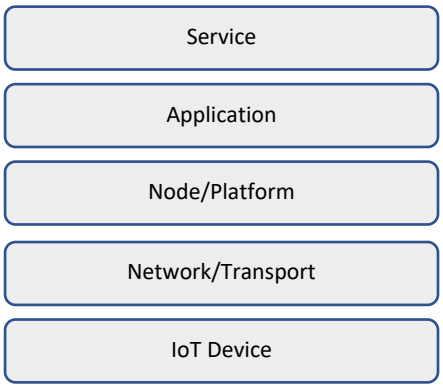


Figure 7. IoT Security Levels.

IoT Device - Many IoT devices will likely reside in exposed and vulnerable environments. Device resident sensitive data can be tampered with. Malicious updates of device firmware and OS pose a significant problem.

Network/Transport - Network connectivity enables secure interaction of device/apps with serving network nodes. To secure this interaction, we need secure identification/authentication (credentials) and data transport. 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 and users and 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 always remain as a focus area. If network connectivity is lost, either because of malfunction or jamming, there needs to be backup mechanisms that on which the service can fall back. There are many other sensor-based services, of various degrees of

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

criticality, that could be enabled by IoT. The path to securing various IoT services will need to consider their uniqueness, as well as criticality of the service itself.

5G THREAT SURFACE FOR MASSIVE IOT

MIoT spans a wide variety of new and exciting opportunities, such as autonomous vehicle communications, smart grids, highway/traffic sensors, drone communications, medical sensors and AR/VR. The MIoT market opportunity, and its 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 8* illustrates this scenario.

.

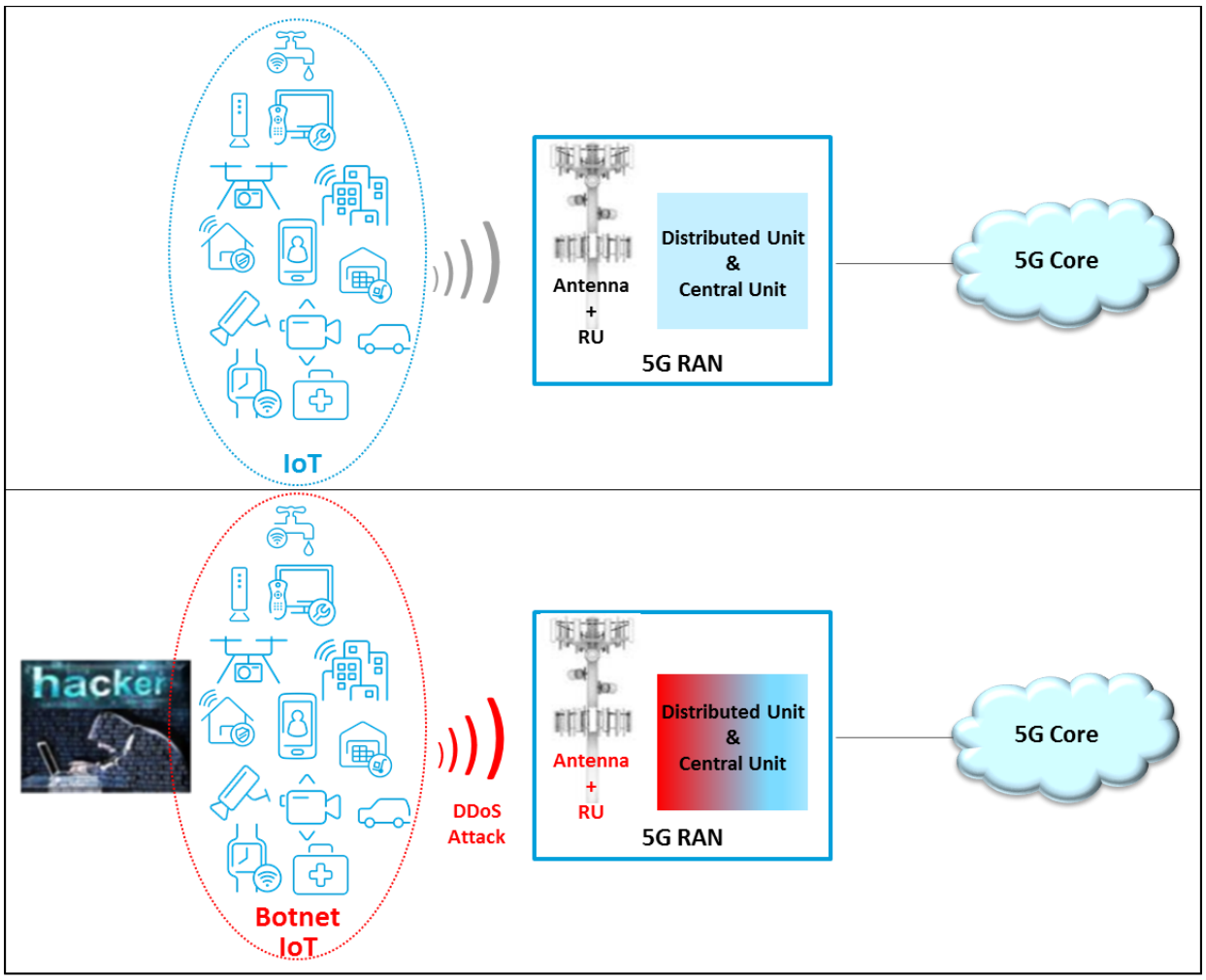


Figure 8. The Network vs. The Hacker.

Figure 9 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.

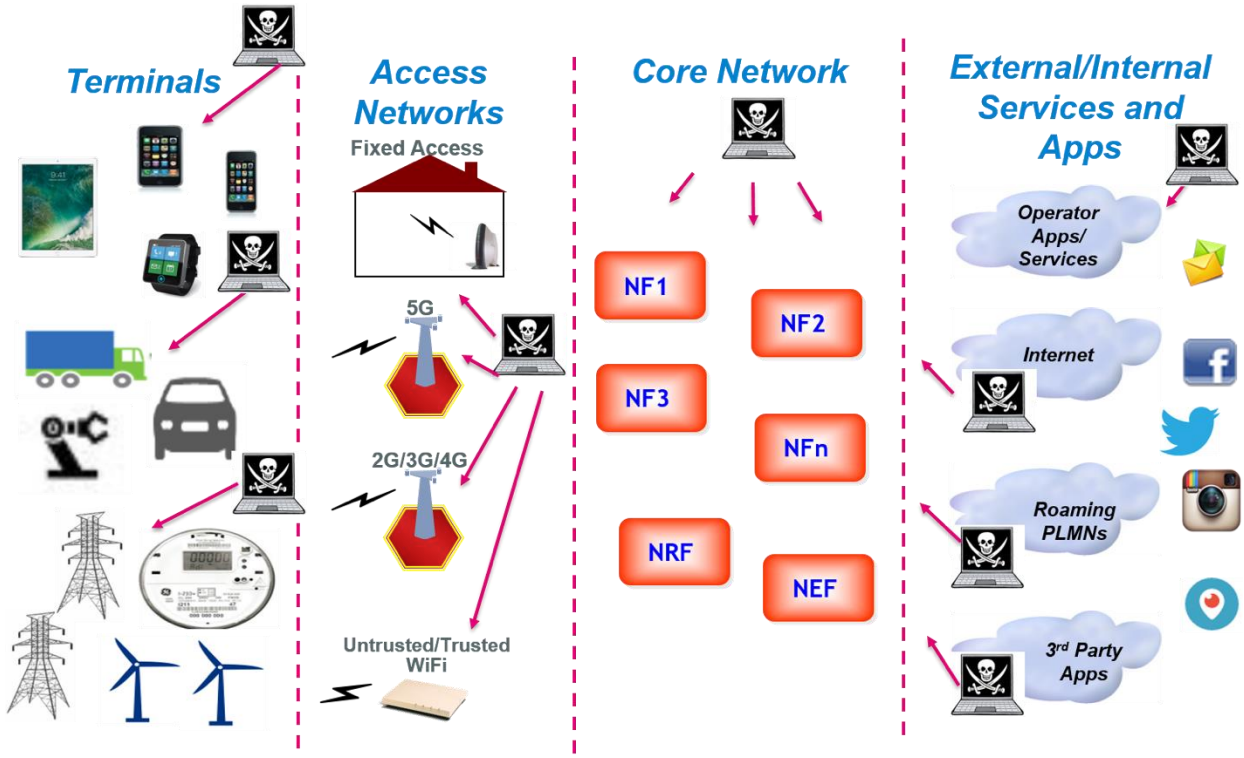


Figure 9. The 5G Threat Landscape.

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 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**: In this attack, 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

RAN THREATS

The fact that 5G will support many different access networks including 2G, 3G, 4G, and Wi-Fi means 5G perhaps inherits all 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 attach 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 not expected that mmWave by itself is less secure than any other part of the spectrum. The data and signaling transmitted and received at the radio layer is expected to be appropriately encrypted and integrity protected at higher layers, whenever possible.

ROGUE BASE STATION THREAT

One of the threats that face the different mobile networks, including potentially 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 use the RBS to launch different attacks on mobile users and networks. These attacks include stealing user information, tampering with transmitted information, tracking users, compromising user privacy or causing DoS for 5G services.

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 to RBS-based threats using, for example, the following threat vectors:

- 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)

SUBSCRIBER PRIVACY THREATS

Subscriber privacy has always been a top concern for the mobile industry, and as the 5G era begins, it's become even more of priority simply 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 of unknown RBSs tracking users in major cities and performing suspicious activities.

Figure 10 shows different potential exposure points for compromising subscriber privacy in 5G networks, using protocol attacks, malware attacks on 5G NFs and insider threats.

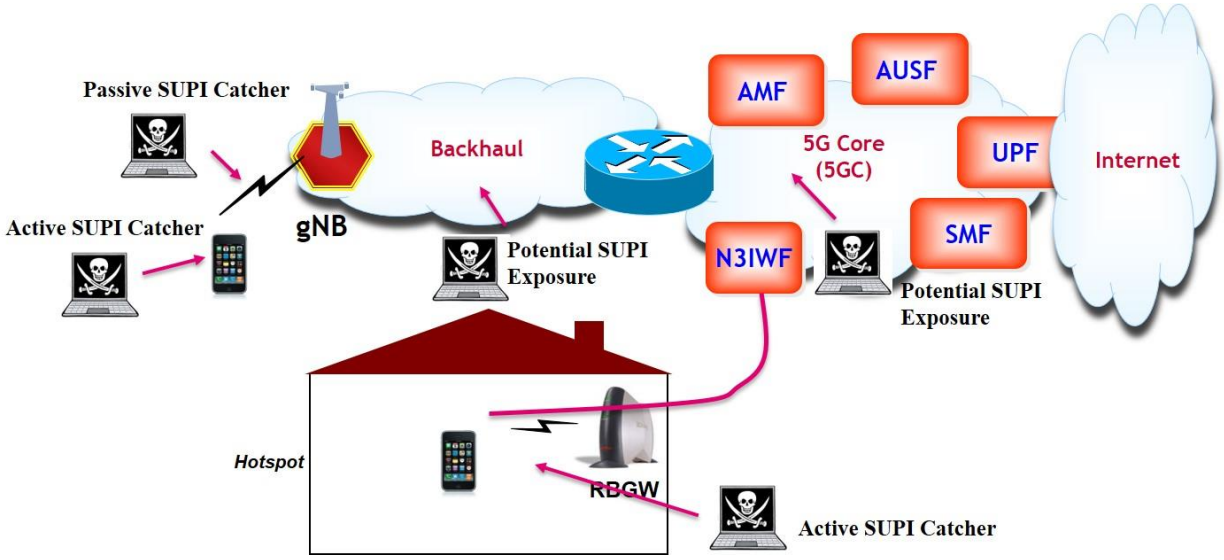


Figure 10. Exposure Points for Compromising Subscriber Privacy.

Note that there have been recent press reports about unknown individuals/groups operating IMSI catchers. Here the attacker takes advantage of an oversight in the original 3GPP mobile standards which 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 force UEs to use no encryption during calls or 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 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.

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 to degrade or make critical services 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. Because 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 IPSec encryption for non-3GPP access. An attacker can exploit the massive number of IPSec tunnel establishment requests by a large number of infected mobile devices simultaneously to launch DDoS on 5G core network functions.

NETWORK SLICING THREATS

A network slice is defined as an independent end-to-end logical network that runs on a shared physical infrastructure, capable of providing a negotiated service quality. Because network slices are a new concept and not yet deployed widely, there are no known physical attacks on them.

Among the features related to network slicing, several have potential security implications, such as the sharing of network functions and the isolation between the different slices. Network slices are expected to be a collection of multiple virtualized functions offering e2e service meant for certain features, such as IoT and eMBMS. When multiple network slices are instantiated over a common hardware platform, isolation of slices from one another is an issue. But this is expected to be addressed by the hypervisor of the virtualization platform hosting the network slice.

The UE needs to be authenticated and authorized for accessing the specific slice. Access to the slice is usually indicated using slice specific identifiers by the UE to the network signaling. If this signaling is not secured, access to the slice itself may be denied, resulting in a DoS attack. Hence, just as any other signaling to the core network, all signaling meant for access to the network slice, as well as UE-to-network slice signaling, needs to be protected.

NFV AND SDN THREATS

To efficiently support the new levels of performance and flexibility required for 5G networks, it is understood that new networking paradigms must be adopted, such as NFV and SDN. At the same time, though, these new techniques also bring new threats. For example, when applying NFV, the integrity of Virtual NFs (VNFs) and the confidentiality of their data may depend to a larger degree 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 quite 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, for its part, bears the threat that control applications may wreak havoc 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 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, DoS attacks or attacks aiming 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, such as usage of IPsec or TLS to cryptographically protect the legitimate communication and exclude communication by malicious third parties.

INTERWORKING AND ROAMING THREATS

Roaming in 5G applies some new protocols delivering new flexibility and new threats as compared to 4G. Following are a few items to consider about roaming in the 5G architecture, specifically pertaining to security delivering embedded security at the 5G roaming links by design:

- 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. MITIGATION CONTROLS FOR 5G NETWORK, IOT THREAT MITIGATION & DETECTION AND MITIGATION OF DDOS ATTACKS

5G NETWORK THREAT MITIGATION

This paper's introduction describes how the 5G threat surface is the widest in scope and the most complex due to a number of factors. This includes a widely distributed network of mobile edge compute, or smaller data centers pushing function closer to the "edge" to serve many of the use cases covered in this paper– specifically ultra-low-latency IoT use cases. One way to look at threat mitigation for a network with a threat surface as expansive as 5G is to break it down into parts and then look at the threats and mitigations for that part of the network.

The first part, shown in *Figure 11* and *Figure 12*, describes the threat surface of the "edge" and mitigation techniques applied at specific points in the network to solve for those threats.

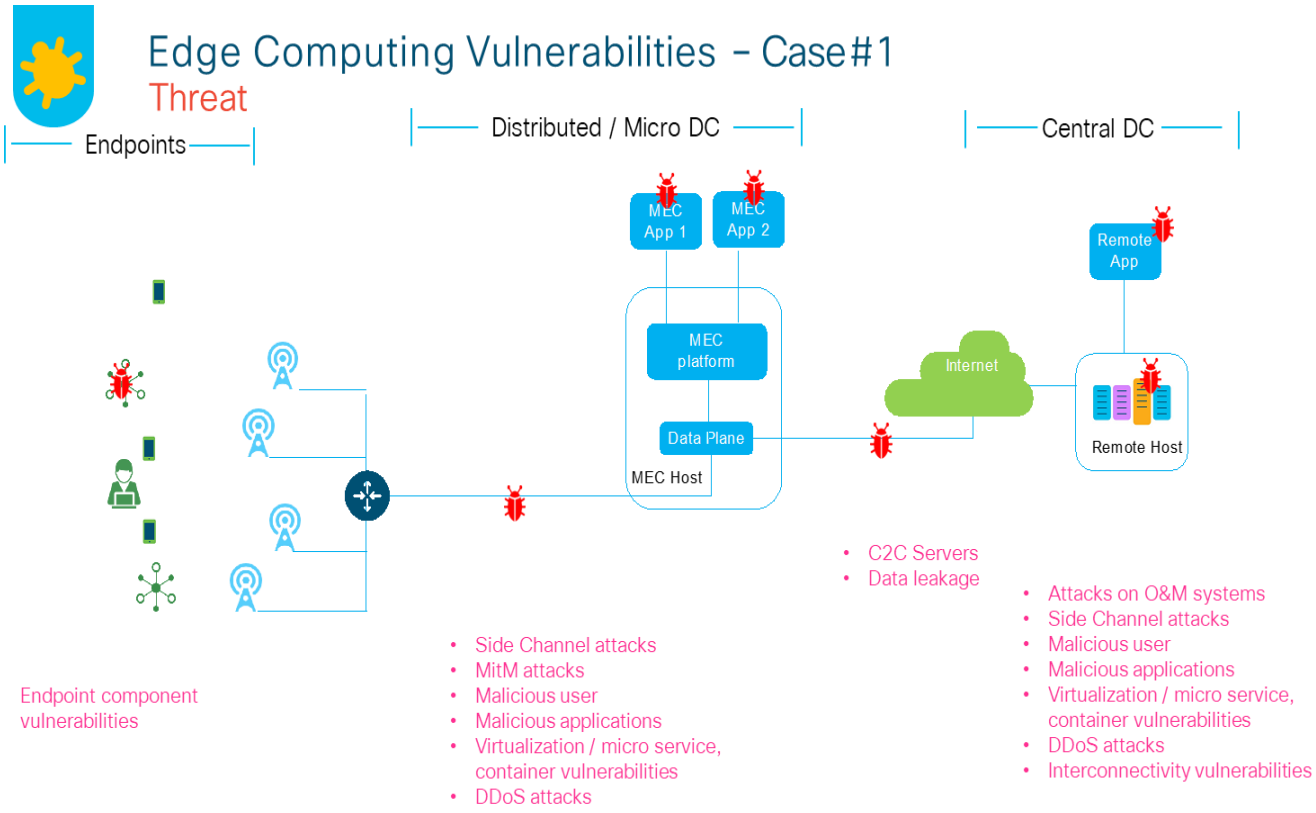


Figure 11. Edge Computing Vulnerabilities.

Figure 12 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 not allowing a botnet to be created to attack the RAN. Commonly used techniques at the DNS protection level allow the attacks to be thwarted at first step in the malware kill chain by stopping C&C communications with known bad talkers. This is just one simple example. Operators will have their own use cases that build on top of this foundation.

Foundational to all security is "visibility." This paper's introduction described the concepts of visibility and controls as the foundational elements of securing 5G. Visibility provides a constantly updated picture of how the network is behaving. Threat feeds make that picture operational against known and unknown threats. We use policy and segmentation to ensure that we know what is abnormal or an anomaly, and then we segment the network to avoid threats from spreading and compromising other functions or workloads. In the grey box in *Figure 12*, various controls that are used at this place in the network are called out 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."

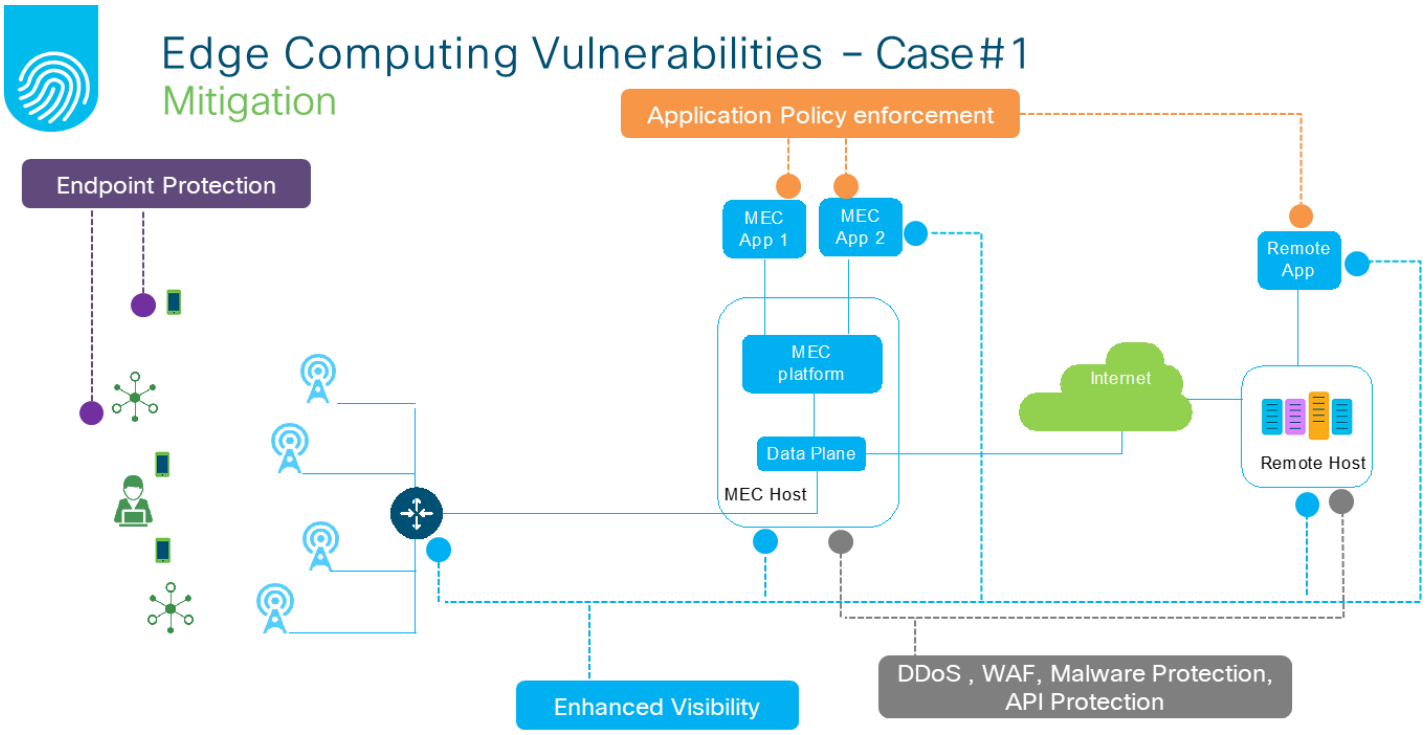


Figure 12. Edge Computing Vulnerabilities – Mitigation.

Figure 13 and *Figure 14* describe the distributed 5G Core and the associated threat surface. 5G brings in layers of orchestration, NFV, containers, micro-services and virtualized implementation of key evolved packet core functions.

5G Distributed Core Vulnerabilities – Case#2
Threats

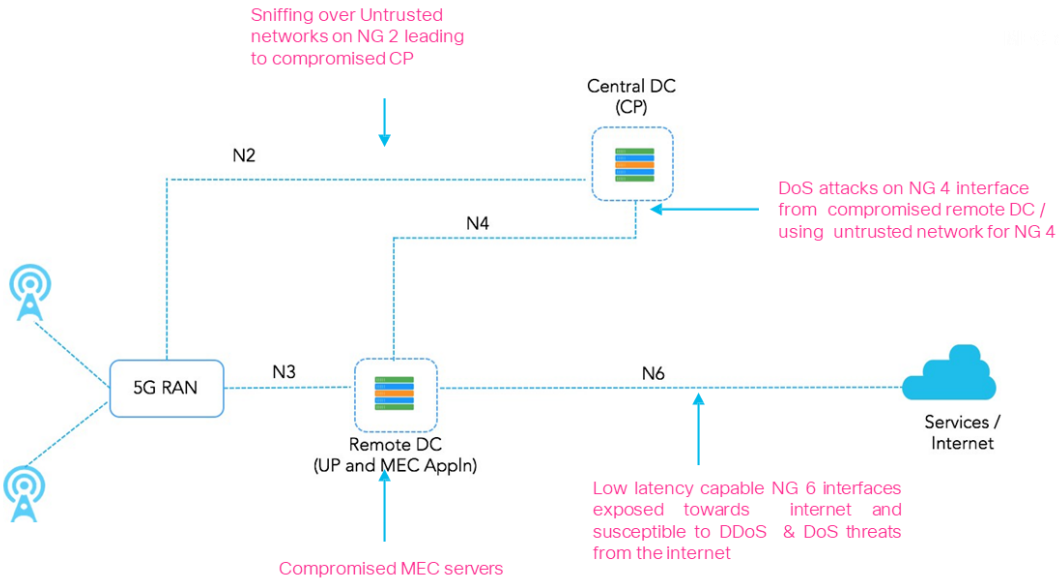


Figure 13. Distributed 5G Core Vulnerabilities.

Interface names in 5G change when compared to 4G with specific mapping in certain cases. One such case is the Gi/SGi in 4G, more commonly referred to as the N6 in 5G. Threats against the distributed core by interface include, but certainly aren't 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 14 shows the visibility points and the mitigation controls. Many of these controls are familiar to operators today. 5G brings challenges of distributed deployment, orchestration and scale-up and scale-out with automation to be able to keep up with threats on a distributed core architecture.

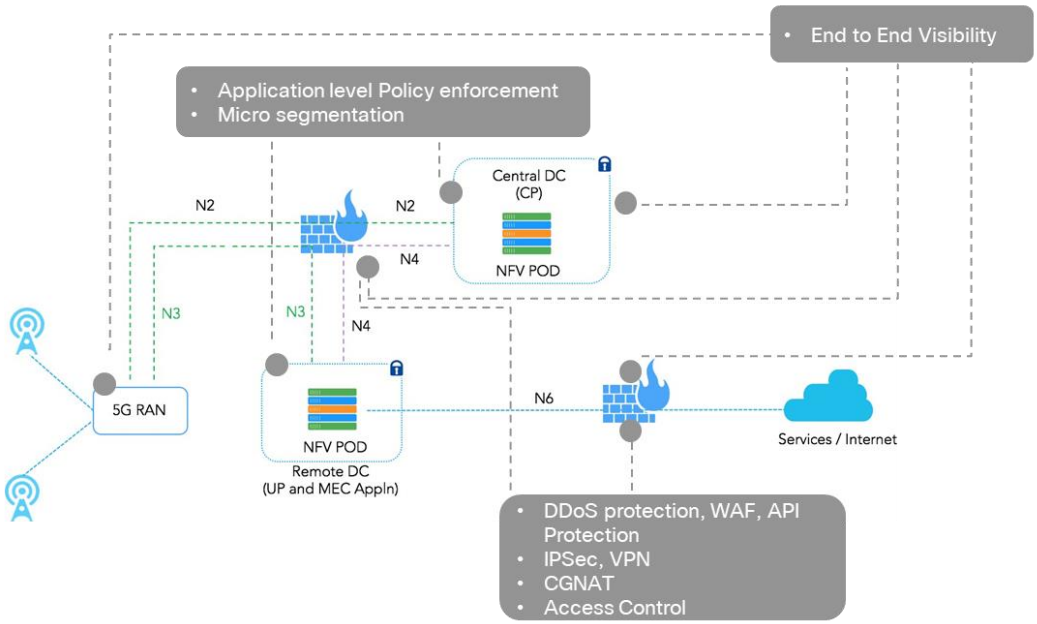


Figure 14 - Distributed 5G Core Vulnerabilities - Mitigation

The next part of the 5G architecture to be addressed is how to mitigate threats at the virtualization layer. Networks built today are highly virtualized in key NFs. 5G takes that all to a completely new level. The operator's back bone 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, which lives in the visibility area as described in Figure 15. On top of that infrastructure is an orchestrated NFVi layer. 5G brings with it a move to highly virtualized workloads and even, in certain cases, movement of certain key parts of the network to the cloud (CUPS model for Control and User Plane Separation). CUPS isn't a 5G feature per se, but it's another aspect of the new trust boundaries and threat surface of the 5G network deployments. Virtualized workloads bring a new set of threats that include, but certainly are not limited to:

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

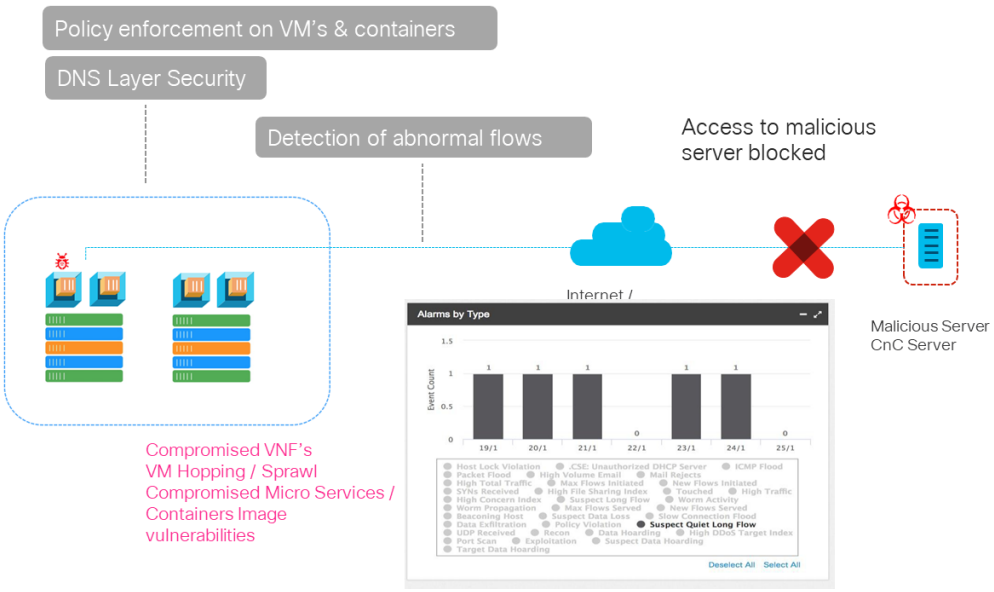


Figure 15. Virtualization Vulnerabilities

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 virtualized part of the 5G architecture.

Figure 16 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.

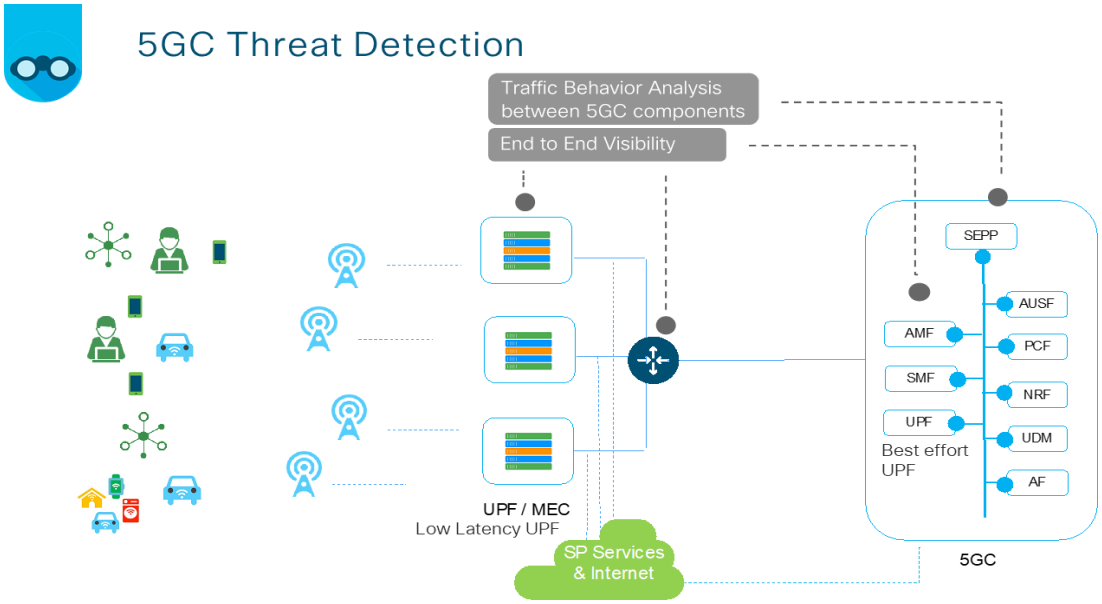


Figure 16. 5G Threat Detection

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

IOT 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.[8]

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 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, involving a 'root-of-trust' approach, to prevent compromised OS in *Figure 17.* Although this functionality has been typically 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.

[8] *IoT Security- Protecting the Networked Society*, Ericsson. https://www.ericsson.com/en/white-papers/iot-security-protecting-the-networked-society

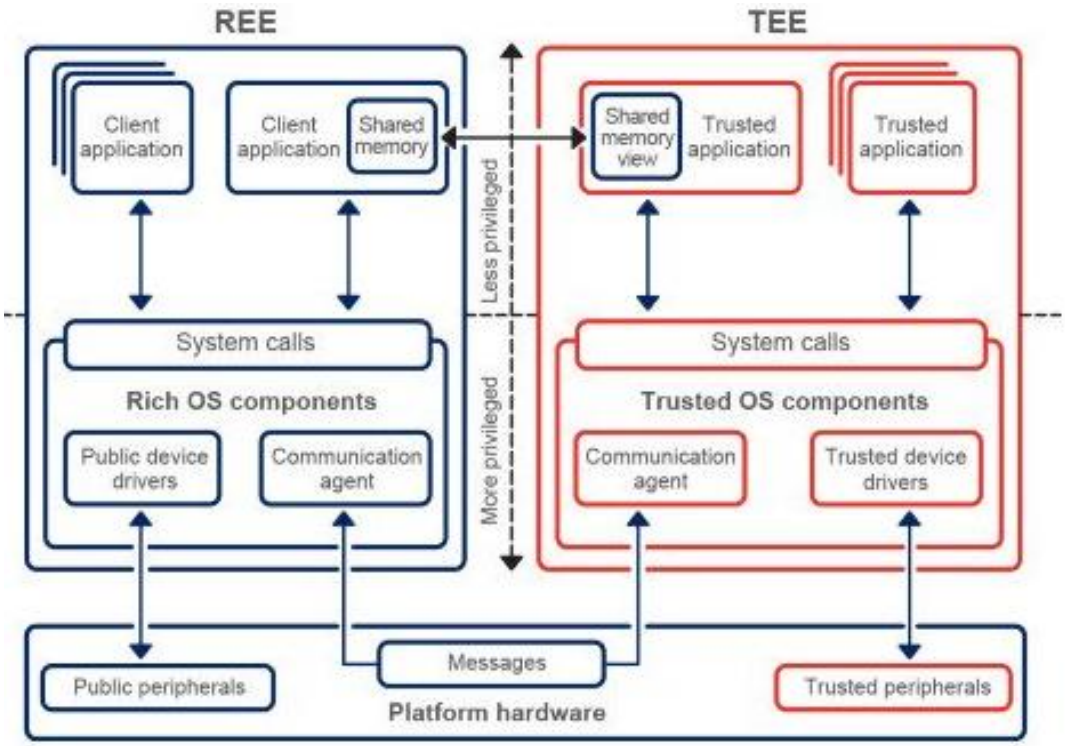


Figure 17. Root-of-Trust Approach.

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.

NETWORK/TRANSPORT

Mobile operators can leverage their unique position in the IoT space as both connectivity and platform providers. Technologies such as LTE-M and NB-IoT are superior solutions designed to provide global connectivity offering far higher robustness compared to unlicensed access. Mobile networks can enhance IoT security by providing device management and 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 management of credentials. By actually generating credentials on the device the risk of security breaches can be reduced. 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.

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 where a customer seeks a single service layer agreement with a single connectivity aggregator.

NODE/PLATFORM

IoT platforms can and should bear the responsibility for 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.

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 use of information containers at the application level, which are capable of confidentiality, integrity and origin authentication, is the preferred solution for protecting message exchanges, rather than at lower layers in the protocol stack.

SERVICE

To illustrate service level security, the modern connected vehicle scenario mentioned previously is used as an example. There exists 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. While accidents caused by malfunctioning machines will probably never be completely eliminated, yet 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, by integrating sensors and cameras into traffic lights, vehicles could become aware of pedestrians in advance. Emergency response is another area where IoT can make a significant positive impact. Free traffic lanes could automatically be created for emergency vehicles, tracking/finding missing children can become easier to find, 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.

SECURITY REQUIREMENTS FOR 5G NETWORK MASSIVE IOT THREATS

To prevent 5G service disruption caused by MIoT botnets used for DDoS RAN attacks, and to ensure 5G service resiliency, deliberate security requirements for the 5G network are needed. The fundamentals of these security requirements are detection and mitigation of DDoS attacks against the 5G RAN, which can also be 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 only be a half measure because the 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.

DETECTION OF DDOS ATTACKS AGAINST THE 5G RAN

To detect a DDoS attack against an operator's 5G RAN caused by MIoT botnets, the 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, which 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 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; and analytics algorithms to determine if it's a DDoS event, based on threshold, volumetric anomaly, timing, Radio Network Temporary Identifiers, and 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.

MITIGATION OF DDOS ATTACKS AGAINST THE 5G RAN

For the mitigation of a DDoS attack against an operator's 5G RAN, the same attack scenario will be considered. 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, thus 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.

PROTECTING 5G NETWORKS AGAINST DDOS AND ZERO DAY ATTACKS

5G networks are vulnerable to attacks on both the control and data planes. Following are some threats to the control and data planes, as well as strategies for mitigation.

The first example to keep in mind is on the control plane. Before the UE has an established connection (for example, to make calls), a series of messages must be exchanged between the eNB, gNB and finally

the MME. If an attacker is able to take control of a large number of devices and cause them to reconnect (for example, by restarting them), this could cause a signaling storm. Note that the 5G era, there can be 100x more devices, and 1000x more bandwidth per unit area, compared to LTE networks.

Another example is, what happens if an attacker uses legitimate devices on an operator's network to target either the operator itself or a third party to produce, for example, a denial of service attack? Such attacks create large amounts of traffic at the level of the data plane.

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

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 DNNs preform extremely well to detect attacks on the KDD99 dataset. 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[9] work well, as do approaches based on the Mahalanobis distance function[10] and auto-encoders.

This approach alone will produce many false positives. The trick to reducing the false positives is to recognize that denial of service attacks produces connections that share some commonality. Simple vertical features, for example, counting the number of anomalous connections per gNB, or with a given User Agent string (if applicable) or Type Allocation Code, can be used to build basic rules to reduce false positives in this stage of the pipeline. A better 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 CNN and used for anomaly detection.

5. 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 our “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 we use 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

[9] *Isolation forest*, Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. Eighth IEEE International Conference on Data Mining, ICDM ‘08. 2008.

[10] 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

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 we build and operate and will remain so into the unforeseeable 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.

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 in this context to note a few characteristics that distinguish 5G security from that of previous generations of cellular technologies.

- Previously in this paper, in the context of IoT, DDoS attacks coming from 5G RAN originated via botnet-controlled compromised devices were explained. 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. Notable examples of such services are automotive and health. The cost of a security breach for such services 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.

For this level of complexity, canned 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. Integrated AI-based defense mechanisms are likely to remain in the realm of research for few more years to come.

6. ACRONYMS

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
AMF	Access and Mobility Management Function
AUSF	Authentication Server Function
C&C	Control and Command
CU	Centralized Unit
CU-CP	Central Unit – Control Plane
CUPS	Control and User Plane Separation
DDoS	Distributed Denial of Service
DU	Distributed Unit
e2e	End to end
EAP	Extensible Authentication Protocol
eMBMS	Evolved Multimedia Broadcast Multicast Services, also known as LTE Broadcast
eUICC	Embedded UICC
FQDN	Fully Qualified Domain Name
HTTP	HyperText Transfer Protocol
IETF	Internet Engineering Task Force
IMSI	International Mobile Subscriber Identity
IPSec	Internet Protocol Security
IPX	Internetwork Packet Exchange
PEI	Permanent Equipment Identifier
ME	Mobile Equipment
MiTM	Man-in-the Middle
MIMO	Multiple-Input Multiple Output
MIoT	Massive Internet of Things

ML	Machine Learning
NAI	Network Access Identifier
NF	Network Function
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NR	New Radio
OS	Operating System
PCE	Path Computation Element
PEI	Permanent Equipment Identifier
RAN	Radio Access Network
RBS	Rogue Base Station
REE	Rich Execution Environment
RRC	Radio Resource Control
RU	Radio Unit
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
SMF	Session Management Function
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
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, 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.