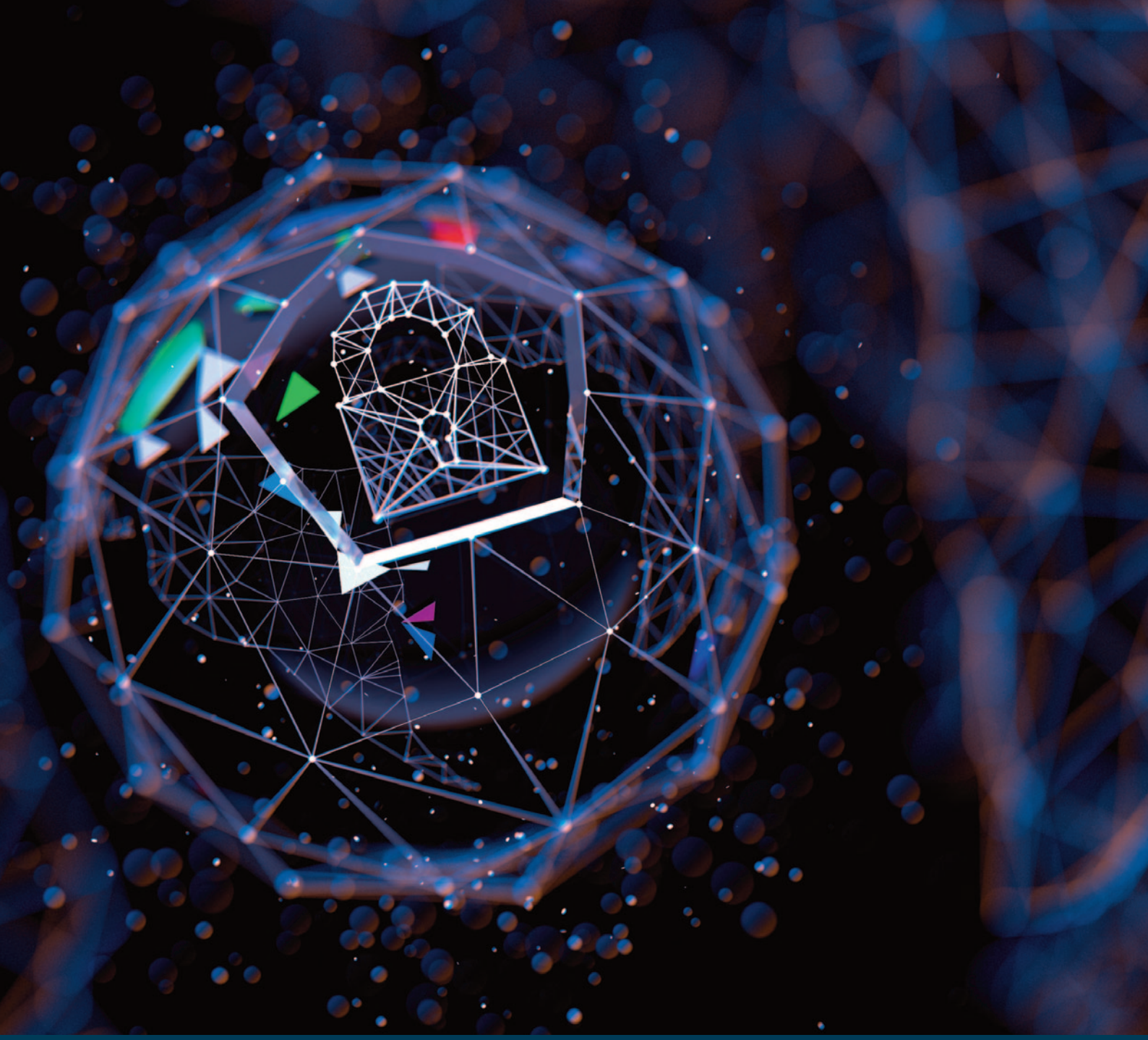


Why 5G Networks and Services demand Security-by-Design

Published by

MOBILE WORLD LIVE

Sponsored by

NetNumber

Introduction

Although 5G can be seen as simply the next generation in the 3GPP cellular progression, it is unlike previous generations because it has been designed as a native, cloud-based network infrastructure to support the challenging demands of low latency, high device density and enhanced mobile broadband. These capabilities present attractive new opportunities for cellular operators and for the world in general but the more 5G enables, the greater the risks. With 5G capabilities being relied upon for sensitive services such as healthcare, logistics and automated driving, it's clear that security breaches cannot be accepted and the ability to monetise effectively depends on users – corporations or individuals – being confident that communication and data are secure.

In many markets data protection is mandated, such as with the General Data Protection Regulation (GDPR) in the European Union, and there are similar regulations in other markets globally. These requirements and the demands of the increasingly network-reliant marketplace mean that security is even more of a priority in the 5G era than in previous cellular generations. For network operators themselves, it's likely that 5G services may demand a premium especially if 5G capabilities such as quality assurance are bundled into the proposition. This, in turn, makes the stakes for fraud and theft from the operators themselves higher and places even greater pressure on securing 5G.

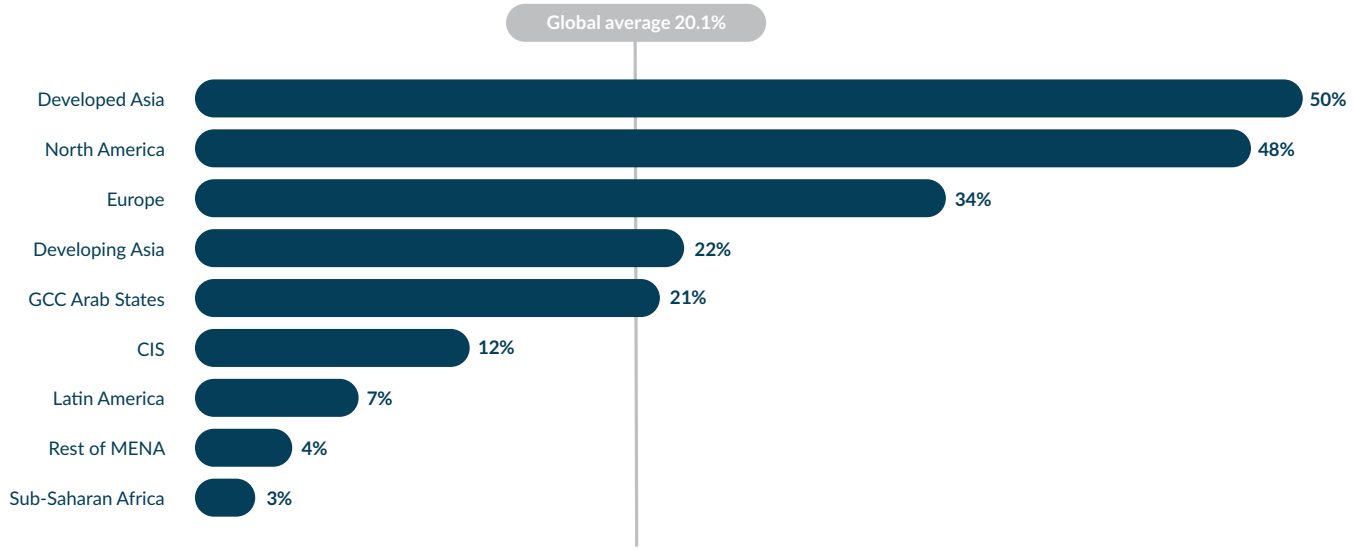
5G is a more complex and multi-layered environment than 2G, 3G and LTE (4G) and therefore security must be achieved by design rather than by a set of reactive, retrofitted solutions. In addition, 5G will not exist in isolation because 2G, 3G and LTE mobile networks will continue to operate and support popular services, such as SMS which 5G does not yet. There may also be issues with spectrum availability and refarming in some markets, that has the potential to delay 5G roll-out. This can also result in a prolonged period, during which multiple generations of cellular networks will operate and need to be secure across the generations. This intergenerational challenge is at the heart of 5G security design because it will be a reality for most of the coming decade, and potentially beyond in order for 5G users to be protected against 2G, 3G and LTE vulnerabilities.

5G security presents a radically extended threat surface because of all the new services and new value it enables, but also because of the new types of security vulnerabilities it introduces. There are both network and service-related security issues that need to be addressed in advance of network launch and service introduction. Concepts such as exposed network application programme interfaces (APIs) and compliance with recent regulation and legislation demand trusted relationships between operators and partners and these must be managed and administered in ways that are accepted and understood across the industry. Importantly, security cannot be achieved by a single operator acting in isolation because the service landscape is dependent in interactions between multiple operators, not just for roaming services, but for many international, pan-regional or global offerings, such as those in the Internet of Things (IoT) market.

5G Security Challenges and Changes

As Figure 1 details, 5G deployment is in its infancy. Current roll-outs utilise the 5G non-standalone architecture (NSA) which connects 5G radios to a LTE core and this is set to continue as deployments of 5G New Radio (NR) roll-out with functionally separated backhaul and fronthaul interfaces. In such NSA deployments services and security is still handled by a LTE core, so users and applications with 5G enabled devices cannot take advantage of the security enhancements with 5G. Later, the full potential of 5G will be realized with deployments of fully-distributed, networking slicing-enabled networks – often described as 5G standalone (5G SA) that feature the decoupled signalling control plane of the 5G next-generation core. It is the 5G core and

Figure 1: 5G deployment status in 2025 (source: GSMA)



service-based architecture that places huge new pressure on security because capabilities such as network slicing and exposed APIs are new threats to be addressed.

Network functions security

The service-based architecture of 5G specifies flat, peer to peer, relationships between network functions via the HTTP/2 service-based interface. Attackers are highly familiar with HTTP/2 and its vulnerabilities, so new 5G security requirements that are designed to ensure that network functions only expose themselves to one another securely are needed. As well as embracing the same requirements as LTE for network domain security using IPsec, the 3GPP standards body has mandated that 5G SA network functions must also support Transport Layer Security (TLS) encryption to secure the information exchange between functions. In addition to this stronger encryption of communications between network functions, 3GPP has specified mature, well-established, authentication and authorisation standards between functions. These are meant to ensure that functions and individuals within the 5G core can only access resources they are authorised to have access to.

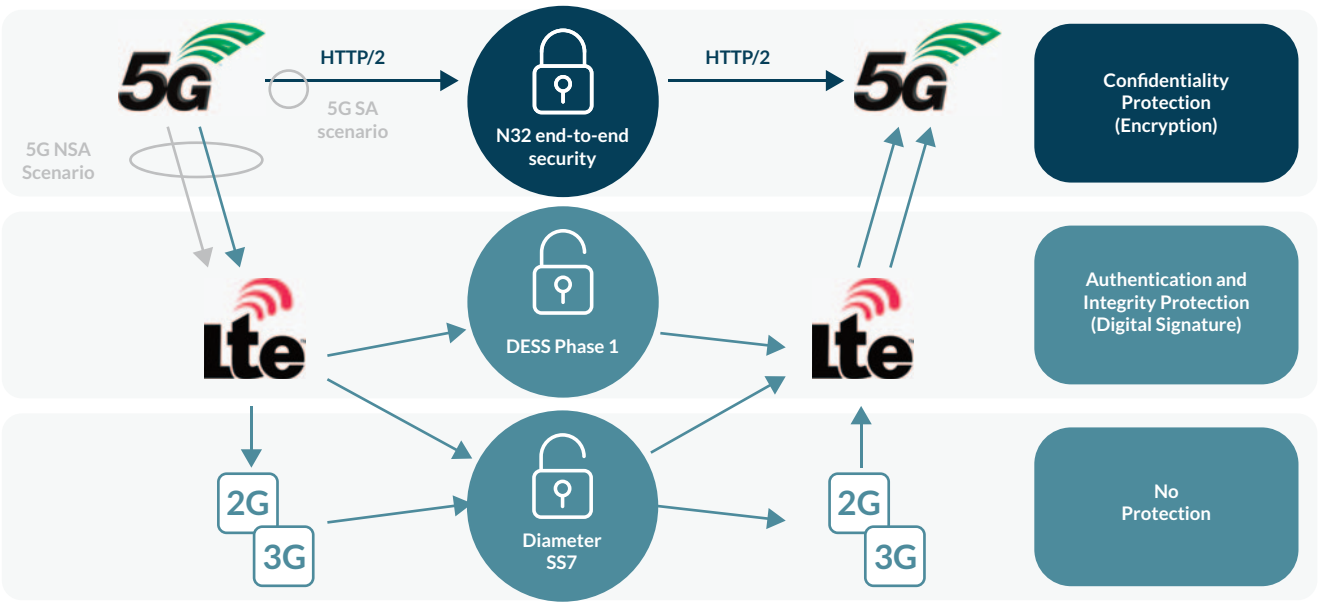
Network slice security

A key part of the value of the 5G core is the capability to spin up and maintain customised network slices for different use cases, support them across transport, the radio access network (RAN) and the device, assuring quality and ensuring end-to-end security across those domains. The 3GPP specification addresses the required security layers for network slicing but these will also need to be augmented according to the specific needs of the different use cases. Beginning with initial slice deployment, 3GPP specifies mutual authentication between the network slice manager and the cloud that the slice is being deployed from. In addition, policies are needed to assure effective isolation of physical and logical networks from one another to ensure threats can't spread between slices.

Exposed network API security

In the 5G core, the interfaces between network functions are APIs rather than traditional network communication methods. Generic IP communications functions and mechanisms such as simple network management protocol (SNMP) or secure shell (SSH) will not be included in the management and orchestration. The API, enabled by the network exposure function (NEF), is the interface between the 5G core and external third-party applications and exposure will therefore be tightly controlled. Operators have experience of this issue having suffered from API vulnerability in the past so are prioritising securing API exposure.

Figure 2: Mobile Network Technologies and Signalling Protocols



Secure inter-operator signalling: from SS7/Diameter firewalls to SEPP

The risks of leaving SS7 signalling messages between 2G and 3G networks unprotected has been known about for many years. Mobile operators have retrospectively specified and deployed SS7 firewalls as well as Diameter firewalls for LTE. 5G provides a means to go further and 3GPP has already specified the Security Edge Protection Proxy (SEPP). This protects HTTP/2 control plane messages between one mobile operator's 5G core and another's, similar to how SS7 and Diameter firewalls protect the roaming interconnections in 3G and LTE. SEPP is already specified and available to operators in advance of 5G core deployment so is an example of security-by-design for 5G and provides a contrast to the retrospective introduction of firewalls in the past. However, operators' existing state of SS7 signalling equipment, plus continuing high volumes of SS7 gateway sales, suggests this migration will, again, take substantial time.

The signalling control planes of 2G/3G and LTE networks have proven themselves robust but 5G increases the complexity and threat level on the control plane and will therefore demand new approaches to meet these more complex security requirements. In previous generations, inter-operator signalling security was difficult to achieve because of the early telephony signalling legacy and retrofitting of firewalls will not suffice in 5G.

Figure 2 sketches the protection applicable in various roaming situations. Ideally 5G users are protected with the advanced 5G security enhancements in a pure 5G SA scenario between 5G Core networks with end-to-end HTTP/2 support. In all other roaming scenarios the protection of 5G users will fall back to LTE or 2G/3G security. This will be applicable, in most cases, for many years and will involve within such situations a transparent - not encrypted - exchange of signalling between mobile networks. Significantly, SS7 does not by default support in-built protection so it should be emphasised that 5G users need protection in the form of signalling firewalls for SS7 and Diameter.

Secure user identity: SUPI and SUCI

The user plane in 5G is better protected because it extends the air interface and integrity encryption of LTE with a new, 5G equivalent of the international mobile subscriber identity (IMSI), the subscriber permanent identifier (SUPI) encrypted and sent over the air as a one-time, temporary identifier, called a subscriber concealed identifier (SUCI). This prevents man-in-the-middle or IMSI-catcher attacks that collect data from devices. Another advantage is 5G's new user authentication framework, which allows the 5G core to serve access requests from Wi-Fi and wireline devices as well as from 5G devices.

Roaming: end-to-end encryption

Network operators have seen that the 5G roaming solutions set out in 3GPP Releases 15 and 16 are far more complex to operate than 2G/3G and LTE roaming and do not necessarily address the existing vulnerabilities with 2G/3G/LTE roaming. Therefore, further work is needed to enhance the current deployment options, which are oriented around SEPP. These include solutions based on transport layer security (TLS) which has all signalling encrypted end-to-end in TLS tunnels between mobile roaming partners. The direct TLS approach nicely works for traditional operator-to-operator roaming relationships among the largest bi-directional operators but does not address the needs of all players in the global roaming ecosystem of approximately 800 mobile network operators.

An alternative approach is the PRotocol for N32 INterconnect Security (PRINS), which is an application layer security solution that involves part of the signalling information being sent in the clear, to enable roaming between VAS operators and hubs, and so that IPX providers can inspect or modify signalling traffic. However, PRINS is not compatible with the direct TLS model and very complex to operate with the need for policies and certificate settlements between about 800 players in the global roaming ecosystem.

As a consequence, mobile operators and IPX operators are intensively collaborating in the GSMA (the global association where mobile operators arrange their roaming connections) on a single, simple and secure solution for 5G roaming with general support in the industry. This joint effort is to avoid what happened with LTE roaming where from the start two different solutions were developed and implemented, and eventually only one solution appeared deployable between real networks. This dual policy resulted in a waste of investments and contributed to a delayed rollout of VoLTE roaming.

Secure caller identity: STIR/SHAKEN

Although not a problem specific to mobile services only, identification is a vital element of security and increasingly prominent in network operators' minds. The slew of robocalling in the US, which saw more than 48 billion robocalls in 2018, has resulted in new legislation, called The Telephone Robocall Abuse Criminal Enforcement and Deterrence Act (TRACED). This could see robocallers fined up to US$10,000 per call by the Federal Communications Commission (FCC). The component of the legislation that requires carriers to

verify caller identities is of greater relevance here and is seeing US network operators roll-out a caller verification service based on the SHAKEN (Secure Handling of Asserted information using toKENs) and STIR (Secure Telephony Identity Revisited) protocols. This security-by-design solution enables a cryptographic method that uses public and private keys so the caller ID can be verified. The high level of encryption means the caller ID info can't be tampered with from end-to-end, thereby enabling the calling number to be verified to the consumer.

The aim behind this approach is to get users to trust voice calls again, but carriers can't act in isolation because caller verification needs to happen for calls that originate and terminate in different carriers' networks. An independent policy administrator is therefore required to ensure only legitimate and trusted entities can participate in the STIR/SHAKEN Caller ID ecosystem. This capability will also feed into the inter-carrier trust mechanisms that operators need to enable roaming.

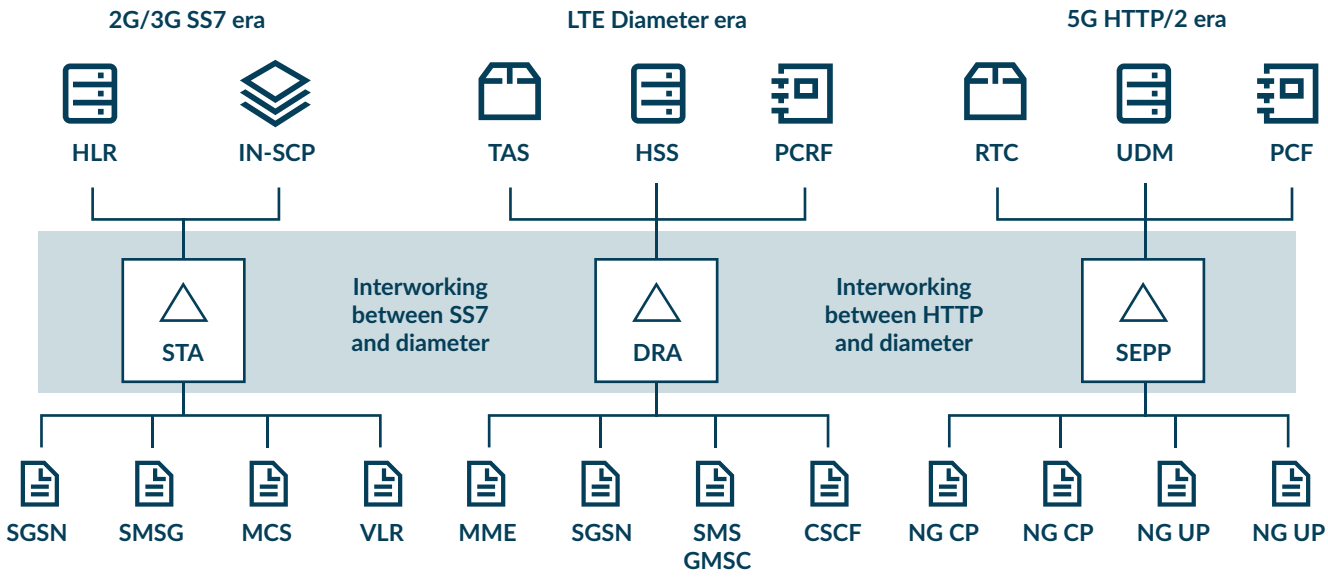
Certifying that trust is a critical challenge and the circle of trust between network operators, service providers, enterprises and customers can remain unbroken in the new ecosystem. However, the STIR/SHAKEN ecosystem is complex and costly to implement and both within and outside of the US the gulf between the current state and TRACED compliance, which is mandated in the US by 30 June 2021, is wide. Important issues such as key management and the political, operational, financial and technical concerns need to be addressed comprehensively to enable international connections and compliance with international regulations as they emerge.

InterGENerational Security

As described above the introduction of 5G is neither rapid nor independent of previous generations of mobile technology. It's well understood that 5G is more of a concept than a finished product and the telecoms industry is fully aware that mobile technology takes decades to evolve because of different rates of adoption, issues with device availability and variables such as security adoption across generations. Security will therefore need to extend the same functionality not only across multiple operators in the 5G environment but also across 2G/3G and LTE. This interworking is complex to establish and to operate because of the substantial differences in architecture between 2G/3G/LTE and cloud-based world of 5G. This is the at the very heart of the 5G security challenge: It's not just 5G networks that need to be secured, it's the entire end-to-end interaction needing an integrated holistic signalling security solution embracing 2G, 3G and LTE (4G) with 5G.

Figure 3 outlines the overall signaling perspective ranging from the legacy 2G/3G SS7 era, the LTE Diameter era and the 5G HTTP/2 era.

Figure 3: InterGENerational signalling coexistence (source: GSMA)



Conclusion

The introduction of 5G provides an opportunity to include security-by-design into specifications in a way that was omitted in previous generations and led to greater complexity and greater workloads to achieve security retrospectively. 5G, however, won't be an idealised landscape even with the adoption of Security-by-Design principles. Nevertheless, these principles are set to introduce a lexicon of concepts that will provide the framework for a radically more secure environment across the 5G landscape – which also includes continued utilisation of previous network generations.

The fundamentals in the Security-by-Design framework are utilisation of mutual authentication so both parties to a session or communication can establish trust and a secure end-to-end relationship. Within this, there are assumptions such as that the network is presumed to be open with no assumed safety of overlaid products or processes and an acknowledgement that all links could be compromised by criminals. This acknowledged insecurity mandates the encryption of both intra and inter networks traffic in a way that ensures encrypted information is worthless if intercepted. This is a substantial development and goes far beyond existing cellular network security practices and delivers improved confidentiality and integrity of user and device data under all circumstances. To succeed, network operators must sustain commitment to seamless interworking with the existing control plane and user plane security measures implemented in 2G, 3G, and LTE networks designed to secure both home network and roaming services.

NetNumber has been instrumental in delivering industry-leading security solutions for the telecoms industry for twenty years, while providing support and continued innovation for legacy technology.

NetNumber security solutions provide comprehensive interGENerational security and fraud protection with real-time threat detection, delivering internetworking encryption and the elimination of attack vectors within the network. Its Signalling Firewall protects against malicious attacks on inbound traffic and data.

Traditional and legacy solutions have been delivered on NetNumber TITAN, a robust centralised signalling and routing control (CSRC) platform, that offers a common, virtualised infrastructure for all signalling control, routing policy enforcement and subscriber database services in the network.

The interGENerational and cloud-native platform TITAN.IUM, is a multi-generation, multi-protocol CSRC. TITAN.IUM aligns to CSPs' network and service evolutions from 2G and 3G, through to 4G. It is also the home for NetNumber's 5G applications and aligned to CSPs' journeys to becoming cloud-native.

NetNumber's Guaranteed Caller™ is a family of standards-compliant STIR/SHAKEN solutions that address all of the common trusted call scenarios, so that legitimate callers can participate in the STIR/SHAKEN trust network, while fraudsters are locked out. Guaranteed Caller was built with one overriding goal in mind – simplifying the journey to STIR/SHAKEN compliance for service providers.

NetNumber maintains its position as a leader in network security though active participation and contribution in a number of industry bodies and workgroups including the Fraud and Security Group (FASG) of the GSMA.

Find out more at www.netnumber.com

Mobile World Live is the premier destination for news, insight and intelligence for the global mobile industry. Armed with a dedicated team of experienced reporters from around the world, we are the industry's most trusted media outlet for breaking news, special features, investigative reporting, and expert analysis of today's biggest stories.

We are firmly committed to delivering accurate, quality journalism to our readers through news articles, video broadcasts, live and digital events, and more. Our engaged audience of mobile, tech and telecom professionals, including C-suite executives, business decision makers and influencers depend on the unrivalled content and analysis Mobile World Live provides to make informed business decisions every day.

Since 2016, Mobile World Live has also had a team of in-house media and marketing experts who work directly with our brand partners to produce bespoke content and deliver it to our audience in strategic yet innovative ways. Our portfolio of custom work - including whitepapers, webinars, live studio interviews, case studies, industry surveys and more – leverage the same level of industry knowledge and perspective that propels our newsroom.

Mobile World Live is published by, but editorially independent from, the GSMA, producing Show Daily publications for all GSMA events and Mobile World Live TV – the award-winning broadcast service of Mobile World Congress and home to GSMA event keynote presentations.

Find out more at www.mobileworldlive.com

Disclaimer: The views and opinions expressed in this whitepaper are those of the authors and do not necessarily reflect the official policy or position of the GSMA or its subsidiaries.

© 2021