WHITE PAPER

THE MOBILE CORE UNDER ATTACK

Securing the 4G/LTE EPC & preparing for 5G migration

TABLE OF CONTENTS

DISCLAIMER

RISING THREAT LEVELS FOR THE MOBILE CORE

Mobile networks are evolving quickly—and so are the vulnerabilities.

Over the years, carrier-grade telecom operators have had a good track record for security. However, new threats and technology changes have made the 4G/LTE mobile network—especially the core—more vulnerable than previous generations. The change from SS7 signaling to IP-based protocols has opened new attack avenues that can potentially penetrate the mobile core. Cyber criminals have increased their capabilities, leveraging automation and cloud tools to easily weaponize Internet of Things (IoT) and other mobile devices to attack both from within the network and via the Internet. Once inside, cyber criminals can exfiltrate sensitive customer data, eavesdrop on subscribers, compromise core resources and even bring down the entire network.

Until recently, most operators focused their security efforts on protecting the internet connection to the Evolved Packet Core (EPC) with a Gi/SGi firewall and DDoS protection. The other EPC interfaces were considered too complex to penetrate because they required deep vendor-specific knowledge. Put another way, the network complexity itself was considered a barrier to potential attackers.

Today, network operators are beginning to question the effectiveness of that strategy and look for more robust security solutions.

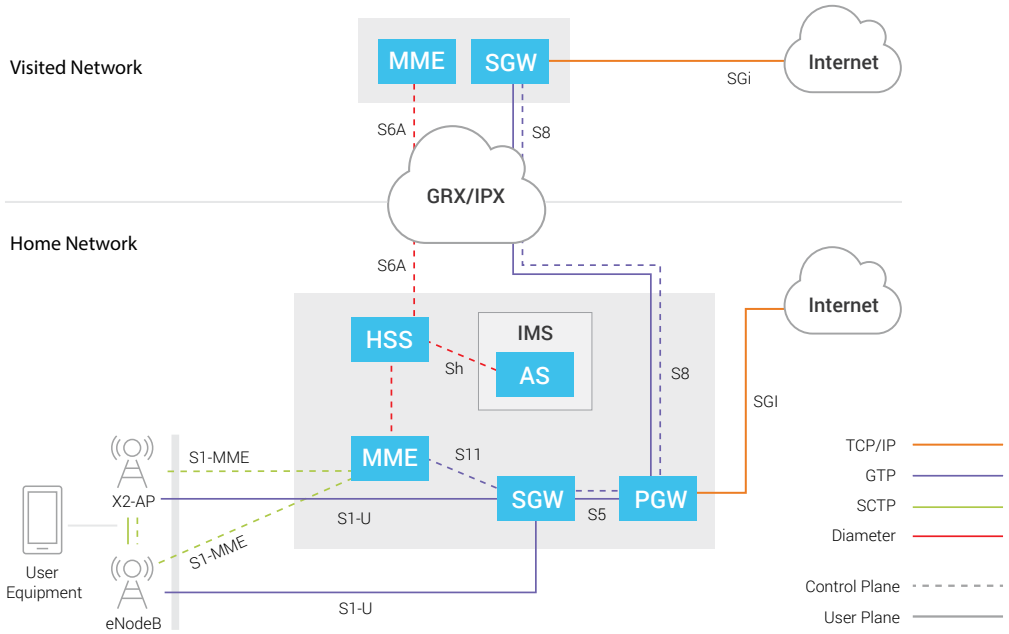
Increasingly the EPC is under attack at all three interfaces: the internet, roaming and radio access network (RAN). The exponential growth in users and traffic volumes and the evolution of network technology are growing the attack surface and multiplying the number of endpoints that can be potentially weaponized.

The coming migration to 5G and the rapid deployment a wide range of 5G use cases such as IoT will only put more pressure on mobile operators.

This white paper describes the escalating threat landscape for mobile operators and outlines a strategy for securing the mobile core using a comprehensive security stack. To get started, let's look inside the EPC to understand the components and interfaces that play into the security discussion.

INSIDE THE 4G EPC

For the purposes of security, the main elements of the EPC are the home subscriber server (HSS), the mobility management entity (MME), the serving gateway (SGW) and the packet data network gateway (PGW). These components interact with each other and the outside world via three primary protocols: GPRS Tunneling Protocol (GTP), Stream Control Transmission Protocol (SCTP) and Diameter (see Figure 1). Any of these can be targets for malicious activity, often exploiting the vulnerabilities of the applicable protocol.



INTERFACE	*PROTOCOL*	*USER PLANE*	*CONTROL PLANE*	
S1-U	GTP	✓		Transfers data from UEs to SGW
S1-MME	SCTP		✓	Controls eNodeB access to EPC via the S1-AP application
S5	GTP	✓	✓	Transfers data and communicates between SGW and internal PGW
S6a	Diameter		✓	Supports LTE access operations
S8	GTP	✓	✓	Transfers data and communicates with external PGW
SGi	Generic IP			Transfers data between PGW and external packet data networks (PDNs)
Sh	Diameter		✓	Communicates subscriber information between HSS and application server
X2-AP	SCTP		✓	Allows eNodeBs to communicate with each other for handoff, load-balancing and congestion control
S11	GTP		✓	Used between MME and SGW to support mobility and bearer management

Figure 1: Key security components and interfaces in the EPC

THE EXPANDING ATTACK SURFACE

The threat landscape is rapidly changing. Attacks can come from any point where the EPC connects to the outside world. In the past decade, numerous academic research papers have shown that attacks from partner networks or RANs were a possibility. Those threats are no longer merely an intellectual exercise—they are beginning to show up in the real world.

One such category of attacks is distributed denial of service (DDoS).[1] Recently, botnets such as WireX and its variants have been found and disarmed. In the past, these attacks targeted hosts on the internet, but now they've started attacking EPC components. In previous generations, core components were hidden behind proprietary and obscure protocols. However, 4G and later generations exposed these components to DDoS attacks via protocols such as TCP/IP, UDP and SCTP.

In addition, the adoption of IoT is exposing the core to the threat of malicious actors taking control and weaponizing devices against a service provider. The exponential growth of IoT has the potential to overwhelm defenses, as the number of connected sensors and devices grows into the billions.

Bottom line: The attack surface of the mobile core is significantly larger than it used to be, and legacy approaches to security are no longer effective.

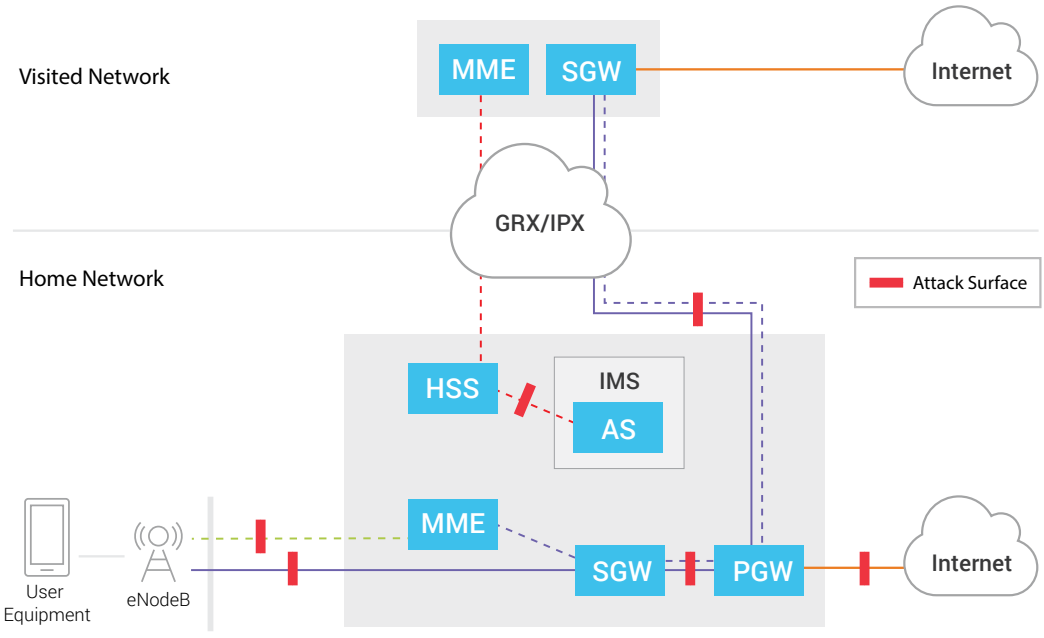


Figure 2: Attack surfaces in 4G/LTE mobile network

1. Bryon Acohido, "Why DDoS Weapons will proliferate with the Expansion of IoT and the coming of 5G," The Last Watchdog, March 29, 2019. https://www.lastwatchdog.com/my-take-why-ddos-weapons-will-proliferate-with-the-expansion-of-iot-and-the-coming-of-5g/

NEW ATTACK VECTORS

An important first step in upgrading core security is clearly understanding the kinds of attacks that leverage the key protocols GTP, SCTP, Diameter and TCP/IP. The following examples are some of the hundreds of protocol-based attacks that have been documented to date.

GTP-BASED ATTACKS

Many GTP attacks start by obtaining information such as the Tunnel Endpoint Identifier (TEID). To obtain a TEID by brute force, an attacker can send a series of GTP-U messages, each with a different TEID. If the TEID in the message is incorrect, the PGW obligingly sends a GTP-C "error indication" message. Therefore, the lack of such a message indicates a legitimate TEID value. Attackers can discover other parameters such as Temporary Mobile Subscriber Identity (TMSI) and International Mobile Subscriber Identity (IMSI) with the help of a fake base station or an IMSI catcher.[2]

- Once they have obtained such credentials, cyber attackers can initiate a wide range of GTP-based exploits such as:
- Eavesdropping and information gathering: Intercepting and snooping into GTP traffic to exfiltrate valuable subscriber information such as user location, APN credentials and session keys.
- Fraud: Leveraging subscriber IMSI to consume services at the expense of the operator or another subscriber.
- Malicious GTP messages: Injecting malicious GTP messages to disrupt sessions, gathering information and redirecting traffic
- Message flooding: Overwhelming the PGW with Create Session Request messages to exhaust the IP addresses assigned to the PGW, causing service degradation and network outages.

SCTP-BASED ATTACKS

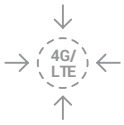
SCTP is used for several important interfaces in the EPC, notably, the S1-MME control plane interface between the eNodeBs and the MME and the X2-AP control and user plane interfaces between peer eNodeBs.

SCTP is susceptible to threats such as association hijacking and address stealing that are launched from a rogue eNodeB. In this type of attack, a rogue eNodeB takes over a legitimate SCTP association—from the standpoint of the MME, nothing has changed. However, now the attacker has direct access to the MME and can disrupt mobile operations and even take down the EPC.

In another tactic, the rogue eNodeB sends SCTP messages to all eNodeBs in the local network, instructing them to disable their radio transmitters and thus preventing subscriber access to the mobile network.

RAN-based attacks require getting access to the back-haul network, either physically or via a microwave, DSL, or carrier Ethernet network segment. The explosive growth in cell towers coupled with the proliferation of less secure small cells will multiply the opportunities to compromise a physical site and initiate these crippling exploits.

2. "Threats to packet core security of 4G network," Positive Technologies, September 28, 2017. https://www.ptsecurity.com/ww-en/analytics/epc-research/



DIAMETER-BASED ATTACKS

Security researchers say the Diameter protocol used in 4G/LTE networks is vulnerable to the same basic kinds of attacks as the SS7 standard used in previous generations. Diameter is inherently safer than SS7 because it supports Transport Layer Security (TLS), Datagram Transport Layer Security (DTLS) and Internet Protocol Security (IPsec) encryption. However, operators don't always take advantage of the benefits of encryption, leading to attacks such as:[3]

- Subscriber/network information disclosure: Gathering operational information about the user's device, subscriber profile and the mobile network itself.
- Traffic interception: Downgrading a Diameter-capable 4G connection to a previous-generation connection and using flaws in SS7 and other protocols to intercept SMS traffic.
- Fraud: Exploiting Diameter flaws to allow free use of the mobile network for a specific subscriber profile.

TCP/IP-BASED ATTACKS

The TCP/IP protocol itself is vulnerable to a wide range of attacks, including password sniffing, IP spoofing, TCP sequence number attack and TCP session hijacking. A particularly common threat is denial of service (DoS) flooding, in which the attacker overwhelms the link with a flood of malicious UDP packets. Mobile users often connect via public IP, which offers convenience but exposes both the user and the network to attacks such as DoS flooding, quota drain and battery drain.

Attackers can also exploit vulnerabilities in the protocols that run on top of TCP/IP, for example, SMTP, NFS, FTP, Telnet, NTP and HTTP. Taken together, the threat level to the EPC from TCP/IP traffic is rising, and operators need to enhance their security strategies to address these vulnerabilities.

3. Catalin Cimpanu, "Newer Diameter Telephony Protocol Just as Vulnerable as SS7," Bleeping Computer, July 2, 2018. https://www.bleepingcomputer.com/news/security/newer-diameter-telephony-protocol-just-as-vulnerable-as-ss7/

DOS/DDOS ATTACKS ON THE RISE

Mobile operators strive for very high levels of network availability. Network outages and disruptions can cost millions of dollars and can contribute to lost revenue and subscriber churn. Malicious DoS and DDoS attacks, as well as other non-malicious events can overwhelm network resources and cause outages or disruption. A recent survey of security professionals in the mobile industry confirmed that more than 90 percent of respondents view this form of attack as a critical issue.[4] Virtually all providers deploy some form of DDoS protection at the Gi/SGi interface.

Less attention has been focused on attacks originating from other parts of the network. Academic researchers have, for years, pointed out vulnerabilities in the RAN and roaming interfaces. A good example can be found in a paper published in 2018 by researchers at Purdue University and the University of Iowa.[5] Using a model-based testing approach, the team identified 10 new attacks that could compromise network security, invade user privacy and disrupt service. One such attack is the authentication relay attack that enables an adversary to spoof the location of a legitimate user to the core network without possessing appropriate credentials.

The roaming (GRX/IPX) interface is also susceptible to DoS attacks. One tactic is to generate malicious GTP messages that interfere with subscriber sessions, disrupt billing and even deny service to all eNodeBs on a given SGW. The PGW is equally susceptible using the GTP flooding attack described earlier. Operators must offer seamless interconnection so large numbers of interconnection agreements are essential. Security of the roaming interface is essential to meeting these business objectives.

Service-impacting DDoS attacks can also be caused by non-malicious events originating through the RAN or roaming interface, for example:

- Roaming partner equipment malfunction or software updates that send invalid or malformed packets or a flood of messages to the visited network's PGW.
- Large volumes of connected mobile or IoT devices that simultaneously request network resources or reconnection due to hardware or software malfunctions or local network outages.

4. Catalin Cimpanu, "Newer Diameter Telephony Protocol Just as Vulnerable as SS7," Bleeping Computer, July 2, 2018. https://www.bleepingcomputer.com/news/security/newer-diameter-telephony-protocol-just-as-vulnerable-as-ss7/

5. Syed Rafiul Hussain et al, "LTEInspector: A Systematic Approach for Adversarial Testing of 4G LTE," January 2018. https://www.researchgate.net/publication/323248235_LTEInspector_A_Systematic_Approach_for_Adversarial_Testing_of_4G_LTE

THE IMPACT OF IOT

The rapid adoption of IoT poses a major security risk to mobile networks. By 2024, there will be an estimated 22.3B IoT devices worldwide – an estimated 4.1B in cellular networks.[6] Unlike smartphones, most IoT devices have limited computing resources and battery reserves. They are often deployed in remote areas or in environments where their behavior is difficult or impossible to monitor.

By 2024, there will be an estimated **22.3B IoT devices worldwide** – an estimated **4.1B in cellular networks**.

The diversity of devices is exploding. One researcher estimates that 127 new devices are connected to the internet every second.[7] Applications range from consumer entertainment wearables to critical medical and infrastructure and power grid monitoring. The operator has even less control over these devices than they have for smartphone users and their applications. A large percentage of these IoT devices will be used in roaming applications. With significant price competition, manufacturers have little incentive to add in security mechanisms, and the devices themselves often don't have the computing power to support it. As a result, large numbers of potentially compromised or poorly designed IoT devices can be deployed within the operator mobile network, and, if compromised, could severely impact network availability. The Mirai DDoS attack several years ago showed how common, lightly-secured devices can be recruited into botnets for malicious purposes.

Faced with these realities, mobile operators are optimizing their networks to support a range of IoT applications, for example, narrow-band Internet of Things (NB-IoT). NB-IoT is one of three low power wide area (LPWA) technologies standardized by 3GPP (Release 13) for operation in licensed radio spectrum. NB-IoT, LTE-M and EC-GSM-IoT are all designed to reduce power consumption and improve coverage ability for IoT applications that have low data rates, require long battery lives and low cost and operate in hard-to-reach locations. Among other characteristics, NB-IoT enables small transfers of data between eNodeBs and MMEs and between the MME and the SGW. LPWA technologies and capabilities are an important driver for the high growth in cellular IoT connections.

In all these cases, the PGW or MME can be overwhelmed or falter, denying service to legitimate subscribers. As IoT devices and network interconnection complexity and density increase through 5G evolution, there is greater risk of disruption or outage from these events.

6. Ericsson 2018 Mobility Report, 4Q, 2018. https://www.ericsson.com/en/mobility-report

7. Mark Patel et al, "What's new with the Internet of Things?" McKinsey & Company, May 2017. https://www.mckinsey.com/industries/semiconductors/our-insights/whats-new-with-the-internet-of-things

SECURING THE MOBILE CORE: A NEW APPROACH

In recent years, operators have added security incrementally in response to specific perceived threats. This reactive strategy has led to unnecessary complexity in the core and Gi-LAN, and worse, gaps in the security posture. In the 4G/LTE world, this approach is risky and inefficient. As operators migrate to 5G, the need to boost security grows significantly.

Mobile operators need a strategy for EPC security designed from the ground up to secure the EPC against today's threats and futureproof to streamline the coming migration to the 5G core. This solution must provide the scale and performance required to grow the business, as well as architectural flexibility to accommodate network function virtualization (NFV), software-defined networking (SDN) and mobile edge computing (MEC). We refer to this solution as the comprehensive security stack (CSS).

This rest of this section describes the recommended Comprehensive Security Stack defenses at each of the three EPC interfaces (see Figure 3).

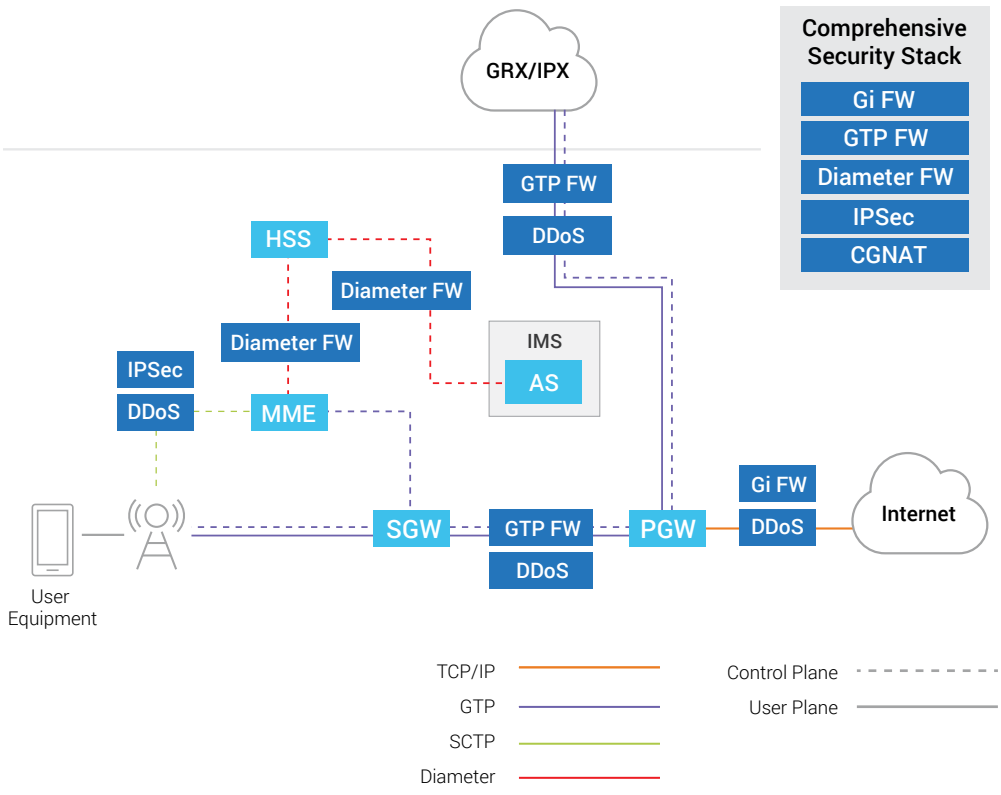


Figure 3: How the Comprehensive Security Stack (CSS) protects the EPC in 4G/LTE networks

GI FIREWALL

The Gi/SGi firewall defends against attacks from the internet, public and private clouds, data center infrastructure and other PDN gateways.

GTP FIREWALL

The GTP (roaming) firewall with granular SCTP filtering defends the EPC against GTP-based attacks initiated from RAN or GRX/IPX networks.

IPSEC

The 3GPP standard requires encryption for the air interface to each eNodeB, but there is no such requirement for the S1-U interface between the eNodeBs and the SGW or the S1-MME interface between eNodeBs and MME. IPsec provides both encryption and authentication in the mobile back haul. This protects against rogue or compromised eNodeBs or small cells.

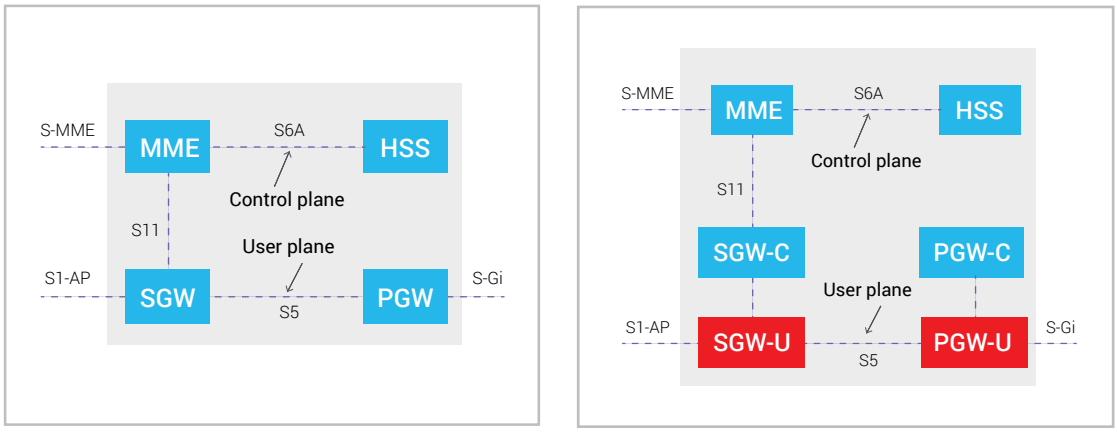
DDOS PROTECTION

As discussed earlier, DDoS protection is required at all three interfaces. The DDoS components must defend against multiple classes of attack vectors, including volumetric, protocol, resource and advanced application-layer attacks, which must be quickly detected and mitigated to prevent service disruption.

EVOLUTION OF THE 5G CORE

The 5G core is an evolution of the 4G EPC that can be thought of as two sequential steps:

1. Separate the control- and user-plane functions of EPC nodes (see Figure 4).
2. Reorganize the EPC functions into 5G services (see Figure 5).



a. EPC Before CUPS

b. EPC After CUPS

Figure 4: Evolved Packet Core, before (a) and after (b) Control and User Plane Separation (CUPS)

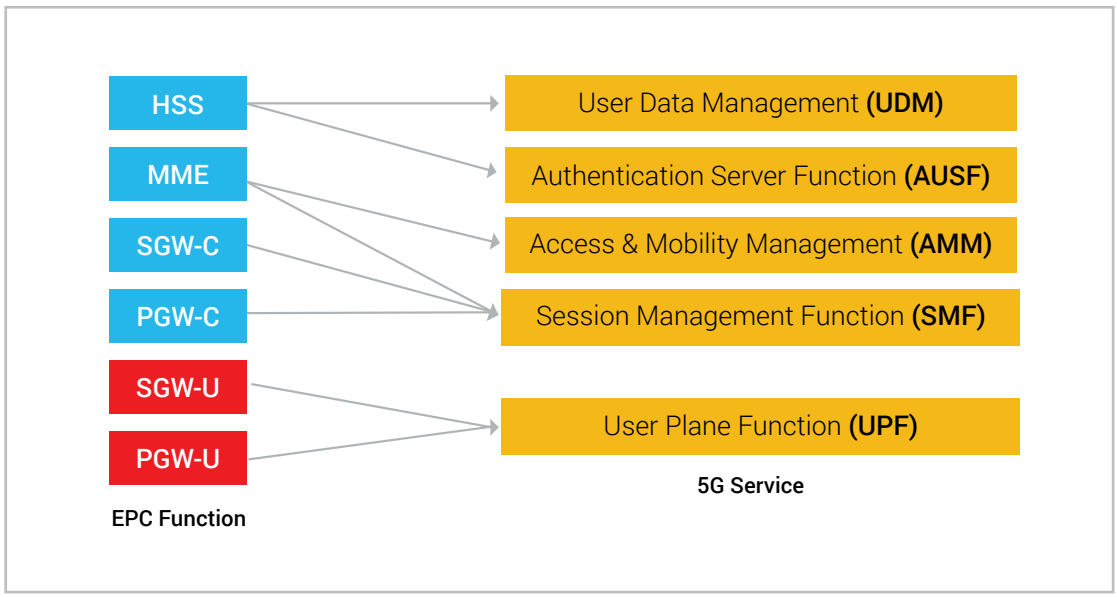
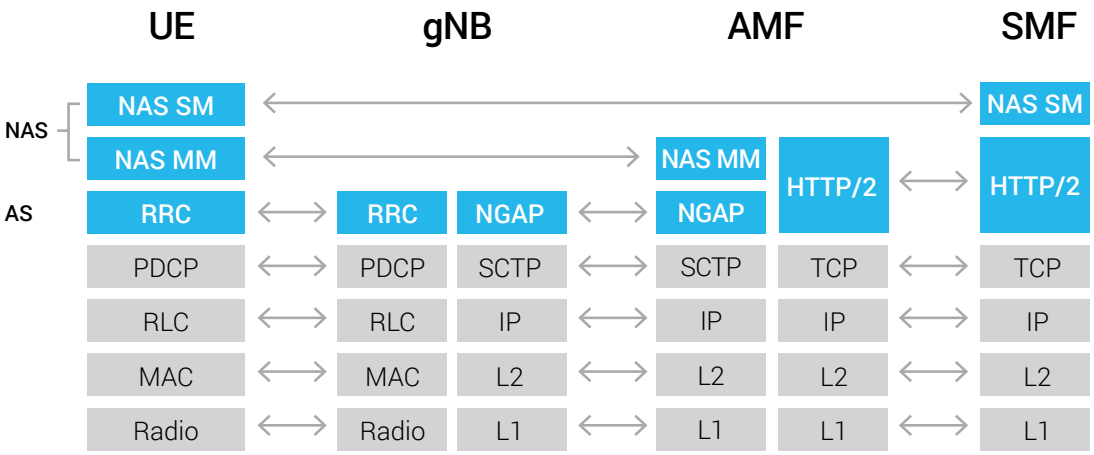


Figure 5: EPC functions mapped to 5G services

5G PROTOCOL STACKS

Three major use-case categories have been identified for 5G: Enhanced Mobile Broadband (eMBB), ultra-reliable and low-latency communications (URLLC) and massive machine-type communications (mMTC) latency. As a result, 5G can potentially support thousands of individual applications, each with varying security, latency and other functional requirements. To enable the flexibility needed to support such a wide range of use cases, the 5G core was designed as a service-based architecture using HTTP/2 and eliminating the use of many of the proprietary mobile network protocols such as GTP, replacing them with standardized control-plane stacks using SCTP, IP, TCP and other standard protocols (see Figure 6).



SYMBOL	*MEANING*	*STANDARD*
AMF	Access and Mobility Management Function	5G
AS	Access Stratum	5G
gNB	Next-generation NodeB	5G
NAS	Non-Access Stratum	5G
NGAP	NG Application Protocol	5G
PDCP	Packet Data Convergence Protocol	LTE
RLC	Radio Link Control	LTE
RRC	Radio Resource Control	LTE
SMF	Session Management Function	5G
UE	User Equipment	5G

Figure 6: 5G control plane protocol stacks

NEW SECURITY CHALLENGES IN 5G

There won't be a point where the service provider throws a switch and the network suddenly transitions from 4G to 5G. Instead, most operators will design a migration path in which they gradually modernize the existing EPC over a period of several years. During that time, the core will be a hybrid of LTE and 5G technology. 3GPP has defined reference architectures for both fully 5G (standalone) and the hybrid LTE-5G (non-standalone) in 3GPP TSG Release 15, knowing that most operators will take a gradual approach to their migration.

Three technology trends that have evolved alongside 5G—network function virtualization, mobile edge computing (MEC), and the shift to Internet-based protocols have an impact on core security.

NETWORK FUNCTION VIRTUALIZATION

Network function virtualization (NFV) is a key requirement for 5G. NFV helps operators drive efficiency and agility in their networks and ultimately create new value. However, it's not as simple as porting all physical network functions (PNFs) to virtual network functions (VNFs). Some functions can easily move from physical to virtual, while others—for example, core routers— may remain physical to meet latency and reliability requirements. In addition, the mix will change over time, so form-factor flexibility is essential during the 4G/5G transition.

Virtualization brings a new set of security challenges that are different from physical security. The management console itself is an attack vector, and so is each virtual machine (VM). Unlike a physical server that has a fixed amount of computing and storage hardware, a compromised VM can be used to consume virtually unlimited amounts of resources from the resource pool. When multiple VMs fall victim to these denial-of-service attacks, the host can become unusable. This attack is difficult to protect against, because it's not easy to identify malicious resource usage. Over usage can take many forms, from pegging a VM's CPU with 100 percent utilization, writing to all the memory assigned to a VM or causing an extremely high volume of disk reads and writes.

To implement effective security in a virtual environment, the main task is to secure the virtual host's management console. In a virtual infrastructure, the management network should be physically and virtually isolated. Hosts and clients should connect to a separate physical network to secure the traffic.

MULTI-ACCESS EDGE COMPUTING

Originally called mobile edge computing, the name change reflects the large number of access points in the modern mobile network, including radio towers, WiFi access points, and small cells. In MEC, the edge becomes a distributed cloud that moves services closer to the end user, resulting in ultra-low latency, higher reliability and greater scalability. When the edge network experiences high traffic, the edge can offload data to the cloud to maintain a reliable connection and avoid performance degradation.

MEC is one element of a general rearchitecting of the mobile network to meet more stringent requirements for individual use cases. For example, video content delivery and IoT applications can be deployed in edge data centers to reduce latency. In contrast, BSS and OSS elements, which are business critical but not latency sensitive, can be centralized in a core data center with strong disaster recovery capability.

Unlike the 4G EPC, the 5G core is distributed between the central office and edge locations. As a result, security must be distributed to each MEC location. Form-factor flexibility is a critical need in distributed security to adapt to a range of facilities and operating environments.

5G PROTOCOLS

5G uses common Internet protocols such as HTTP, TLS and REST API, replacing mobile protocols like SS7, Diameter and GTP. While this may eliminate some of the known protocol vulnerabilities of Diameter and GTP, it also makes the mobile network more accessible to the large base of cybercriminals already well versed in HTTP-based attacks. According to ENISA report, "Vulnerabilities for those type of protocols are quickly discovered and exploits are integrated into all kind of penetration testing tools readily available."

CHOOSING A PARTNER FOR CORE SECURITY

As network architects create their 5G migration plans, they must choose vendors that know how to facilitate an orderly and secure transition from 4G/LTE to 5G. Vendor selection is a complex process involving both technical and business considerations. However, many companies rely heavily on feeds and speeds and overlook the intangibles that ultimately determine the success or failure of a major infrastructure initiative. Here are three recommended criteria to incorporate into the vendor search process.

FORM-FACTOR FLEXIBILITY

The evolution of the mobile core during the 4G/5G transition make form-factor flexibility an imperative. The same network functions should be available in physical and virtual form and be able to run in a microservices/containers environment. The mix will change over time. Appliances with several physical functions can be gradually disaggregated into virtual form, which enables cloudification. Considerations such as latency and reliability for individual use cases may require that a given function exist in physical, virtual, or container form simultaneously.

OPERATIONAL INTEGRATION

Given the distributed nature of the mobile core, the security functions must naturally be disaggregated in an architectural sense. However, in the transition from a traditional physical environment to 4G and 5G, the emphasis must be on operationalizing efficiency and cloudification. The goal is to operationalize the function once and then deploy the function wherever needed in the architecture. This approach speeds redeployment of security functions as the network architecture changes and lowers operating costs.

FULL-SPECTRUM SECURITY

A comprehensive security strategy for the mobile core requires multiple technologies, including DDoS protection, IPsec, CGNAT and L4-L7 firewalls. Look for vendors that have proven expertise in each of these technologies and a substantial track record of successful deployments.

CONCLUSION

Service providers are challenged with evolving security vulnerabilities and scaling requirements in explosively-growing 4G and 5G mobile networks. These are threatening service availability, customer retention and potential obsolescence. The move from 4G to 5G is a journey, not a leap, enabling strategic investment now to maintain, and gain, competitive advantage, while reducing overall TCO.

As you plan your 5G migration, consider A10 Networks. A10 Networks delivers advanced security, reliability and performance for user and control-plane protection, with significantly lower latency. A10 Networks Thunder CFW is designed to meet the hyperscale needs of new 5G service applications, including enhanced Mobile Broadband (eMBB), ultra-reliable and low-latency communications (URLLC) and massive machine-type communications (MMTC).

ABOUT A10 NETWORKS

A10 Networks (NYSE: ATEN) provides Reliable Security Always™ through a range of high-performance solutions that enable intelligent automation with deep machine learning to ensure business critical applications are protected, reliable and always available. Founded in 2004, A10 Networks is based in San Jose, Calif., and serves customers globally with offices worldwide.

For more information, visit: a10networks.com or tweet @A10Networks.

LEARN MORE

ABOUT A10 NETWORKS

CONTACT US

a10networks.com/contact

Part Number: A10-WP-21154-EN-02 MAY 2019