White paper

Five Steps to 5G Deployment

Version No.: 1.1

May 23, 2019

Authors:

Milan Stolic (mistolic@cisco.com)

Inderpal Singh (insingh@cisco.com)

Jiming Shen (jishen@cisco.com)

Nitin Bhasin (nbhasin@cisco.com)

Contents

Planning and design considerations

Before diving into the five deployment steps, it is important to discuss planning and design as a prerequisite for any deployment.

The transition to 5G requires an extremely well-planned and focused approach for operators to effectively execute their technical strategy and meet business goals. The diverse nature of the use cases demands extremely stringent requirements related to high capacity, faster data rates, ultra-low latency, densification, and reliability. The use case selection and timelines will be key to prioritize the infrastructure transformation.

Before any efforts are made to plan and design 5G infrastructure, let's review a few high-level architectural changes that will take place in different network domains for 5G transformation:

Radio access network

- Implementation of new radio functions (5G New Radio or "NR" is a new air interface specified by 3GPP)
- RAN disaggregation and virtualization
- Integration with 5G Core Network (CN), legacy packet core, and transport networks

Transport

- xHaul implementation
- Data center interconnects in core transport and xHaul networks
- Integration with different radio access technologies like 5G, 4G, and Wi-Fi
- Capacity augmentation and IP fabric evolution

Packet core

- Introduction of new cloud-native network functions
- 5G CN integration with legacy packet core, 5G RAN, and other 3GPP elements
- Disaggregation of core nodes

Additionally, OSS/BSS integration and SDN/NFV capabilities in each network domain are implicit for cost optimization, enhanced user experience, and agile service lifecycle management. Next, key design principles will be addressed that deserve careful planning for the successful implementation of 5G networks.

Early end-to-end planning by use case

The planning and design of a 5G infrastructure should consider end-to-end transformation to include RAN, transport, and packet core. One of the critical steps in this phase is to conduct a readiness assessment of the current infrastructure to analyze gaps and requirements. The following methodology is recommended to conduct the readiness assessment and build a plan to design a next-generation network for 5G services:

- Establish technical requirements for 5G use cases
- Impact analysis on current network infrastructure
- List identified gaps and prioritized recommendations
- Target design to implement the recommendations
- Build a high-level implementation project plan

The outcome of the readiness assessment should provide a solid foundation to accelerate the network transformation for 5G. Additionally, the comprehensive planning done upfront will enable identification of risks that could later become potential roadblocks and impact intended business targets.

Functionality

It is generally understood that readiness for 5G will require a major architectural change in the network infrastructure. The recommendation is to leverage gaps identified in the previous section along with a prioritized use case list to determine the implementation of new functions and features for the target network architecture. The following factors will impact the target network design:

- New functions in each domain (for example, UPF, RAN split function like CU/DU, MEC applications)
- Placement of new functional nodes
- Technology and feature selection for each new function
- Virtualization and cloud-native support
- Integration requirements

Capacity

One of the capabilities defined in IMT-2020 vision for the 5G network is a massive capacity increase in 5G compared to legacy networks. However, the capacity planning will be driven based on use case requirements, implementation timelines, and forecasted user growth. There are two deployment models related to capacity management that will impact network design:

a. **Centralized** – Currently deployed widely and requires the bulk of the traffic routed through central data centers. Massive backhaul and data center interconnect capacity is required due to aggregation of multiple access networks.

b. **Distributed** – Future model to address massive capacity and low latency requirements of 5G while potentially increasing complexity.

At present, the need for additional capacity is addressed by implementing traditional solutions like carrier aggregation, additional fiber / bandwidth in transport, and deploying incremental nodes in the packet core. While this approach has worked well for operators in the past, it will not be adequate considering the scale of 5G use cases.

Implementation of high-frequency spectrum, RAN split, transport evolution, and decoupling of packet core functions like UPF will be key to augment the overall capacity for 5G. With the distributed design, the large capacity requirements will target the xHaul or traditional access/aggregation portions of the network by terminating certain network functions and user services closer to the subscriber via micro-edge DCs. Also, a trend is to offload Internet and peering traffic closer to the subscriber within these distributed micro-edge DCs. These architectural shifts result in the lower bandwidth capacity needed for more expensive core and backbone links.

Resiliency

The ability to handle and recover from network and network function failures is one of the primary requirements for 5G network. For instance, URLLC use cases may have an extremely low tolerance for changing network conditions like node failures, path failure, packet drops, and packet delay. Therefore, the network design must be resilient and provide real-time KPI monitoring, self-healing functions, and other advanced capabilities in the following areas for demanding 5G use cases:

- Packet core with microservices architecture
- Transport network high availability and fast reroute capabilities
- Advanced traffic engineering

- Closed-loop service assurance
- Geo-redundancy
- Self-Organizing Network (SON) solutions for RAN

Scaling

The 5G network will run services with diverse requirements at a large scale. A service may be short lived or long lived, it may have constantly changing constraints, a service profile may be optimized by the end customer due to their changing requirements, and all of these dynamic changes should be delivered by the network in an agile manner.

By implementing NFV and SDN capabilities, network functions can be quickly deployed, configured, and scaled on demand. For example, when a new premium, low-latency service is required, an additional UPF instance can be dynamically instantiated and orchestrated in the remote data centers. Similarly, traffic for a noncritical, high-bandwidth use case can be dynamically routed through a high-capacity but longer path in the network. In addition, the network must be service aware for auto-scaling of available resources based on current and forecasted performance requirements of a 5G service.

Management

In a hybrid environment with 5G and legacy networking technologies, it is cost prohibitive and inefficient to manage the networks with existing tools and processes. Therefore, a new network management framework must be designed to address the following key objectives:

- Model-driven operations based on standards-defined YANG models
- Network automation to orchestrate a multivendor network and standard APIs for seamless integration with various OSS and BSS applications. The end goal is to drive operating expenses gains, accelerate time to market, and improve reliability
- Analytics for real-time network visibility by leveraging streaming telemetry and to drive actionable insights related to fault management, optimization, and scaling
- Workflow manager to implement business logic for the service lifecycle

Backup and recovery

Having a well-defined process for backup and recovery is an integral part of service management. Whether it is 5G or legacy networks, it is vital to clearly understand different failure points and scenarios to develop a comprehensive recovery plan. In addition, automation must be leveraged to expedite recovery efforts and to eliminate manual execution of repetitive tasks. Following key components must be kept in mind for an efficient backup and restore strategy:

- Configuration of each network function in RAN, transport, core, and other functions in the service chain including the NFVI layer
- Time and frequency of backup including data protection from security breaches and other damages
- Location of backup repository – cloud or physical
- Restoration procedure
- Disaster recovery plan for data center failures
- Employee training

Interworking with legacy networks

The road to 5G transition needs to be evolutionary and must consider seamless interoperability of 5G CN with the legacy packet core. For a long time, legacy networks will continue to co-exist with 5G for several reasons such as coverage, technology

maturity, and use case requirements. It is generally understood that in the initial deployment stage, 5G RAN will not provide 100% geographic coverage currently provided by legacy radio access technologies. There will be numerous scenarios that require session handover back and forth between new and existing radio nodes for service continuity.

Therefore, the 5G network design must account for all integration requirements and dependencies with the legacy packet core and RAN. The implementation approach can be evolutionary based on the use case prioritization and requirements identified earlier.

Similarly, the transport networks must also ensure that new technologies like segment routing seamlessly interwork with existing MPLS and access technologies.

Organizational alignment

Because the 5G CN and other services are going to be virtual, the boundaries of responsibilities within an organization will blur. Therefore, emphasis must be given to evolve the organization and align with the business objectives. A very common approach is to focus on the technology aspects in silos and later care about organizational changes. Although this approach may have worked in the past, 5G requires a high level of synergy between organization and technology transformation.

The organizational alignment needs to address the following:

- Skills and role development around new technologies and Agile/DevOps methodologies
- Operational process optimization and tools readiness with a focus on automation
- Business process changes, workflow automation, and deep integration with OSS
- Governance model cultural shifts and, new modes of operations

What can be automated?

5G will transform the way networks are built and managed to deliver next-generation services for the digital economy. The scale and speed of 5G services will pose new network management challenges for the operators, and those challenges will require automation as a key enabler for closed-loop operations.

Real-time monitoring of the network will no longer be a choice but an absolute requirement that should be fulfilled for efficient management of 5G services. Similarly, automated traffic management solutions must be implemented taking into account changing network conditions and service requirements to optimize end use experience. In addition, data management capabilities like data collection, transformation, analysis, and reporting must be built for closed-loop operations. Automation should also address some of the most common operational tasks related to network function provisioning and configuration, network upgrades, and service lifecycle. In the context of 5G, network slicing is seen as one of the strong candidates for automation to build and manage an end-to-end slice given the use case requirements.

Step 1: Infrastructure and network functions deployment

1.1 Access

5G radio access deployments will be characterized by their dense, throughput-focused, and software-driven nature. Deployment of access infrastructure in 5G will be significantly different from that seen in previous generations of mobile networks. There are three major reasons for this shift:

1. Radio split
2. Coverage challenges caused by higher radio frequencies
3. Radio-agnostic access specific to 5G

As mobile technologies evolved from one generation to the next, the functions performed by base stations and other devices have been distributed using cloud technologies. Cloud-RAN decouples the baseband processing from the radio units, allowing the processing power to be pooled at a central location, thus reducing the required redundancy. One of the most important aspects of C-RAN architecture is splitting of base station functions at distributed locations and centralized C-RAN servers. There are eight standardized options for this split, with two (options 2 and 7) being seen most frequently. Details about the radio split options are documented in 3GPP specifications 38.801 and 38.816. Chosen radio split option will have a direct impact on the transport (xHaul) design and deployment because the transport requirements vary dramatically with radio split options. The figure below shows these split options and general requirements.

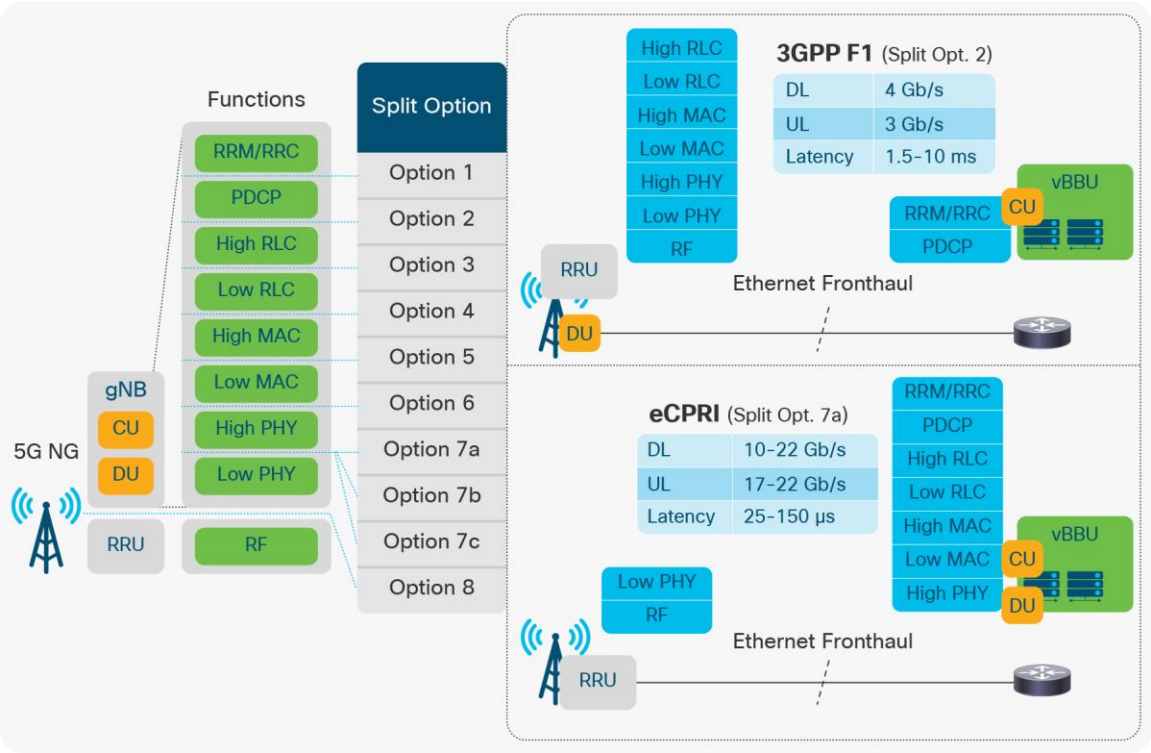

Figure 1.
Radio split options

Higher throughput required in many use cases of 5G networks is made possible by using a higher spectrum. The new 5G bands that regulators are making available will affect how networks are deployed. Prime 5G mid-bands (for example, 3.5 GHz) and millimeter wave bands (for example, 26 and 28 GHz) will suit dense 5G small cell networks in urban hotspots where additional capacity is vital. However, these frequency bands can also suit macrocells for wider area coverage, including fixed wireless access, using beamforming. These technological advancements mean that the 3.5-GHz band can provide the same coverage, and use the same cell sites, as the current 2.6-GHz and 1800-MHz mobile bands. [1]

One consequence of high frequencies is significant propagation attenuation, which means smaller coverage area and, in many cases, poor penetration inside the buildings. All of this has to be factored in when planning and deploying 5G access networks. Densification may be necessary, and it can be accomplished by a combination of deploying additional cell sites (small cells), deploying C-RAN as described in the previous paragraph (multiple radio heads will feed into a pre-agg DU/CU site), and using alternative access methods (for example, outdoors 5G receiver feeding into indoors Wi-Fi network).

Lastly, while in previous generations of mobile networks only cellular coverage was considered, 5G networks are access agnostic. Just like any other aspect of a 5G network deployment, access is also directly dependent on a chosen use case. Depending on this use case, different radio technology may be deployed as appropriate; in many cases, it will be cellular radio, but it can also be Wi-Fi, LPWAN, LoRaWAN, and so on.

Nokia studies [2] on dense deployments in Madrid and Tokyo have shown that a 10,000-fold capacity can be provided in a dense urban environment as well as dense indoor areas. 5G will require a coverage layer that could be provided by macro cells and a coverage layer of small cells providing capacity using the available spectrum range from below 1 GHz to 100 GHz. The indoor capacity will require dedicated indoor small cells. While 5G will provide a significant boost in capacity, the deployment density of 5G outdoor small cells can be limited to approximately 75-m ISD (intersite distance), and for an indoor deployment, an access point in every room is required for coverage and capacity.

The specification of 5G will include the development of a new flexible air interface, NX, which will be directed to extreme mobile broadband deployments in terms of reliability and latency.

1.2 Transport

Transport infrastructure plays an important role for 5G services spanning multiple network domains. The end-to-end service lifecycle should be seamless, and hence an approach towards building a unified transport architecture is essential.

Operators must evolve their current transport networks to provide the following key capabilities in order to deliver 5G requirements discussed in the previous sections:

- Converged IP fabric
- Latency and capacity optimization
- High availability
- Service awareness
- Programmability
- Simplification

The proposed solution is to evolve the transport networks with a common Segment Routing (SR) underlay for service delivery and common EVPN overlay for customer and service segmentation. SR is the next-generation technology for unified network fabric across multiple network domains. It simplifies current transport network architecture and provides the foundation for elastic programmable networks. To maintain consistency, the SR network is referred to as an "SDN fabric" in this document.

In this section, key 5G transport network domains will be covered along with deployment considerations for operators to evolve their transport networks.

1.2.1 xHaul

Traditionally, backhaul provided transport network between monolithic radio infrastructure (evolved NodeB for 4G) and centralized packet core functions. However, with the RAN split covered earlier in the document, xHaul emerges that includes fronthaul, midhaul, and backhaul networks as shown in the figure below. RAN split will enforce very strict requirements on the xHaul for bandwidth, latency, and high availability to support precise clock synchronization between disaggregated radio functions, additional capacity due to radio densification, and mmWave deployment. The xHaul must also deliver on foundational capabilities mentioned earlier for seamless integration with radio and core networks.

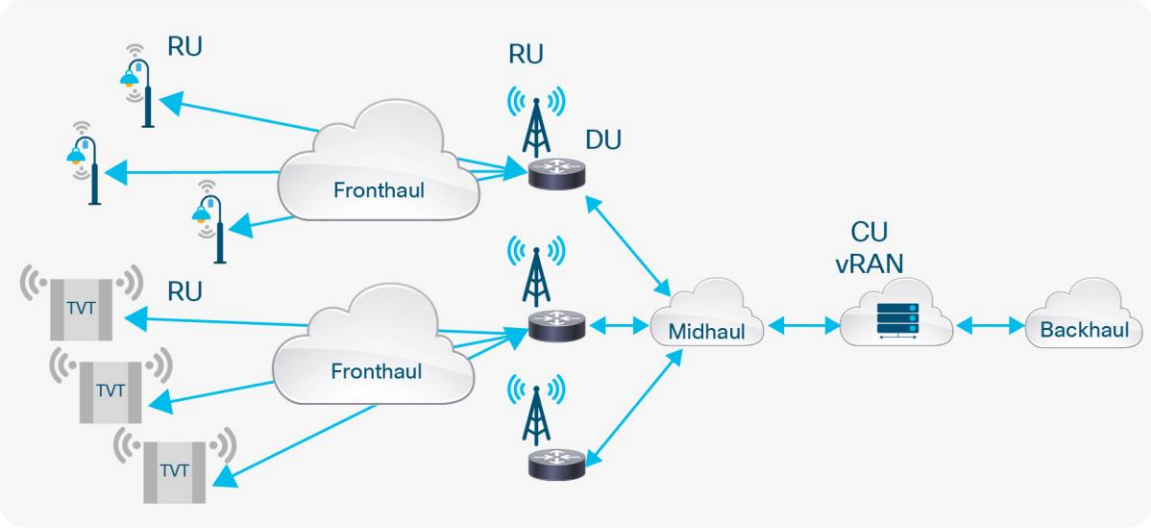

Figure 2.
Transport infrastructure

Deployment of xHaul will likely be driven by 5G services timing, technology maturity, and vendor readiness. It will be a daunting task for operators to evolve and implement xHaul network architecture with a single-step approach due to inherited risks and technology readiness. Hence, it is logical to break xHaul deployment into smaller manageable steps and align the progress with an operator's business goals. Here is a summary of keys steps for xHaul deployment:

1.2.1.1 Requirements

The deployment of xHaul will be driven based on the use case requirement, and not every use case will require RAN disaggregation, which means current fronthaul and backhaul may be sufficient with capacity augmentation and technology upgrades. However, for those 5G use cases with strict SLA, the design of xHaul depends on factors like RAN split, service functions placement, and use case constraints. Examples of such factors are:

- RAN disaggregation
- Quality of service per 5G use case
- Mobile edge computing and traffic offloading

1.2.1.2 Technology selection

Converged IP transport architecture with programmable capabilities is one of the key requirements to implement main features like cloud-RAN, CUPS, network slicing, and MEC for 5G services.

The backhaul network typically combines multiple network domains between the cell site and core transport network. These networks are stitched together for end to end service provisioning using static and complex operational processes. Any ongoing change related to service optimization or maintenance requires an operator to undertake time-consuming and lengthy cross team collaboration, which often can be error prone. With 5G, operators need to focus on agile service delivery and quality rather than managing the complexity of the xHaul infrastructure, and that calls for network simplification coupled with the SDN fabric discussed earlier in the document. It is recommended that the technology selection for xHaul networks delivers the following key attributes:

- End to end application control and segmentation
- Dynamic traffic steering for RAN split options
- Intelligent capacity management and visibility
- API support for automation and assurance
- High resilience and security
- Simplification

It is extremely important that the selected technology for xHaul evolution is standards based and future ready and provides backward compatibility. SDN fabric built with SR has seen industry-wide adoption in just a few years and addresses all key attributes for the xHaul. It provides common underlay and EVPN-based overlays that can span all network domains between the service endpoints. This significantly reduces network touchpoints, service provisioning time, and complexity due to its programmable capabilities.

1.2.1.3 Impact assessment

Before implementing a new xHaul architecture, the impact of change must be carefully reviewed and understood by the organization. The impact assessment may be done to cover all technical and nontechnical aspects like deployment costs, skill readiness, operational processes, and so on. A few key areas important for operators to assess before implementing xHaul changes should be addressed:

- Capacity assessment for new use cases
- Placement of virtual radio, packet core, NFs, and MEC applications; compute availability
- Integration with existing RAN and transport networks
- Technical specifications of xHaul transport nodes (synchronization for RAN split)

- Skill readiness of engineering and operations for new technology adoption
- Cost and ease of new technology implementation
- Operational process changes
- OSS readiness and integration

Once, key areas for xHaul deployment have been looked at, all risks and gaps that could impact the intended design of xHaul networks to deliver expected 5G service performance must be addressed or mitigated. The importance of operational change is often overlooked during a significant technology shift. However, in the case of 5G, technology transformation needs to be lock stepped with the operational transformation.

1.2.1.4 Implementation

xHaul implementation is expected to be driven by 5G use case prioritization done during the preparation phase. Disaggregated packet core and RAN functions will be deployed closer to the cell site locations that will trigger the deployment of fronthaul, midhaul, and backhaul networks. Fronthaul routing nodes will be introduced to provide ultra-low latency communication between Remote Radio Unit (RRU) and Baseband Unit (BBU) functions of the disaggregated RAN. In addition, BBU may be further split into separate functions that are connected via the midhaul network as shown in the section 5.2.1.

After the first use case is selected and target cell sites along with the xHaul architecture have been finalized, the following guidelines shall be considered for the implementation:

- Deployment or reconfiguration of fronthaul, midhaul, and backhaul routing nodes
- Underlay network provisioning to enable routing between the transport nodes and segment routing configuration
- Enabling TI-LFA for link and node failure resiliency
- Enabling interoperability between SR and MPLS nodes with the SR mapping server
- Configuring SR on legacy nodes and gradually expanding the SR footprint in the rest of the transport network
- Integrating new nodes, SR, and overlays with the OSS
- Removing legacy configuration for features implemented with SR and related advanced features

The approach outlined above could vary depending upon current transport network architecture, technology selection, and business objectives. Also, the steps described are overly simplified and may not cover every single aspect of the deployment. However, whatever approach operators end up defining must deliver the xHaul requirements and technology capabilities discussed earlier to provide superior quality of experience for a 5G service. Additionally, xHaul networks could be optimized with advanced features like SR Traffic Engineering (SRTE), SR-PCE for path computation, Quality-of-Service (QoS) policies, fast convergence, traffic segmentation with Flex-Algo, and more to meet diverse application requirements. It is recommended to refer to the Cisco Evolved Programmable Network Transport and Metro Fabric design documents for best practices and additional details to design and implement xHaul transport networks. Last but not the least, there is a tremendous opportunity to leverage automation capabilities of a programmable infrastructure to accelerate deployment and migration activities, which not only helps to optimize costs, but also improves time to market for the new service. In the next section, the core transport network and technology changes required for consistent E2E experience for 5G services will be discussed.

1.2.2 Transport core

The transport core network aggregates several xHaul access networks and provides a conduit to the packet core functions and other services hosted in the data center. This part of the transport network is provisioned with adequate capacity to handle aggregated traffic to and from multiple access networks. The focus is on reliable fast delivery of packets, honoring embedded traffic prioritization and routing instructions as well as high availability.

The transport core network architecture change follows similar guiding principles and foundational requirements discussed in the xHaul section. Therefore, the implementation of SR underlay and other related advanced features will be extended from the backhaul to the transport core network elements. Following this approach, the complexity of running different underlay and overlay protocols in multiple transport network domains can be significantly eased. This provides operational simplicity and cost efficiencies as transport core nodes are engineered for specific functions and do not participate in several control plane functions, state reservation, and so on that are otherwise required with some legacy protocols.

From a routing architecture perspective, the transport core will continue to operate in a separate routing domain in most deployments for manageability and scalability and to limit the impact radius during possible failures. The interdomain forwarding is established using SR-PCE-computed, on-demand SRTE policy or via traditional BGP-Labeled Unicast (BGP-LU) implementation depending on network scale and preferred design choice. As also mentioned in the previous section, several reference documents are available for additional design and implementation details to build a 5G-ready transport core.

1.2.3 Data center connectivity

Traditionally, operators implemented packet core and network functions in the Mobile Switching Offices (MSOs) and Central Offices (COs). Data centers for telco operators were either nonexistent or mainly built to host BSS and OSS functions. However, with growing virtualization of network and application functions, many of these MSOs and COs are being converted into edge data centers. The adoption of virtual functions driven by 5G is expected to accelerate in the telco environment due to disaggregation of packet core, radio, and other network functions. Some of the common use case applications areas in the data center are:

- **Ultra-low latency:** Probably the most important requirements for this use case are disaggregated radio functions and gateway services closer to the edge.
- **Traffic offloading:** Mostly related to Internet-destined, best-effort traffic routed through the distributed gateways. The goal for this is to reduce the load on the backhaul and dependency on central data centers.
- **Mobile edge computing:** Video caching and Internet of Things (IoT) application use cases that are bandwidth hungry and latency sensitive and require frequent access to computing systems.
- **Enterprise VPN:** An enterprise customer with mobile UEs hosting applications in their private data center.

To address some of the use cases defined above, micro-edge DCs may start to pop up in the fronthaul or midhaul aggregation networks to host disaggregated radio and core functions (like CU, DU, and UPF). MEC applications and dedicated enterprise gateways may also be deployed in those micro-edge DCs.

Given the importance of micro-edge DCs for 5G, the integration of data centers edge nodes with transport networks must be seamless. There might also be a requirement to enable inter-DC communication for certain functions. Therefore, the data center edge nodes must at a bare minimum provide the following capabilities:

- Border leaf functions to enable connectivity between transport core, xHaul, data center, and networks
- Extending unified underlay and overlay technologies implemented in xHaul and transport core networks
- Service stitching in case different underlays or overlays are implemented (for example, SR <-> VXLAN and L3VPN <-> EVPN)
- Capacity to cater several aggregated networks' bandwidth demand
- Data Center Interconnect (DCI)

In some cases, the common SR/EVPN transport may extend further into the data centers terminating on Top-of-the-Rack (ToR) nodes or even at the host level (VM/Container).

1.3 Packet core

Packet core as the heart of a mobile network is expected to go through significant changes as we move from 4G to 5G networks. Standards for this transformation are in place, and 3GPP documents them in detail as of its release 15.

1.3.1 Requirements

Unlike in previous generations, requirements for a 5G network will directly depend on a use case and may vary dramatically from one to another.

Based on early research by industry, academia, ITU specifications, and guidelines set up by various industry forums and standard bodies, the 5G infrastructure is targeting to deliver bandwidth of approximately 1G-bps download and approximately 500-Mbps upload to end users while maintaining latency of 1 to 10 ms for most applications. 5G envisions to be the mobile standard, delivering a unified connectivity fabric converging all wireline and wireless technologies to support a vast diversity of devices and services.

5G market drivers can be roughly classified into three broad categories: eMBB (Enhanced Mobile Broadband), MIoT (Massive Internet of Things), and URLLC (Ultra-Reliable Low-Latency Communications).

1.3.2 Impact assessment

Before implementing a new 5G network, a detailed impact assessment needs to be conducted to confirm the effect the transition to it would have on the existing 4G network. Only after that a migration plan can be executed. This assessment would confirm, based on the requirements per use case, what are the shortcomings of the existing network, how can the gaps be mitigated, and in which order based on priorities.

While the existing network may be insufficient to meet the new requirements, its existing subscriber base will continue to be supported for many years. Over some period of time, new features will be introduced, followed by the new 5G network that will eventually take over any remaining 4G functionality. Either way, the two networks will have to coexist for a significant period of time, and this coexistence will have an impact on the 4G network that can and has to be assessed. More details about the implementation and migration will be covered in the next subsection.

1.3.3 Implementation

Deployment of the new 5G mobile core will have to go through phases and will differ based on the customer's existing network and requirements for the new one. These phases of deployment are likely to be as follows, and timeline will expect of the mobile core feature availability and SP's timeline.

1. Virtualize today's packet core and shift to cloud-native delivery models. This step provides a standardized approach decoupling hardware and software. It does not prevent an operator from reusing existing physical nodes in the transitional period. Virtualization does not necessarily imply using only legacy virtualization technologies based on VMware, OpenStack, and similar solutions. Cloud-native network functions provide serious advantages with better performance. It is a difficult transition with cloud native promising to be better in the long run. Automation should be introduced in this step as well.
2. Introduce CUPS. CUPS introduction can be a fairly easy step in existing 4G networks and can be executed with minimal or no downtime. It will lay the groundwork for an architectural transformation required for a 4G-5G transition, improve performance, and simplify future steps.
3. Implement 5G NSA architecture (option 3). In this step, a 5G NSA radio should be deployed alongside 4G radio. It will be served by an existing 4G core, while allowing for high-speed connection use case deployment. The focus is on the radio and transport changes, new MME, HSS, and policy features.
4. Deploy a non-4G interworking 5G SA. This phase is designed to provide FWA and limited 5G services in dense urban areas without 4G interoperability. IMS voice is typically not supported in these deployments. The most important upgrades required for this phase are modified HSS, new charging subsystem, and HTTP2 protocol implementation in the control plane.
5. Deploy a 5G SA mobile core. This 5G core provides interoperability with a 4G system in terms of IP address preservation and seamless interworking. Eventually 4G-only devices would be moved onto the converged core in a controlled manner.

The figure below depicts a combined 4G/5G network during the transition period.

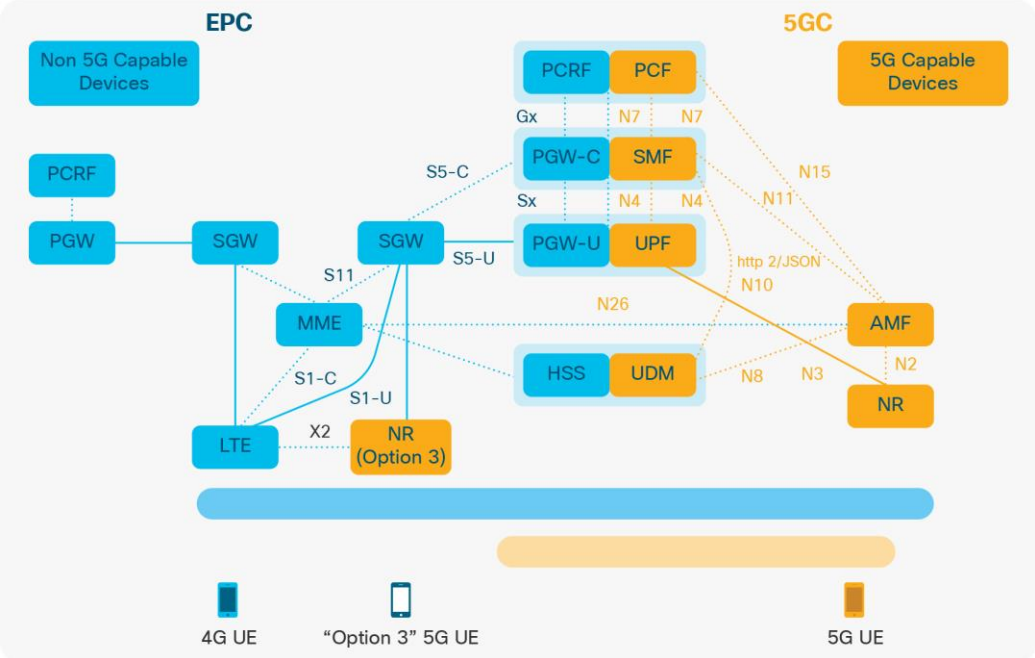

Figure 3.
4G/5G interworking solution

Another important transition path does not involve 5G-NSA deployment. Many operators are considering migrating directly from a 4G network to a 4G+5G core network, especially if they have deployed a virtualized or cloud-native 4G network with automation. These functionalities, required in a 5G core, would provide for an easier transition without 5G-NSA consideration. Service separation in a 5G core based on use case (such as eMBB, URLLC, and IoT) would follow.

Cloud-native principles and technology will help service providers to achieve web scale. Virtualization and VNFs helped us in getting started in moving toward cloud-native applications and being cloud native continues this journey. Service providers must fully automate the deployment and operations of the network. There are specific considerations to realizing being cloud native in the network that are not inherent to web-based, cloud-native solutions such as user plane and protocol considerations.

Deployment of cloud-native network functions provides significant benefits. Some of these benefits are: [3]

- Distributed microservices
- Lightweight footprint
- Service discovery that provides a real-time service registry for all available services
- Lifecycle management applied to each service separately
- Availability and resiliency through service discovery and load-balancing transactions across stateless application containers
- Operational benefits (containerized applications running on bare metal perform better than those running on virtual machines because there is not the overhead of a hypervisor)
- Scalability of each microservice independently

1.4 Programmability of infrastructure and VNFs

A 5G network will be significantly bigger, spanning multiple domains and carrying a large number of highly scalable and agile services. Network functions including 5G CN will be deployed as containers in a cloud-native microservices architecture. The scale of change management and cross-domain orchestration expected for 5G network has not been experienced before by the operators. 5G users expect reliable services provisioned in minutes or seconds instead of days or months. In addition, the traditional network management techniques and tools will struggle to cope with the operational overhead and new technologies in the 5G network. Therefore, a solid automation framework built with programmable capabilities and standard APIs to provide closed-loop operations is at the center of focus for a successful 5G network.

Below is a list of few common tasks that are candidates for operational automation and leverage programmable network capabilities:

- Network or application function deployment
- Day N configuration management
- Service creation, optimization, and removal (for example, network slice)
- Traffic or capacity optimization
- Service assurance and streaming telemetry

To build a highly efficient automation framework capable of automating an end-to-end 5G service lifecycle, following foundational components may be considered for the deployment:

- **Orchestrator:** To orchestrate cross domain activities related to design and configuration of a service. The orchestrator leverages YANG-based data models and standards-based communication protocols (NETCONF, RESTCONF, etc.) for service lifecycle management. The example of a 5G service orchestration is a template-driven automated creation of a network slice covering xHaul, core, data center, 5G packet core, and programmable radio functions.
- **VNF Manager:** For lifecycle management of multivendor VNFs. Integrates with orchestrator(s) and virtual infrastructure manager (for example, OpenStack) for service orchestration.
- **Virtualized Infrastructure Manager (VIM):** To control and manage NFVI resources like compute, storage, and networks. In the future, this could evolve to cloud-native implementation when ETSI NFV specifications are updated and container management and orchestration tools in the automation framework are introduced.
- **Data Collection and Analytics Platform:** The platform enables telemetry-based data collection, ingestion, normalization, consumption, and analysis to build a closed-loop solution. In addition, it supports integration of various operational support applications for use cases like real-time monitoring, network auto-healing, and on-demand capacity optimization.
- **SDN Controller:** Software-Defined Networking (SDN) in the transport network implements the concept of distributed control plane and data plane functions. With this approach, the networks built with segment routing underlay are provisioned and optimized dynamically based on service constraints. For example, an on-demand low latency path is computed and provisioned for an eligible traffic flow using SR traffic engineering and network slicing. Therefore, an SDN controller coupled with segment routing helps an operator deliver operational SLA requirements for 5G services.
- **Workflow Manager:** Provides systematic approach to implement workflow and process automation within the target framework. The workflow manager provides the frontend for service creation and change management to the operator and hides the underlying complexity of multiple operational and change management processes. It may be integrated with multiple orchestrators, SDN controllers, and OSS tools to provide a comprehensive lifecycle automation framework. A simple example of a workflow manager use case is a service provisioning that requires a ticket creation, order placement, MOP creation, MOP implementation and validation, and ticket closure, including multiple checkpoints and rollback options, and all of this executed in a certain order.

Step 2: Network slicing, automation, and onboarding

5G technology shapes the industry with a set of new capabilities. To achieve these capabilities, as well as the scale that 5G needs to provide, network slicing, end-to-end automation, and advanced onboarding techniques should be leveraged in various stages of the deployment.

2.1 Slicing capabilities per domain

Network slicing is a distinctive feature that differentiates 5G deployment from previous generations. It is evolved from functionalities of the dedicated core network and spreads the concepts from end to end. The true network slicing should consider the capabilities to be provided to various domains in the full converged 5G network.

From a service use case perspective, a set of unique metrics is always defined before the deployment. This is the criteria to measure if the provisioned service can meet the end user requirements. The most important and frequently used metrics are:

- End-to-end packet latency
- End-to-end bandwidth
- End-to-end packet loss
- Separation of use case traffic

End-to-end 5G deployment can be divided into access, transport, and core domains, as stated in the previous sections. Therefore, the capabilities of each domain need to be defined so that the key network slice characteristics can be satisfied.

2.1.1 Access domain

3GPP and non-3GPP access can be fully converged in 5G. The access network type itself can be the first level of isolation of the slice. Traffic within the same access network type can be further segregated logically by the traffic types, such as real time and non-real time. The capability is normally provided by a robust MAC scheduler to control the latency and bandwidth.

When dealing with the RAN, advanced capabilities can be further arranged by planning of functional split options with centralized/cloud-RAN solution. This includes careful positioning of Centralized Units (CUs) and Distributed Units (DUs) for different slice metrics. The virtualization of these components can also offer flexibility of capability positioning with orchestration.

2.1.2 Transport domain

The key change in transport domain for 5G slicing deployment is to provide the right resources at the right time without compromising other tasks. Traditional transport infrastructure in mobility is mostly static or semistatic with planning in advance. 5G deployment calls for the transport flexibility in the following aspects:

- Arrange resources for a new network slice in a short turnaround time
- Adjust or stabilize resources for the existing network slice during certain network conditions
- Modify or release resources when required by the network

The techniques behind these capabilities require the use of transport solutions such as segment routing v4/v6 and x-VPN. The core value of these advanced solutions is to ensure that the needed QoS and traffic isolation can be achieved. The transport domain should be able to get differentiated conditions as to the current state of the system and desired state of the system in a timely manner. The detection can be as a result of a network condition probe, application feedback, or even manual intervention.

2.1.3 Core domain

In 5G, the core domain is revamped with service-based architecture and some new disaggregation of network functions (for example, mobility and session management, or control and user plane). These changes provide the much-needed network slicing capabilities.

In deployment, e network functions can be shared or dedicated for different slices, he consumption model is the key to decide what scheme to use. When a use case is performance driven and intensive, it is always better to arrange dedicated NF resources for it.

The 5G core service-based architecture enables the new way to build system procedures in the core. AMF, SMF, and UPF can be selected individually based on standard processes. This makes the slice creation and change much easier to meet end user requirements.

2.2 End-to-End slicing

With all the network domains discussed above providing necessary capabilities, the end-to-end network slicing deployment can be achieved. However, obviously, a cross-domain coordination mechanism shall be set for this task.

The cross-domain activities mainly focused on the following:

- Automation and orchestration
- Slice monitoring and management
- Slice security

Automation and orchestration capabilities are important to anything related to the network slice lifecycle. When a regular 5G slice is deployed, it is expected that the time it takes for it to go alive is measured in minutes or hours, instead of weeks or months. Service providers need a robust and responsive cross-domain automation and orchestration to get it done. This is equally important for the modification or decommissioning of the slice. In order to achieve this goal, domain-specific automation plans have to be prepared, cross-checked, and tested with another domain to ensure a smooth implementation. With more and more service providers and vendors moving to DevOps, the automation and orchestration system should also consider the integration with these processes when needed.

Slice monitoring and management systems are evolved from the traditional NMS and OSS/BSS system. It is the window for multiple parties to understand the operational status of the slices. Deployment of such systems can consider these pieces below:

- Re-use and upgrade existing OSS/BSS system to support cross-domain correlation of stats and other analytics
- Balance the reactive management and proactive management
- Define sets of feedbacks that can be used for automation or orchestration trigger
- Employ intelligent or advanced techniques for reporting, service assurance, and even self-healing support
- Provide for being cross-domain, which means no conflict of information to the operation staffs and end users

Slice security is a new area that every 5G deployment should pay attention to. Given network slices are dynamic and end to end, any security breach in one domain can cause disastrous impact to the whole deployment. With tools and techniques improving by time, the following slice security measures can be considered:

- Interdomain interface security should be examined cautiously as the security implementation and loopholes are quite different
- For architecture with multiple level of abstraction, such as NFV and SDN, it is advisable to provide a clear, standard, and detailed definition of information exchange

Service providers need to pay attention to the inventory system that registers all NFs, of which a majority are dynamically instantiated or allocated. Security policies that must be employed for NFs should be aligned with its lifecycle. In some cases, the lifecycle is relatively short.

2.3 Deployment automation

5G deployment is expected to be automated, in some degree from the beginning, to fully automated when the solution is matured.

3GPP and other domain standard bodies defined many specifications and references for this process. In this section, the deployment is equal to the traditional "day-1" implementation. The automation and/or orchestration should cover one or more of the following processes:

- Design
- Environment preparation
- Preprovisioning of 5G service or slice
- Instantiation (or installation if PNF)
- Configuration
- Activation

It is recommended to seek tools to automate the repeated work first, such as IP planning, common routing configuration, security provision, and so on. These automation tasks can be further combined to form work flows for orchestration with different scenarios.

Another productive approach is to leverage templates or blueprints that are predefined. Coupled by a customized CIQ, new services or slice specifications can be made in a very short period of time and human errors can be maximally reduced. These templates/blueprints can also be integrated into the OSS/BSS or network slice management tools for further "automation-over-automation" solutions.

The methodologies above should be expanded to a multidomain multilevel of abstraction and fully dynamic nature.

2.3.1 Operations automation

Once the deployment is completed, the network moves to the state to "operational," or in traditional terms, "day-N" state. Service providers historically spend huge engineering and technology resources to maintain and excel the condition of the network to satisfy the consumers, especially the critical business users. This effort obviously will hit a bottleneck given the variety and scale of the services that 5G can provide. Consequently, operation automation in 5G is a serious topic for 5G deployment.

Given we are still in the early stage of 5G deployment, we foresee the following automation that can be used in network operations:

- Due to customer requirement changes, security alerts, or service provider internal change, Configuration modifications may be required. upgrades
- Automation in production environments can be dry-run and rolled back when needed and in timely manner. The roll-back should minimize any manual steps
- NFs are recommended to be designed with same programmability mechanism. This can simplify the automation system overhead to the service provider
- Software upgrade was a use case in the past. But in 5G, the expectation is that DevOps and CI/CD will be widely used. It is advisable for service providers to consider the system from the beginning phase of the 5G implementation
- Analytics of the service or slice, especially performance and scale related, should be placed on the high-priority list to implement for operations
- Resiliency, recovery, and remediation automation become the basic requirement in 5G network

2.3.2 Lifecycle automation

The full lifecycle of a service or a slice is typically divided into four phases based on 3GPP specification:

- Preparation phase
- Instantiation, configuration, and activation phase
- Run-time phase
- Decommissioning phase

The following components are recommended for a 5G service or slice lifecycle management system to have:

- A comprehensive network design, ordering, and operation user portal that service provider teams can access to design, implement, maintain, and optimize network slices with an intuitive interface. This portal shall be able to convert the design information into proper network slice instance templates. It can also take requirement parameters of a certain network slice and insert the data into the relevant NSI template to form instructions that can be used by the workflow engine. The portal shall be able to execute operational instructions to monitor slices and accept commands from SP operation staffs to resolve incidents
- A provisioning workflow engine that can accept the command actions in every phase of the lifecycle. This engine can take the instruction from the user portal and work with back-end orchestrations, most of the time multidomain, to implement bespoken lifecycle events of network slices. This engine also needs to track the provisioning status and is able to handle expected or unexpected errors from underlying orchestration activities
- The multidomain orchestrator is the soul to pull all pieces together into a live slice. It interacts with multidomain, multilayer resource stakeholders. Based on each flow requirement's form workflow engine, the orchestrator can execute individual tasks such as instantiation of VNF, configuration of VNF/PNF, and creating links and connections between network slice subnetwork instances. It is also expected that this orchestrator will provide capabilities of automated validation testing of slices to prove the flows are executed successfully
- The assurance platform can pick up the operational side of the slices during its life span. When a slice is created, the assurance platform is able to discover it and starts to monitor the KPIs and manage incidents associated with the slice. This platform should interact with the user portal, or sometimes part of the user portal, to provide up-to-date status information in a very visualized manner to SPs, so that incidents can be highlighted by its severity and handled appropriately. The data from this platform can be used for the judgment of modification or other lifecycle actions of network slices. On break-fix perspective, the assurance platform can also act as a point that works for automated remediation or Root Cause Analysis (RCA)

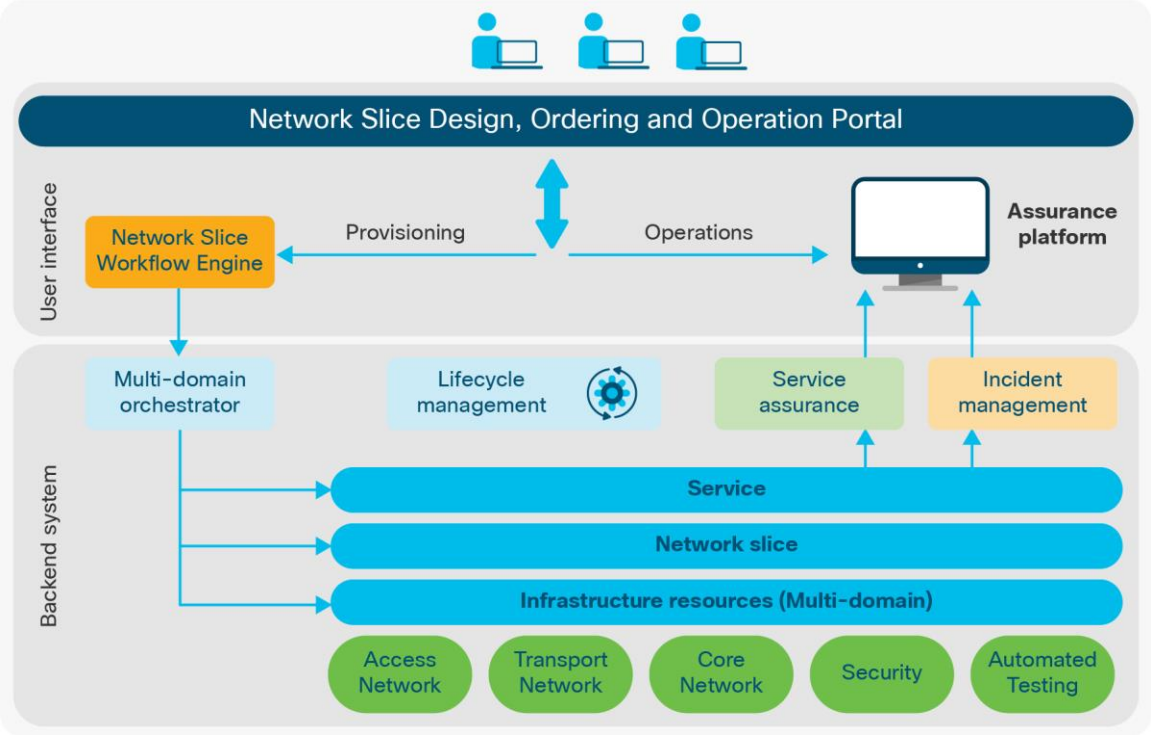

Figure 4.
Lifecycle management architecture of network slices

This lifecycle management architecture offers a benign loop for SPs to provide and adjust services to their end customers, as well as bring efficiency to save engineering efforts and maximize resource utilization. The real-world system is expected to vary and be more comprehensive to be associated with more parts of the 5G network.

Step 3: Application servers and MEC

Application servers have been deployed in service provider networks for a long time. The typical use case is the Telecommunication Application Server (TAS) that is widely used in IMS/VoLTE solutions. As the industry moves to a more service-based architecture in 5G, a wider use of different application servers is expected.

The traditional application servers are deployed with the following aspects:

- They are centralized and normally located in a main data center
- They have static interfaces defined to communicate with a core network
- The core network can share predefined information (for example, reading from an NF database with the application servers)
- They serve highly telco-centric applications

The role of an application server changes when 5G deployment is in place. There are a few considerations below, when operators will deal with the integration of a new variety of application servers.

Besides the traditional telco applications, the industry will see more Machine-Type Communications (MTC)-related application servers emerging. These application servers can provide services for real-time or non-real-time use cases:

- The IoT
- Media broadcast
- Vehicle to everything
- Critical communications

For example, when working with real-time or URLLC use cases, application servers can be distributed with the location of the UPFs, among central, regional, or even edge data centers. This can help the low-latency processing of application data, as well as reducing the operational cost.

In order to make such a solution work in a production scenario, operators need to consider looking outside the traditional telco framework and extend the plan and design effort to the application servers that serve the use cases. The handling of routing/switching, relocation processing (for example, inter-UPF), and load/performance assessment should be made well in advance.

On the implementation side, as the 5G NFs all built with APIs, the communication should be much easier to define. This means not only the static information from database can be shared, but also the dynamic network information is available to application servers. As a matter of fact, the Network Exposure Function (NEF) is a focal point to provide core network information, which is not available in previous-G solutions, to the application servers. This exposure opens a door for more responsive and intelligent applications to be available for different use cases.

Such flexibility opens a lot of deployment work for the operators:

- What group of APIs can be opened for application servers?
- What information, in what frequency, can be shared with application servers?
- What kind of exposure method can be implemented? Passive/semiactive/fully active?
- How should you differentiate the treatment of internal and external application servers?
- What security measures are needed?

Looking a bit deeper into the integration side, the industry is putting together efforts to define a common API framework for application servers to communicate with 5G components. It is advisable for operators to follow the standard approach (3GPP TS 23.222) when working with application server integration with the 5G network.

Multiaccess Edge Computing (MEC) has evolved as one of the leading technology areas that provide 5G with the advantages of low-latency use case capability, as well as edge offloading capabilities to reduce capital and operation costs.

The scale and function of MEC deployment are not always the same. "Edge" is a logical term that defines a set of functions and services to be grouped and rendered in the proximity of the access facility. These functions and services would have been centralized in a core facility during previous-G solutions. The placement of the MEC resources will be unique to the specific use case, customer network, and data center implementation and expansion plan.

In positioning of the MEC, the following locations can be candidates:

- Base station
- IOT closet
- MTSO
- Cable headend
- Central office

The determination of the location and/or the MEC suitability may rely on the following considerations:

- Latency reduction requirements
- xHaul cost efficiency
- Location awareness requirements
- IOT battery conservation
- Regulatory or compliance
- Localized impact
- Introduction of operational complexity
- Additional infrastructure cost
- Security concerns
- Environmental impact
- Technology maturity

Once the scope and location of the MEC is determined, the operator will work on the details about deployment. The following options should be considered when available:

- Light-weight virtualization platform vs. full-fledged

  Virtualization was known for a relatively fixed overhead. The MEC solution seeks optimal solution with the least footprint with regard to physical space and resource consumption in edge sites. For example, a containerized solution, or multitenancy, may provide similar functionality with less resources and a smaller footprint. However, on the other hand, ruggedized implementation may be needed in certain environments in which MEC is deployed

- MEC management

This includes provisioning, monitoring/analytics, lifecycle management, and fault management. While the main facility has the luxury to install all these management entities, the MEC site should carefully select what to be added locally and what to be provided remotely or asynchronously

- Security

  Both physical and network security are important in the MEC. The MEC should employ the same or higher standard of security, preferably in multilayer, to prevent possible compromise of sensitive information and impact to operations

- Third-party integration

  The integration should follow consistent rules, such as well-defined restrictive APIs. Unnecessary connection points, either physical or logical, should be avoided

Overall, MEC is an extension of a traditional core and a consolidation of disaggregated RAN. Resources are realigned and reorganized to provide new capabilities in 5G. Careful planning for the deployment shall achieve optimal cost with the best monetization results.

Step 4: Securing the network

Securing the network is a part of a "vertical" integration. Planning for security deployment has to be done in parallel with horizontal planning. As the network nodes' deployment progresses, so must their security integration. Security integration in a 5G network should be done at the following levels:

- Node
- Domain
- End to end
- Service/use case

While these approaches are well known and documented in prior generation networks, end-to-end planning, deployment, and testing have never been as important as it is now in 5G. The advent of IoT devices that are expected to flood the networks in billions will pose an unprecedented challenge. Addressing this, and similar challenges, after the network has been deployed will dramatically raise the cost and risk.

Another aspect that has to be considered is protection from DoS- or dDos-based attacks. With 5G providing access to more network bandwidth and especially over time a larger end-point footprint, the impact of these types of attacks is amplified. From a transport perspective, ideally, we would reroute traffic to a blackhole in an automated or closed-loop remediated fashion.

Securing data center and cloud components is becoming critical as the mobile network components are being virtualized and can be deployed on an NFVI or as a microservice component in the cloud. For securing the data center, all the components within the data center should be secured apart from the securing the perimeter. The related components are the NFVI infrastructure security that relates to hardening the NFVI hardware, securing E-W security, VNF and container security such as isolation between the VNFs, detecting malicious behavior of the virtual functions, securing the third-party application and API, securing and segmenting the network interfaces, roaming and peering interfaces, and then securing the user access and the orchestration layer. [4]

The security architecture leverages a foundation of NFVI to provide a layer in the architecture that all other vertical applications (be they EPC or other) plug in to. Therefore, the eight-step process described in the graphic below (to secure the NFVI and the applications that run on top of it and that are orchestrated for it) provides the systemic approach to security in this context (for slicing). Specific features that are part of each network slice will have their own domain-specific functions and security concerns, but the overall approach to controlling the "impact domain" for network slicing is described in the graphic below. [4]

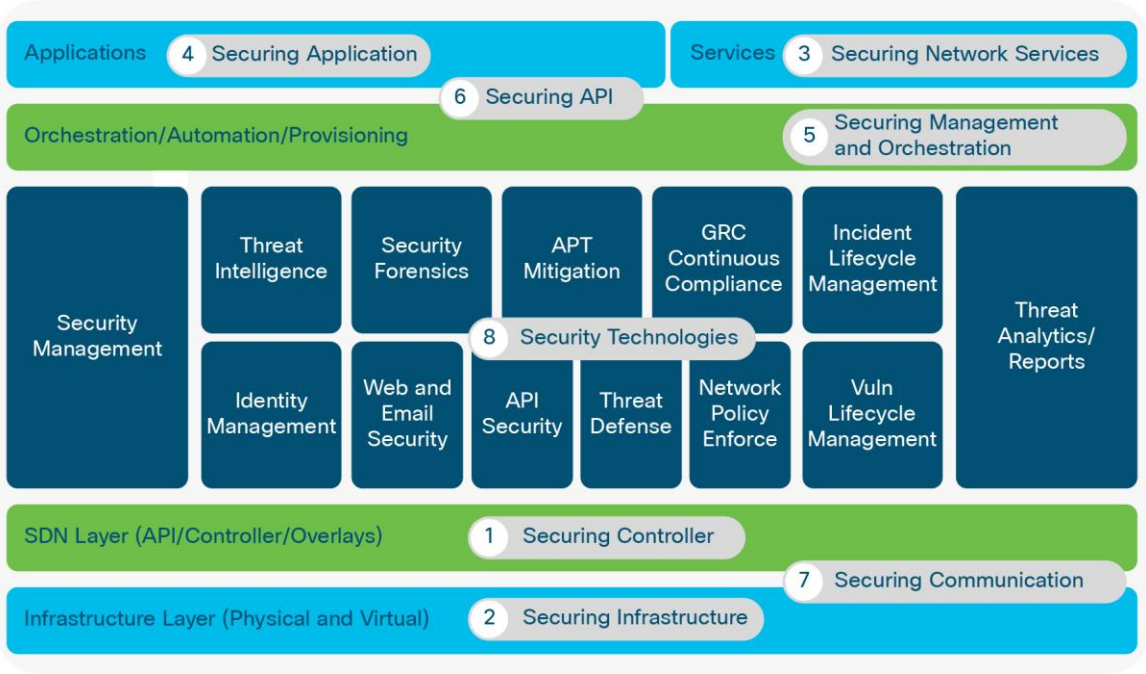

Figure 5.
Eight-step security implementation approach

For mMTC, very low data rates going down to a few bits per day, the extent of security becomes an important consideration (be it authentication, confidentiality, integrity, or otherwise) that can be provided. Several IoT or Machine-to-Machine (M2M) services and devices fall under this category. Examples include temperature sensors giving hourly updates, sensors on farm animals giving vital status couple of times a day, and so on. Such devices will also be resource constrained in terms of battery, computation, and memory. For security, the requirement will be to reduce the overhead of security-related bits (for example, for integrity or for every communication). [5]

At the other end of the scale, URLLC devices will call for high data rates with potentially higher battery and computational resources. Examples include cars, Industrial IoT (IIoT) devices like factory machinery, and Virtual or Augmented Reality (VR or AR) devices used for gaming or real-time services. Providing higher data rates also means that throughput of security functions needs to be considered to avoid processing delay. [5]

Different services would require different levels of security. For example, lightweight security can suit the requirements of IoT while remote healthcare services will demand resilient security. Different services would require differentiated security mechanisms that would rely on flexible security architecture to support end-to-end protection. In a cloud environment, where multiple vendors provide software and equipment for network infrastructure, security concerns may get more complicated. This scenario can be addressed by building an E2E data security chain that would not only reduce dependence on individual link security, but also streamline security management. [6]

It is important for networks to separate virtual network slices to protect the confidentiality of information and prevent one user's resources from being accessed by other users in other slices. For example, company A may choose to block other companies from using its resources even though similar virtual network slices are serving the needs of these companies. [6]

Slice security plays an important role in overall security deployment. Considering the dynamic nature and end-to-end span of network slices, security breaches can have a wide impact. More details can be found in the network slicing section (step 2).

Step 5: Operating the network

Traditional approaches to operations, where the network is designed, put in production, and then transferred to the operations team after the initial soak time and a handover, are no longer valid. There is a number of reasons for that, starting from complexity of the new networks and multiple layers of the stack brought in by virtualization and automation. Complexity, magnitude, and importance of 5G networks from the revenue generation perspective add onto that, forcing a new look at how operations are involved.

It is of an utmost importance to include operations team in the architecture and design process. Also, during the deployment planning and execution, operation aspects step up as very important considerations. Networks can no longer be built with a subsequent consideration about the operations handover; methods, expertise, and tools that comprise operations have to be planned and built in parallel with the network deployment.

The characteristics of 5G technology and services will require a fundamental shift toward a new operating and monetization paradigm, which enables SPs to participate in unprecedented digital value creation chains. 5G-driven services and business models require architectural modernization, which needs to be supported by SP BSS/OSS ecosystems. Following are some of the considerations necessary for a successful deployment of the operations processes: [7]

- Building a monetization infrastructure that allows monetization based on events, transactions, and value of interactions. Monetization parameters include time, priority, or other measurable units
- Strategic investments in real-time, AI-data-driven, autonomous operations comprising automated customer engagement system architectures for 5G business models and use cases
- Digital autonomous intelligent operations (enabled by predictive analytics, machine learning, or AI) will be pivotal to drive dynamic customer interactions comprising various aspects of management and configuration for 5G services
- A dynamically changing partner ecosystem will become a key element of complex multicomponent 5G offerings (such as connected cars, smart homes, and vertical solutions); it is an integral part of workforce, culture, operational processes, and product/service development in wider digital value chain ecosystems [7]

Standard CI/CD and DevOps principles and tooling are used to allow for increased feature velocity as well as consistent deployments. Continuous delivery and DevOps are methods used to automate the process of building, validating, and deploying services into a production network.

Continuous delivery makes an individual application change ready for release as soon as it is ready, without waiting for bundling with other changes into a release. Continuous delivery makes releases easy and reliable, so organizations can deliver frequently, at less risk, and with immediate feedback from end users. Eventually, deployment becomes an integral part of the business process and enterprise competitiveness, taking advantage of testing in the real world rather than artificial labs.

DevOps is the utilization of lean and agile techniques to combine development and operations into a single IT value stream. DevOps enables organizations to build, test, and release software more rapidly and iteratively by applying continuous integration and delivery. To fully realize DevOps, service providers must establish automated continuous integration and deliver pipelines with its vendors. [3]

AI and ML deployments can ease the pain of operating complex networks. Also, deploying predictive analytics capabilities would provide the ability to remediate prior to an event taking place and reduce MTTK/MTTR. The number of the alarms seen by a NOC can easily be in the tens of thousands; reducing the noise by ignoring the insignificant ones and grouping the important ones by correlation while understanding the impact chain is critical. When deploying an OSS system, this is the aspect that has to be considered with high priority.

Deploying a closed loop between the orchestration and fault management systems should also be high on the consideration list. Closed-loop actions enable orchestrated action or implement workflows through a service orchestrator based on the fault category and severity. Closed-loop remediation can automate processes like updating tickets or executing scripts to auto-remediate situations with certain conditions by providing a trigger to the orchestrator. Closed-loop introduction can have a direct impact on the speed of a 5G service lifecycle.

The traditional FCAPS system is still applicable; the challenge that has to be addressed during the network deployment is its expansion to the layers of the stack and further inclusion of security management. In addition to that, operations is responsible for onboarding, whether a new customer, tool, or VNF is being brought in. Considering a direct impact of the onboarding speed to the revenue, the importance of engaging a proper process at a deployment time becomes obvious. For all components of the FCAPS system in the environment where software and hardware are decoupled, correlation is a must in order to be able to extract actionable conclusions. This requires new processes, skills, tools, and organizational changes.

Conclusion

Deployment of a 5G network can be a very challenging task because of the network complexity, dependency on a use case, and migration that needs to take place in most cases to transform the existing network. This deployment, discussed in general terms that apply to deployments irrelevant of a specific use case, can be split into 5 steps. Every single one of the 5 steps shown in previous sections has to be considered from an end-to-end perspective. While some steps may seem to be more extensive and/or more important than others, they all work in close synergy and affect each other. In other words, they cannot be executed in sequential order, but simultaneously in many cases and always end to end. That is what makes both planning and deployment challenging. Understanding these interdependencies between domains and functions from an end-to-end perspective, and following the steps, should ease a deployment.

References

[1] 5G Spectrum; GSMA Public Policy Position, November 2018.

[2] Ten Key Rules of 5G Deployment, Nokia white paper.

[3] Cloud-Native Network Functions, Cisco white paper, June 2018.

[4] 5G Security Innovation with Cisco, M. Geller and P. Nair, 2018.

[5] 3GPP 5G Security, August 6, 2018.

[6] 5G Security Challenges and Ways to Overcome Them, Radware blog, Fabio Palozza, April 19, 2018.

[7] 5G Digital Value Incubation Hinges on Shifting Operating and Monetization Paradigm, Gartner blog, Martina Kurth, October 2, 2018.

Glossary of terms

AMF	Access and Mobility Management Function
ARPU	Average revenue per user
BBU	Baseband Unit
CI/CD	Continuous Integration/Continuous Delivery
CN	Cloud-Native
C-RAN	Cloud Radio Access Network
CU	Centralized Unit
CUPS	Control and User Plane Separation
DCI	Data Center Interconnect
DU	Distributed Unit
EPC	Evolved Packet Core
EVPN	Ethernet Virtual Private Network
IMS	IP Multimedia Subsystem
LoRaWAN	Long-Range Wide-Area Network
LPWAN	Low-Power WAN
MEC	Multiaccess Edge Compute
mMTC	Massive Machine-Type Communications
NFVI	Network Function Virtualization Infrastructure
NSA	Non-Standalone
(R)RU	(Remote) Radio Unit
SA	Standalone
SDN	Software-Defined Networking
SMF	Session Management Function
SP	Service Provider
SR	Segment Routing
URLLC	Ultra-Reliable Low-Latency Communications
UPF	User Plane Function
(V)NF	(Virtualized) Network Function

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **https://www.cisco.com/go/offices**.

Printed in USAs

C11-742416-00 06/19