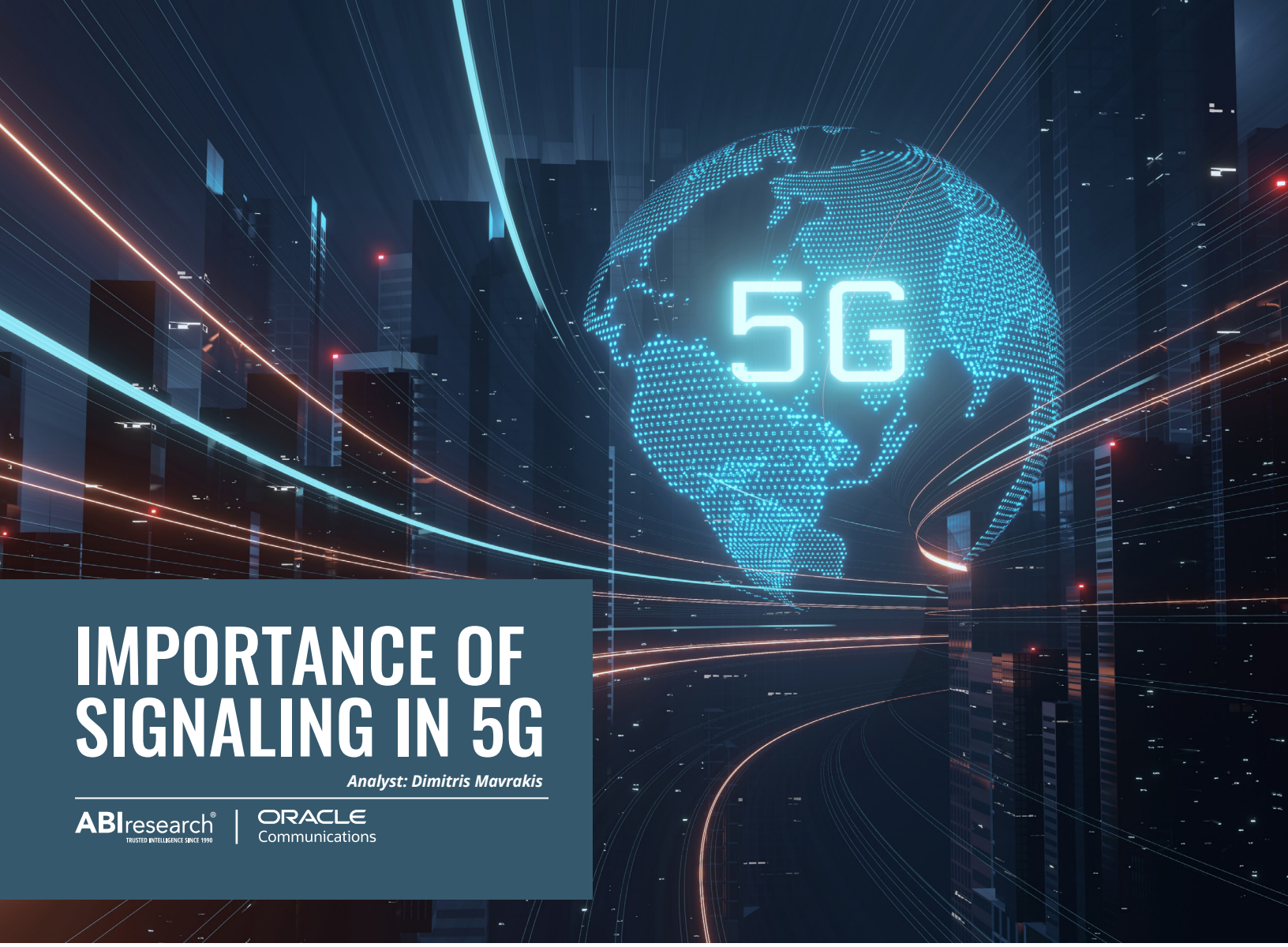


IMPORTANCE OF SIGNALING IN 5G

Analyst: Dimitris Mavrakis

ABIresearch® TRUSTED INTELLIGENCE SINCE 1990 | ORACLE Communications

TABLE OF CONTENTS

INTRODUCTION AND MARKET OVERVIEW

The 5G market is currently growing rapidly with mobile operators deploying networks aggressively in North America, Western Europe, and Asia-Pacific. Early adopters in South Korea have already completed the first phase of their 5G network deployments, while operators in China are deploying 5G New Radio (NR) base stations by the thousands every week. In the meantime, there were hundreds of 5G devices in the market at the end of 2020, signaling that 5G has reached a critical mass. At the same time, several operators are also now deploying 5G in Standalone (SA) mode, where the NR is no longer tethered to a Long Term Evolution (LTE) Core Network (CN), but a brand-new CN designed for advanced services, including network slicing, microservices, edge computing, and high levels of automation. The beginning of 2021 will truly see 5G reaching the mass market and slowly starting to become the key cellular communication technology.

It has not been a smooth introduction for 5G though. Several operators are coming to realize that 5G cannot become successful with additional fees over 4G, meaning that this new generation will not result in new revenue in the consumer market, which is already most familiar with Mobile Broadband (MBB) connectivity. In a way, 5G is not introducing anything new in terms of user experiences in its infancy, but the higher speeds it offers will surely create a new stream of consumer applications. Moreover, 5G infrastructure is costly and the first-generation equipment that is being deployed in 2020 consumes more energy than previous generations, leading to some operators switching these new 5G base stations off at night to save on power costs.

Despite these early challenges, 5G shows great promise for enterprise use cases. There are already several trials and early deployments of the new cellular generation in the manufacturing, energy, transportation, and mining sectors. 5G can introduce carrier-grade networking for mission-critical applications and introduce tangible benefits for enterprises, especially when mobility is a necessity for automated use cases. The market is still expanding for enterprise use cases, but 5G is a platform that many enterprise verticals can use to accelerate their digital transformation.

Mobile Network Operators (MNOs) are currently deploying their networks with these future prospects in mind. Early 5G radio networks were tethered to a 4G core, but many operators are now deploying in SA mode, deploying a new CN in parallel with the NR network. This CN will enable future use cases and new applications, and it has now become important to deploy these networks with the future in mind.

5G CORE NETWORK DEPLOYMENTS

The arrival of 5G, cloud technologies, and virtualization architectures has necessitated revisiting the overall system architecture of both the Radio-Access Network (RAN) and the CN. The 5G CN is designed from the onset to natively introduce three new significant advancements: a Service-Based Architecture (SBA), native support for network slicing, and control plane/user plane split. All of these represent significant improvements over previous CN deployments, with many existing functions renamed and either split or merged depending on the functions that fall within the user or control plane. Some select key changes are as follows:

- **User Plane**: The user plane consists of the User Plane Function (UPF) that is responsible for packet routing, packet forwarding, and Quality of Service (QoS), among many other functionalities.
- **Control Plane**: This consists of several parts, namely the Access Management Function (AMF) and the Session Management Function (SMF). The AMF oversees authentication, connection, and mobility management between the network and the device. The SMF handles session management, Internet Protocol (IP) address allocation, and control of policy enforcement.
- **Other Functions**: The 5G CN also includes the Authentication Server Function (AUSF) responsible for authentication functionality, the Unified Data Management (UDM) responsible for authentication of subscriber data, and the Policy Control Function (PCF) responsible for QoS policies. The Network Exposure Function (NEF) and NR Repository Function (NRF) handle exposure events and service delivery capabilities, respectively. A full list of these elements is presented in Figure 1.

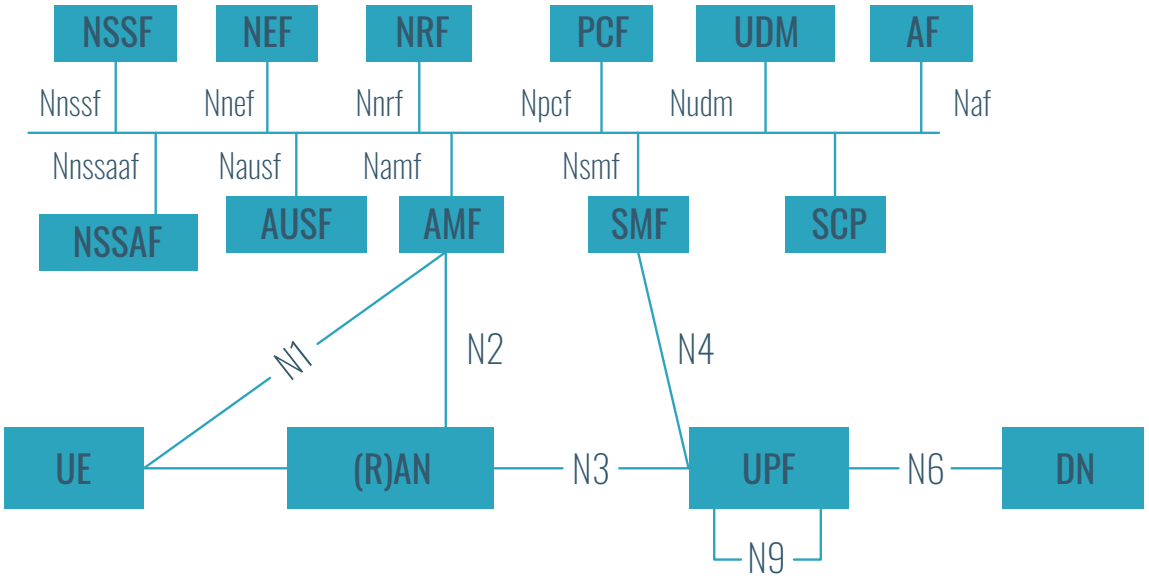
THE SERVICE-BASED ARCHITECTURE

The underlying pillar of the 5G core is the SBA architecture, a fine-grained arrangement of loosely coupled and autonomous components intended to fully use the efficiency and flexibility of virtualization and cloud technologies. In SBA architecture, services interact with other services in a lightweight fashion across Application Programming Interfaces (APIs). The underlying premise of an SBA is service atomicity, which is the ability for services to run in isolation from concurrent processes. By extension, an SBA is conducive to a 5G network that is reduced to being more granular and decoupled, all of which is associated with the following benefits:

- **Introduce IT-Oriented Nimbleness and Scale**: With SBA as a predominant architecture in 5G deployments, network services of the 5G core can efficiently be exposed to third-party providers, which paves the way for innovation and new business models (*e.g.*, Business-to-Business-to-Consumer (B2B2C)).
- **Incorporate, Elastic, and Modular Components**: SBA's elastic nature promotes accessible and recomposable applications and services that can grow and shrink on demand. This bodes well for an increasing demand for consumption-based commercial models.
- **Open and Disaggregated Network**: A service-centric delivery calls for data, functionality, and API exposure to external entities (*e.g.*, enterprise back ends) without a cumbersome protocol/process conversion. SBA is a sound step in that direction.
- **Open Source and Cloud-Native Technologies**: An SBA 5G network relies on open-source technologies and cloud-native architectures, which introduces cloud computing benefits in the telco domain. Almost all mobile operators have now understood that cloud-native architectures, containers, and micro-services are founding elements for their Next-Generation (NG) networks and services.

Figure 1: 5G System Architecture

(Source: 3GPP TS23.501)



BUS COMMUNICATION MODEL

The point-to-point control plane communication model that was being used in 4G CNs is now replaced by a "service bus" that connects the individual Network Functions (NFs). This is a control plane bus that will allow NFs to exchange information through common REST APIs and exchange signaling messages through the HTTP/2 protocol. NFs that are connected to the service bus can either access the services (consumers) that other NFs offer (producers). NFs are now classified in these two groups: consumers and producers that connect through APIs in a consistent interface bus. The NRF holds information and the list of individual network functions so that a new NF can discover the rest of the network and access information accordingly.

CONTROL AND USER PLANE SEPARATION

Control and User Plane Separation (CUPS) was first introduced as part of the 3rd Generation Partnership Project (3GPP) Release 14 standard and it is a key pillar for 5G CNs. The utility of CUPS lies in the separation between the control and user plane, where the latter can be scaled independently of the former. The control plane can be hosted in a centralized function, whereas the use plane can be placed close to edge locations. This arrangement avoids the need to backhaul traffic to a central location, thereby improving latency and bandwidth requirements for applications like video, connected cars, and other enterprise vertical use cases. CUPS enables the following benefits:

- **Deploy a Distributed Architecture**: CUPS aids deployment of 5G core elements across both public and private clouds, in turn paving the way for economic gains. At peak times, traffic can be scaled to the public cloud. Off-peak Mobile Service Providers (MSPs) can scale back to the private cloud.
- **Promote Cloud Platform Compatibility**: VNFs for management, data, service, and latency-insensitive can now be hosted in multiple cloud platforms, such as Amazon Web Services (AWS), Azure, third-party OpenStack, a vendor's own cloud, or a combination thereof.
- **Centralized and Simplified Operations and Management (O&M)**: CUPS is conducive of a centralized O&M configuration in which one control plane resource pool configures and controls several user plane pools, in turn simplifying network configuration.

CUPS and many other 5G features will create the opportunity for more applications, while lowering the burden for the edge of the network. For example, a future feature of 5G is Device-to-Device (D2D) communication that allows content to be shared between devices directly, without traffic entering the core or even the edge network. However, these devices will need to be tethered to the CN to establish communication between them, meaning that the signaling network will likely be burdened.

In many cases, 5G CN infrastructure becomes critical to the carrier-grade operation of the network, especially when enterprises and new use cases are considered.

SIGNALING INFRASTRUCTURE

Signaling is a critical component of any cellular network and a topic of debate since the first digital network was deployed. A short history on signaling infrastructure follows:

- Signaling System No. 7 (SS7): This is a system of protocols designed in 1975 that are used predominantly to establish and tear down circuit-switched voice calls. SS7 is the established signaling protocol for 2G and 3G, and several networks that use it are still online today, especially in developing markets. The protocol can also perform additional functions, such as number portability, Short Messaging Service (SMS), and prepaid services.

  Although SS7 has been widely deployed, it does not include authentication or access control, because at the time of its inception, a handful of telco providers were present in the market, which trusted each other. Since then, the deregulation of the market has resulted in an increasing number of threat vectors, targeting SS7 for malicious intents.

  SS7 was also positioned toward establishing
  circuit-switched voice calls, rather than data-oriented services, and certainly not packet-based voice calls (Voice over LTE (VoLTE)). As such, it was replaced in LTE with Diameter.

- **Diameter**: Diameter was based on the Radius protocol and provides authorization, authentication, and accounting functionality in Information Technology (IT) networks. It is the predominant signaling protocol in 4G LTE networks and widely deployed throughout the world. Diameter is also the foundational protocol in IP Multimedia Subsystem (IMS) deployments and is used to enable VoLTE in 4G, as well as many advanced features, including online charging, interfacing with roaming partners, and more. Unlike SS7, Diameter is inspired by Internet Engineering Task Force (IETF) and IP networks, adding security in its core design. For example, Diameter traffic must either be transported with Transport Layer Security (TLS) or Internet Protocol Security (IPsec) tunnels.

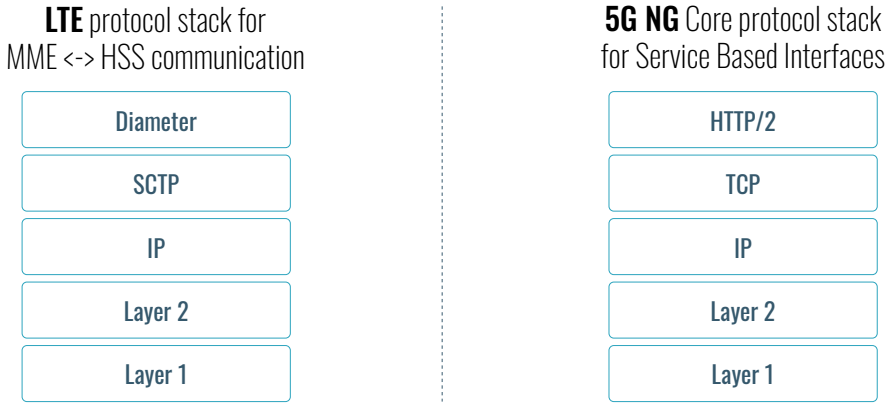
Regardless of the benefits of Diameter, 5G NG CNs have been designed with an SBA in mind, so they require web interfaces that include APIs for service exposure and communication between network elements. 3GPP's standards group has decided to adopt HTTP/2 as the main signaling protocol.

HTTP/2 SIGNALING IN 5G CNS

As discussed in Section 2.1, the 5G SBA introduces the consumer-producer interaction model that can theoretically scale for several micro-services and containerized network elements. The 4G and 5G protocol stacks are depicted in Figure 2.

Figure 2: LTE and 5G Core Network Signaling Protocol Stacks

(Source: ABI Research)



The 3GPP group has decided to use the HTTP/2 protocol for signaling, taking advantage of the maturity of the protocol and the good understanding the Internet community has for it. Moreover, HTTP/2 means that vendors developing signaling products will be able to develop products much faster than previous generations, when new protocols were chosen. Nevertheless, the choice of HTTP/2 does introduce new challenges.

5G NETWORK SIGNALING CHALLENGES

As with any new or even seasoned signaling infrastructure, the 5G SBA network is subject to new and existing challenges, some of which are described in Table 1.

Table 1: Challenges Facing 5G SBA Signaling Infrastructure

(Source: ABI Research)

NETWORK OR PROTOCOL PROPERTY	SIGNALING CHALLENGE
Choice of HTTP/2 Protocol	The HTTP protocol is well established and prone to malicious attacks. Previous signaling protocols were not as widely deployed as HTTP, potentially opening new threat vectors for the 5G control plane.
5G Services	The 5G network is designed to enable a plethora of enterprise use cases that will vary among high bandwidth, low number of devices to low bandwidth, and extreme numbers of devices. The signaling infrastructure will likely face increasing pressure when enterprise 5G use cases are deployed.
CUPS	CUPS will potentially ease the burden in the user plane, which may create an influx of new types of services and applications. These will likely use the NG core control plane, leading the new levels of congestion.
Signaling Storms and Overload	5G CNs will be subject to the very same challenges previous networks have faced, especially given the producer/consumer interface model and the increased number of network elements due to microservices.
Backwards Compatibility	NG CNs will still have to interface with previous signaling protocols going back to SS7 for inbound roaming. 5G signaling networks need to be designed appropriately from the start.

Although NG core is indeed introducing a new signaling protocol, it is still subject to the very same challenges previous signaling infrastructure faced, and adding new challenges due to the popularity of HTTP/2 and the architecture of the SBA. The challenges facing this new signaling infrastructure are now two-fold:

- Intrinsic challenges introduced by the selection of the HTTP/2 protocol itself. For example, there are several HTTP/2 attack vectors that led to Denial of Service (DoS) conditions, including data dribble, ping flood, resource loop, reset flood, empty frames flood, and many others. If a malicious actor can access an operator's signaling infrastructure, then there are established methods and tools that can cause harm.
- The HTTP/2 framework is still subject to the very same challenges previous network generations faced with SS7 and Diameter, including routing and optimization, traffic management, robustness and scalability, network visibility, core security, and authorization and authentication. All of these will still create implementation challenges that all operator Chief Technical Officers (CTOs) will need to consider before scaling up their 5G CN deployments.

Adding the fact that this new signaling infrastructure needs to interface with previous generations and translate between protocols means that signaling challenges in 5G may be considerably more than for previous generations.

NETWORK REPOSITORY FUNCTION

3GPP Release 15 introduces the NRF that is responsible for maintaining a record of all network elements operating in the 5G network, as well as the services these elements can provide. The NRF is the key element to enabling the instantiation, scale-in, scale-out, and termination of network services, as well as to enabling element and service discovery. Through the NRF, new network components can discover the rest of the network and get an updated status of said network elements. In short, the NRF provides the following functionality:

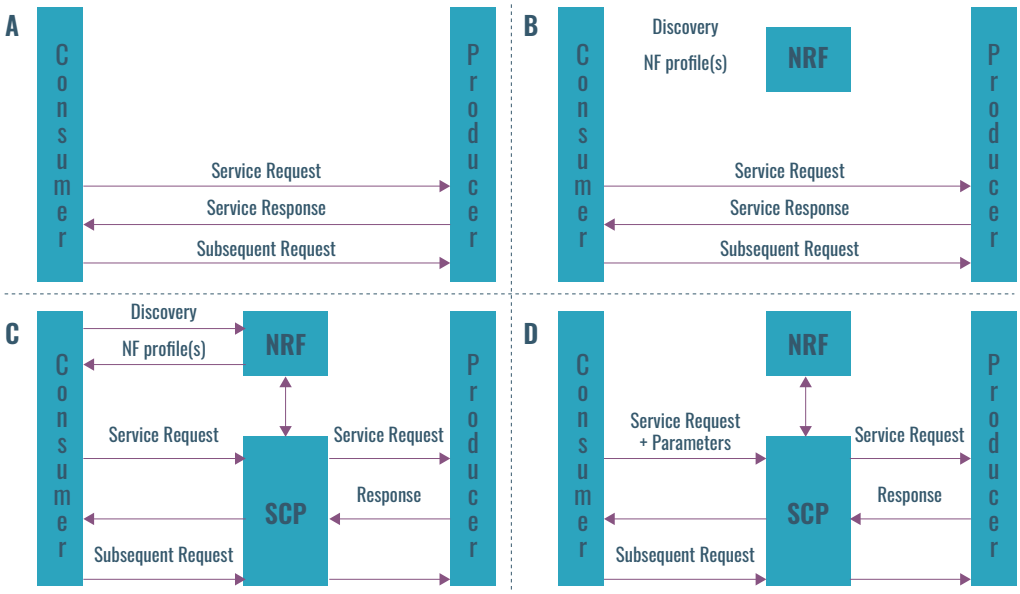
- Maintains the profiles of the available NF instances and their supported services in the 5G CN
- Allows consumer NF instances to discover other providers' NF instances in the 5G CN
- Allows NF instances to track the status of other NF instances

A challenge that the NRF could face in advanced 5G deployments is that control plane scalability, security, and granular management will likely create a bottleneck. The NRF is responsible for service discovery and maintaining a record of available capabilities, rather than managing the control plane. Therefore, a new component was introduced in Release 16.

SERVICE COMMUNICATION PROXY

The Service Communication Proxy (SCP) was introduced in 3GPP Release 16 and is a vital network element if NG CNs reach a large scale, and edge computing, network slicing, and other services are considered for the future. Figure 3 illustrates signaling deployment types with and without an NRF and an SCP.

Figure 3: Different Types of NG Core Signaling Infrastructure Deployments
(Source: 3GPP TS 23.501 V16.6.0)



The SCP includes one or more of the following functionalities:

- Indirect communication between NFs
- Delegated discovery
- Message forwarding and routing between NFs
- Message forwarding and routing to a next hop SCP
- Communication security, load balancing, monitoring, overload control, *etc*.

Although an NG CN may be deployed without an SCP, it is very likely that every single 5G CN will require one once large-scale deployments are achieved. Moreover, enterprise use cases will likely necessitate the use of an SCP even more. An SCP may also address signaling challenges that took place in the past and are agnostic of the underlying network generation. Thus, it is vital to learn from past challenges and plan ahead accordingly.

LEARNING FROM THE PAST

Previous network generations, especially 4G, illustrated that signaling infrastructure should not be ignored. Indeed, Diameter used in 4G eventually necessitated the deployment of Diameter Routing Agents (DRAs), Diameter Edge Agents (DEAs) and Diameter Signaling Controllers (DSCs) to perform control plane message routing, interface with roaming networks, and perform Inter-Working Functions (IWFs). These network elements were not specified in the official 3GPP specification and had to be tested and deployed in the market. Nevertheless, practical deployments of 4G illustrated that these were indeed vital components to secure these networks, to ensure that "signaling storms" did not happen and that signaling messages were delivered consistently.

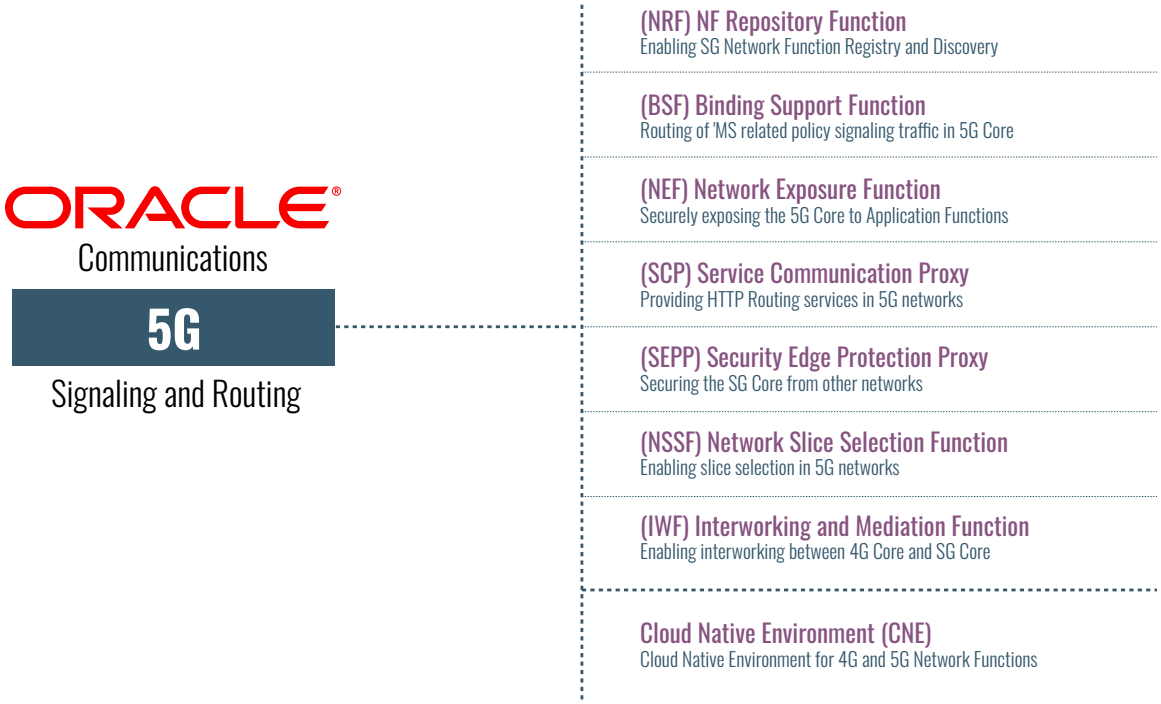
Very similar discussions took place in 3GPP between Release 15 and 16 that argued the necessity of an SCP. In theory, the NRF is the only network element needed for service discovery and for NFs to communicate, but in practice, the NRF may quickly become congested with signaling messages and may also not be able to handle security challenges. Therefore, the SCP was introduced in R16 and may now be deployed in a centralized or distributed manner. The most important priority for any operator that deploys an NG core will be to plan for use cases that have not been discussed yet. An SCP will likely be a vital component for the future of 5G networks.

DEPLOYING 5G SIGNALING INFRASTRUCTURE (BASED ON INPUT FROM ORACLE)

Oracle Communications provides the complete set of microservices-based cloud-native NFs to help operators design a reliable signaling and routing strategy to enable elastic growth, interoperability, and rapid introduction of new services in its network.

Figure 4: Oracle Communications Signaling and Routing

(Source: Oracle Communications)



Oracle Communications' signaling and routing portfolio is designed as microservices and is based on cloud-native principles. It integrates seamlessly with Oracle Communications' Automated Test Tool and Script for automating the testing of 5G NFs. Oracle Communications' NFs also support Oracle Communications' Cloud Native Core Console (CNCC), a Graphical User Interface (GUI) for configuring 5G NFs. The signaling and routing NFs are designed on the principle of best-in-class technology with laser focus on quality and security parameters. The NFs can be readily deployed in a customer's environment, support logging and tracing, and integrate easily with DevOps workflows and operators' Continuous Integration (CI)/Continuous Deliver (CD) pipelines.

Table 3: Oracle Communications Signaling and Routing Features

(Source: Oracle)

NF	FEATURES
Binding Support Function (BSF)	Provides secure interaction of application functions with the policy framework. Comes with embedded database for storing binding information. Basic Key Performance Indicator (KPI) support from BSF services for registration/deregistration/ discovery.
Interworking and Mediation Function (IWF)	Protocol translation feature of IWF enables interworking between Evolved Packet Core (EPC) network elements and 5G core NFs by performing mapping of HTTP/2 to Diameter messages and *vice versa*. Mediation framework is a rules-based engine that alleviates interoperability issues through easy creation of mediation rules *via* IWF Mediation Wizard and allows for HTTP/2 extensions

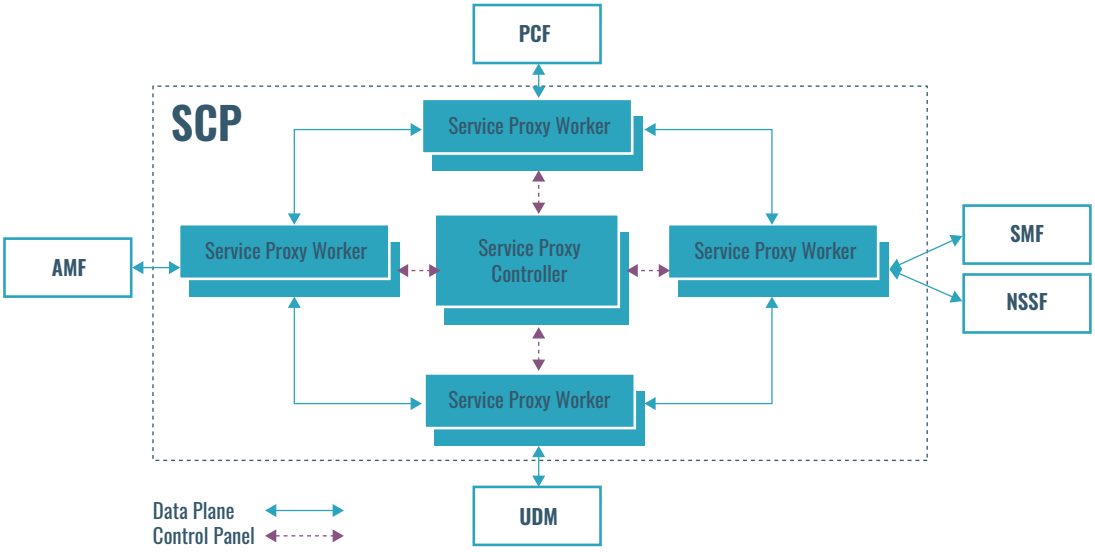
NF	FEATURES
Network Exposure Function (NEF)	The NEF acts as a centralized point for service exposure and plays a key role in authorizing all access requests originating from outside 3GPP networks to enable cellular Internet of Things (IoT), non-IoT, edge computing, and API gateway use cases for operators. Enables monitoring event procedure by allowing IoT AF to monitor the device state location information, roaming status, IoT device reachability, connectivity status, etc.
Cloud Native Environment (CNE)	NF CNE is a collection of services that provides a common platform for CSP applications' operation and life cycle management. Supports industry-leading open-source components for logging, monitoring, and tracing collection, storage, and visualization. This allows the NF CNE platform to leverage the innovation and velocity of third-party elements and developments across service providers and IT.
Network Repository Function (NRF)	Supports NF management, discovery, and authorization services. Can be combined with Oracle SCP to enable more optimized NF and be implemented as standalone NRF.
Network Slice Selection Function (NSSF)	Provides slice selection capabilities by determining authorized slice information based on Network Slice Selection Assistance Information (NSSAIs) provided by the Access and Mobility Management Function (AMF) as part of registration procedure. Enhanced slice selection policy: This feature enables the NSSF to be intelligently select the Network Slice Instance ID (NSI-ID) based on Public Land Mobile Network (PLMN), Tracking Area (TA), time of day, day of week.
Service Communication Proxy (SCP)	Reduces connections to/from NFs by being deployed alongside NF instances and acting as an outbound proxy for the NF instances. The SCP eliminates the need for 5G NFs to set up direct connections between each other. Improves load balancing: The SCP has a complete view of all the messages arriving for a given NF type. Improves routing control and resiliency: The SCP provides enhanced routing control based on routing rules created using NF notifications received from NRFs. Provides congestion control: The SCP has the ability to reprioritize traffic, as well as protect the network from flooding by malicious or rogue consumer NFs, and at the same time protect provider NFs from being overloaded. Supports canary testing: The SCP plays a crucial role in the rollout of new NF releases.
Security Edge Protection Proxy (SEPP)	SEPP provide authentication, confidentiality, and integration protection for inter-PLMN Service-Based Interfaces (SBIs) signaling traffic between 5GC NFs. Supports robust congestion control features to protect against signaling storms.

Oracle Communications' Cloud Native Core SCP is a decentralized solution that provides signaling control to a 5G CN. It is composed of Service Proxy Controllers and Service Proxy Workers and is deployed alongside other 5G network functions. The SCP provides routing control mechanism by creating traffic routing rules based on interactions with the Network Repository Function (NRF). The Oracle Communications SCP also provides resiliency and observability to the 5G CN, while enabling elastic growth, interoperability, and rapid introduction of new services. This allows CSPs to more effectively and efficiently operate their 5G networks, while reducing complexity and maintenance costs.

Oracle Communications SCP is modeled after the cloud-native service mesh solution and is made up of a control plane and data plane that scale independently. The control plane is used to transfer routing rules from the controller to the worker, while the data plane is used to transport 5G SBA messages.

Figure 5: Oracle Communication Cloud Native Core, SCP System Architecture

(Source: Oracle Communications)



Oracle Communications' SCP not only resolves the challenges introduced by the 5G SBA, but also optimizes signaling controls. It provides service providers with better visibility into the CN by:

- **Reducing Connections to/from NFs**: By being deployed alongside NF instances and acting as an outbound proxy for the NF instances, the SCP eliminates the need for 5G NFs to set up direct connections between each other. Connections can be optimized so that each NF instance maintains a set of redundant connections to the SCP and uses those connections for all outbound requests.
- **Improving Load Balancing**: The SCP has a complete view of all the messages arriving for a given NF type. It supports schemes like round robin and weighted round robin, and factors in current load and NF availability to improve load balancing.
- **Improving Routing Control and Resiliency**: The SCP provides enhanced routing control based on routing rules created using NF notifications received from the NRF. The SCP boosts resiliency in 5G networks by providing features like alternate routing, outlier detection, and circuit breaking. It relieves consumer NFs from remembering and interpreting complex routing rules associated with next hop selection and, at the same time, makes re-routing decisions based on load conditions and the health status of NF providers.
- **Acting as a Circuit Breaker**: In the absence of an alternate route, the SCP will quickly reject requests destined to become a failed or degraded NF, thereby acting as a circuit breaker. This prevents valuable resources at the consumer's NFs from being tied up waiting for responses from providers. The SCP also performs retries on behalf of the service consumer, thereby relieving the service consumer from this burden and leaving it to focus on the application.
- **Supporting Metrics, KPIs, and Reports**: As services requests are proxied *via* the SCP, the SCP collects metrics and KPI related to message processing, such as request and response counts, messages/sec or average transaction latency, etc. With this information, the SCP is in a unique position to provide a view of the network health indicators at any given time.
- **Providing Congestion Control**: The SCP has the abilities to reprioritize traffic and protect the network from flooding by malicious or rogue consumer NFs, and at the same time, protect provider NFs from being overloaded. In the event of an overload, the SCP can identify and prioritize the important messages over others and proxy toward the overloaded provider NF.
- **Supporting Canary Testing**: The SCP plays a crucial role in the rollout of new NF releases. It supports mechanisms that allow for a new release to be exposed to a fraction of the users or friendly users. Once successful, the SCP slowly opens up additional users to the new release in a controlled manner, providing confidence to the operator during the rollout.

CONCLUSIONS AND RECOMMENDATIONS

Past experiences indicate that although signaling infrastructure is typically treated as a deployment afterthought and a secondary priority, it is, in fact, one of the most important and critical parts of any mobile network. This is even more important for 5G, with enterprise services and critical communications expected to provide a generous revenue stream for mobile operators and the entire supply chain.

5G signaling infrastructure will migrate to the HTTP/2 domain, but will still be subject to known attack vectors that have been present in both SS7 and Diameter frameworks. HTTP will likely introduce new vulnerabilities that have been well researched and documented in the web domain, where malicious attacks are on a larger scale than in the telco domain. 5G CNs will also need to interface with legacy infrastructure, namely SS7 and Diameter, making the integration of these networks a much more complicated process.

ABI Research urges mobile operators to consider their signaling infrastructure strategies immediately, and not postpone this discussion until 5G enterprise use cases or Standalone (SA) mode become prevalent. Planning for a signaling platform that can scale to include hundreds of different use cases—and services we cannot yet fully understand—will be vital for the success of any 5G network.

Published March 2021

249 South Street
Oyster Bay, New York 11771 USA
Tel: +1 516-624-2500

www.abiresearch.com

About ABI Research

ABI Research helps organizations–and visionaries within those organizations–successfully conquer digital transformation. Since 1990, we have partnered with hundreds of leading technology brands, cutting-edge companies, forward-thinking government agencies, and innovative trade groups around the globe. Through our leading-edge research and worldwide team of analysts, we deliver actionable insight and strategic guidance on the transformative technologies that are reshaping industries, economies, and workforces today.