ERICSSON

# Future network architecture

## Description

Good architecture forms the foundation of communication networks. It powers the most innovative platform for the ecosystem, provides an overview of all system activity, plus supports future innovations. This is a description of the network architecture and how it will evolve over time.

Contents

# 1 Executive summary

The rapidly evolving digital transformation has given rise to a new paradigm across a wide range of industries. This includes an ever-increasing number of highly successful digital companies that do not deploy their own infrastructure.

The ability to support other industries is a major challenge for communication service providers going forward. Different needs apply to different industries as they go digital. Entirely new business models will be possible through advances in connectivity, software, mobile devices and cloud. The term user will be seen from the widest context, it can be anything from a person to an actuator, a sensor, a data center, a content creator, etc.

The future network will need to evolve into a network platform that is characterized by its capability to instantaneously meet any application needs and offers a wide range of capabilities to all its users. It provides a seamless universal connectivity fabric with almost unlimited, scalable and affordable distributed compute and storage. Sensors and actuators can be attached anywhere throughout the network. Latency can be optimized by interacting with the control of access, compute and storage. Embedded into the platform is a distributed intelligence that supports users with insights and reasoning.

The addressability and reachability capabilities make it possible to connect anyone or anything regardless of location and time. Together with the inherent security and availability, the network platform can also meet communication needs relating to secure identification of users and networks. It also provides the scalability to automatically adapt to the exact needs of individual users and applications.

The network platform offering is consumed through an automated digital marketplace. Network services and data are available through consistent and open business interfaces for the applications (APIs). Data, such as location, connectivity conditions and user behavior, can be made available from the network platform.

The vision is of a future network, built on hardware controlled by software, that requires very little, if any, manual intervention. Most task will happen automatically supported by artificial intelligence that is guided via policies that describe the business priorities, and the control software finds the optimal way of achieving it. Technology evolution will take small steps towards this future vision and there are several challenges on the way.

The figure below is an illustration of a model of the future network:

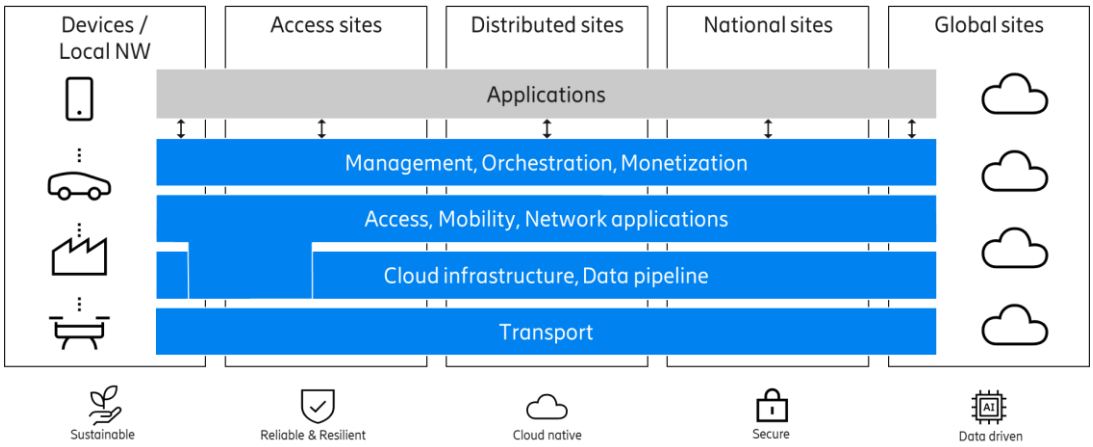


Figure 1 High-level network architecture

Starting from the bottom, in order to lower manual intervention, software defined networking (SDN) is used to manage the transport networks, optical switching will reduce the need for higher layer networking in the access networks and, in parts of the network, white boxes are slowly changing the routing and switching hardware. During the transformation, there will be a mixture of new and legacy equipment, so initially much of the new intelligence will be located to the management part of the network.

The cloud has come to transform how networks, network functions and applications are built and managed. From the initial centralization of the cloud infrastructure there is now an increasing focus on solving the performance challenges through distribution of workloads across the network to where it makes sense to have them. The distributed cloud will span across the network and support many types of workloads. It is important to have a clear separation of concern regarding managing the basic infrastructure and managing the workloads running on top of it. It is however quite possible to create a homogeneous exposure of both workload APIs and cloud resources to other tenants. The data pipeline supports all network domains with collection, storage, distribution and processing of data.

When considering the access-mobility-network applications layer, many of the earlier vertically integrated function are now being transferred to virtual machines and containers, but during a long transition period these will exist in parallel. Moving to 5G will have different migration paths, either based on Evolved Packet Core (EPC) or based on a newly defined core network architecture, 5G Core. New functions will in a flexible way be deployed where it is optimal from commercial, performance or other reasons and the future network will enable this. There will however be parts that for a foreseeable future will remain as native/vertically integrated deployments, e.g. the antenna near parts of the radio functions. Network slicing will be used to realize specific and dedicated end-to-end services to build logical networks on top of a common and shared infrastructure.

The management and orchestration will be where much of the network intelligence will initially evolve towards the zero-touch vision. The network capabilities will be exposed in an efficient and automatic way to other industries. The exposure and monetization will have to attract developers, tenants and the service providers and all will have to be able to innovate fast and make money. The architecture needs to provide simple and stable API that makes the whole network appear like a programmable entity. For the management and orchestration part, ONAP is now gaining momentum and will influence the evolution of this layer.

The future networks will utilize artificial intelligence to become a fully autonomous network with closed loop control and policy governance for dynamic behavior. The automation loops will exist on all levels of the network, from the extremely fast radio loops where the analytical data gets old in milliseconds to the cross-domain optimizations that predicts network traffic and load over long time periods.

Energy efficiency will be vital in the networks of the future. The closed loop automatic optimization on the different layers of the architecture will not only be used for performance but also to save energy and cloud resources. Predictive analytics will forecast the need and take measures automatically to move workloads or power up and scale out when needed.

The open and cloudified networks will be more exposed to threats than the closed systems of today. Open source as well as the exposure of the network resources to multiple industries will open for attacks and there is a need for an even higher degree of security considerations. The componentization and horizontalization of network functions and infrastructure resources moves part of the security handling from product characteristics to deployment choices. The importance of analytics and artificial intelligence will increase for both detection and automatic remedy of security incidents.

All-in-all; Networks are moving towards a hardware agnostic and software defined architecture that will support full automation, short TTM and exposure of the network as a platform for innovation.

# 2 Network Trends

The rapidly evolving digital transformation has given rise to a new paradigm across a wide range of industries. This includes an ever-increasing number of highly successful digital companies that do not deploy their own infrastructure. This revolution in Information and Communications Technology (ICT) is changing the value chain of all industries. Specifically, two fundamental aspects have changed:

- A dramatic digital business transformation enabling unprecedented efficiency through process automation.
- A radical shift to a services economy enabled through cloud-based delivery that result in significant pressure on traditional product value and cost.

The ability to support other industries is a major challenge for service providers going forward. Different needs apply to different industries as they go digital. Future competitiveness will rely on a much broader set of capabilities impacting company strategy. Entirely new business models will be possible through advances in connectivity, software, mobile devices and cloud.

The future network will need to be able to provide for different usage scenarios: high to low capacity, wide to local area, very dense coverage to spotty coverage, private to public, indoor and outdoor etc. We usually refer to these usage scenarios as the 5G use cases. This calls for a scalable, flexible and a modular approach when creating the network. The scalability will be able to go both upwards and downwards depending on the strategy chosen by the owner of the network.

It will be very important that the network provides interfaces that are built with business needs as the primary driver. Open interfaces that support a flexible approach to building new services will provide for a shorter time to customer.

The cost of acquiring and installing a telecom network must be taken into consideration, as well as the cost of operating the network. It calls for automation, machine learning and simplification as well as supporting the important use of abstraction to hide potential architecture complexity.

There are a couple of areas which are important to focus on to be competitive in the transformation to the future new ecosystems:

- Adaptive network quality must be enabled to handle very different use cases. The industry is moving from a consumer-led, smartphone-driven network era to increasing complexity, volume and diversity of requirements. Use cases range from extreme availability and performance (critical machine type communication) to scalability and volume (massive machine type communication).
- Efficient spectrum management will be more essential than today because spectrum will become an even more scarce resource. System coverage and capacity will continue to be a key driver of network design expanding to new higher frequency bands in licensed, licensed shared and unlicensed scenarios.
- Customer service optimization, both from cost and end user satisfaction perspective. Analytics and precise customer knowledge are essential to make this a mutually beneficial experience.
- Flexible business models. The final consideration is the ability for service provider to act in many different business roles, for example as a provider, supplier, customer, broker, etc. Each role dictates different requirements of functionality and business logic and each role requires different insights and segmentations for each business actor at the other end of the transaction.

With increased openness we also see an increasingly complex world with standardization organizations aiming to adopt to more flexible environments, for example 3GPP defining service-based architectures. We also see several open source initiatives in the communication areas such as ONAP and ORAN. All these initiatives aim for increased speed of evolution and reduced cost but also introduce multiple choices which can complicate network evolution.

To address these trends and challenges we have the following high-level view on the future network.

Starting with the transport network, software defined networking (SDN) is changing the way transport networks are handled, optical switching will over time take over more and more of the traditional networking and white boxes are slowly changing the routing and switching hardware. During the transformation, there will be a mixture of new and legacy equipment, so initially much of the new intelligence will be located to the management part of the network.

The cloud has come to transform how applications are built and managed. From the initial centralization of the cloud infrastructure there is now an increasing focus on solving the performance challenges through distribution of workloads across the network to where it makes sense to have them. The distributed cloud will span across the network and support many types of workloads. It is important to have a clear separation of concern regarding managing the basic infrastructure and managing the workloads running on top of it. It is however quite possible to create a homogeneous exposure of both workload API's and cloud resources to other tenants.

When considering the access-mobility-network application layer, many of the earlier vertically integrated function are now being transferred to virtual machines and containers. New functions will in a flexible way be deployed where it is optimal from commercial, performance or other reasons and the future network will enable this. There will however be parts that for a foreseeable future will remain as native/vertically integrated deployments, e.g. the antenna near parts of the radio functions. The architecture supports network slicing which allows networks to be logically separated, with each slice providing customized connectivity, and all slices running on the same, shared infrastructure. Virtualization and SDN are the key technologies that make network slicing possible. Network slices are logically separated and isolated systems that can be designed with different architectures but can share functional components.

All these functionalities are managed by a highly automated management and orchestration layer. This layer handles life-cycle management, day 0 operations of the network slices and network functions, and day 1+ operation. The aim is to achieve close to zero-touch operations and to achieve this, technology such as machine learning, and artificial intelligence are used.

The end goal is to provide an automation stack where intents can be defined to describe the desired state of the network. This intent is translated by the automation platform in an optimized representation of the services and network resources which is then deployed on the network itself. From a networking perspective, this means to enable advanced closed loop automation logic to allow the automation platform to make sure the intent is always ensured. It is essential to provide a high degree of monitoring and control on the automation decisions. The future automation platform will need to provide a supervision enabled model of operation where the user can review and approve the proposed automation steps to build trust in the automation logic.

Intents will be expressed by a model and will be executed by the orchestration platform. Models are becoming the most important artefact describing automation and versioning of models will become essential. Strong Continuous Integration and Continuous Delivery (CI/CD) infrastructure is needed so that new versions of models can be verified and automatically deployed. It can be foreseen to have extended integrations between the vendors and customers CI/CD pipelines.

The management and orchestration layer will be where much of the network intelligence will initially evolve towards the zero-touch vision. It will also in an efficient and automatic way expose the network capabilities to other industries. The exposure and monetization will have to attract developers, tenants and the service providers and all will have to be able to innovate fast and make money. The architecture needs to provide simple and stable API's that makes the whole network appear like a programmable entity. For the management & Orchestration part ONAP is now gaining momentum and will influence the evolution of this layer.

The future networks will utilize artificial intelligence to become a fully autonomous network with closed loop control and policy governance for dynamic behavior. The automation loops will exist on all levels of the network, from the extremely fast radio loops where the analytical data gets old in milliseconds to the cross-domain optimizations that predicts network traffic and load over long time periods.

In a similar way as for the network applications, the non-network applications benefit from the Distributed Cloud Infrastructure. There are several types of physical sites playing different roles, ranging from larger datacenters (national and regional datacenters) where the focus is on compute and storage, to medium sized sites (central office, local switching centers) where wide-area networking plays a greater role, to smaller sites (hub and antenna sites) which are optimized from an access networking perspective

Global connectivity and services have by tradition been deployed in a federated model, where the interfaces are well standardized and offered by one service providers. The complexity with multiple networks has been hidden through interoperability and inter service providers exchange models. However, the rapid deployment of new features makes the traditional standardized federated model hard to use. New methods of enabling exposure of assets from multiple networks is needed, like network asset facilitation and exchange or, on service providers request, aggregation into a single offer.

See also [26] Six key trends manifesting the platform for innovation.

# 3 Network capabilities

## 3.1 Network slicing

Network slicing is a concept to realize specific and dedicated end-to-end services over communication service provider networks. Instead of the prevailing notion of a single and monolithic network serving multiple purposes, it allows to build logical networks on top of a common and shared infrastructure layer. Network slicing is, though outlined as a "5G concept" in the industry, not tied to 5G and variants of network slicing are feasible already with earlier technologies.

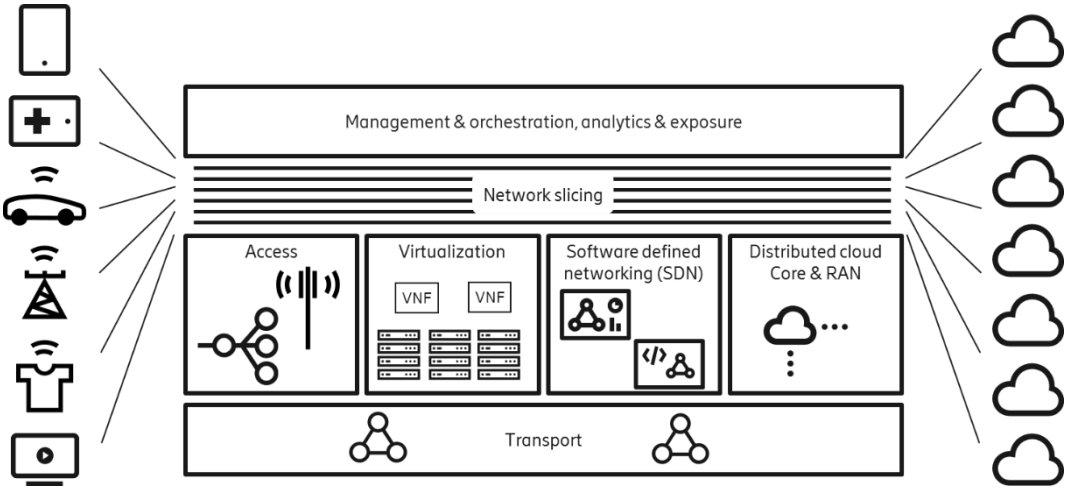


Figure 2 Network slices enabling industry verticals specific use cases

The network slicing paradigm relies on a few enabling technologies. Most important are Network Function Virtualization (NFV), Distributed Cloud, and Software Defined Networks (SDN). A high degree of automation and orchestration is essential, along with equal flexibility in the BSS and service exposure/enablement layer, to create a matching business flexibility.

In addition, network slicing relies on enablers in the underlying network. In EPC-based networks (either 4G or 5G NSA), slicing can for example make use of traditional mechanisms like PLMN ids and APN (Access Point Names), which are supported in 2-4G. With 4G/EPC, the DECOR (Dedicated Core) standard can be used for better control and scalability of network slicing. These are possible to combine with functions such as Radio Resource Partitioning, based on SPID and/or PLMN-ID, to get entry level "end to end" network slicing.

5G Core standards i.e. used with 5G SA, brings additional values. With 5G Core, a device can simultaneously be connected to multiple slices supporting new use cases. Furthermore, 5G Core comes with new signaling functions (Network Slice Selection Function) and procedures that extends to the device. These allows for a better control of how devices connect to slices and for e2e monitoring using common identifiers in RAN and core. Network slicing also extends into IMS and UDM allowing these to be partitioned per slice.

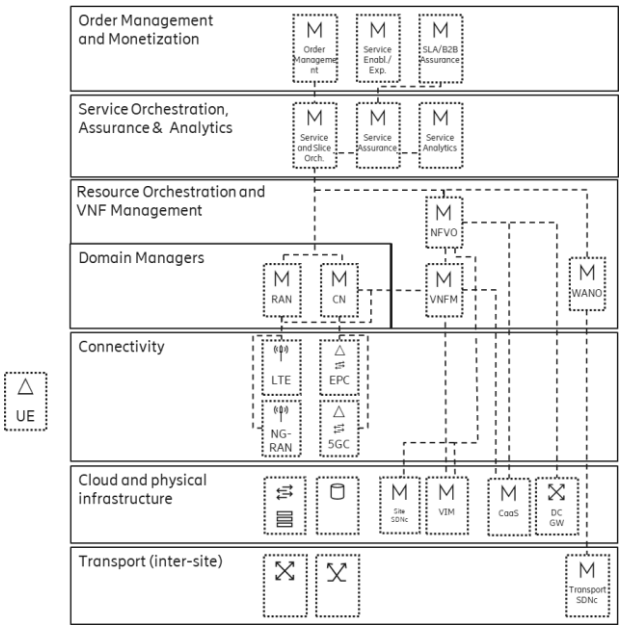


Figure 3 High level Network Slicing functional view

Network slices and 5G services typically involve multiple network domains:

- Order Management & Monetization in the BSS domain makes it possible to monetize the Network Slices and the hosted services. It provides a flexible layer of capabilities for billing, exposure, ordering etc. enabling new business models customized to the needs for 5G customers.
- Service Orchestration, Assurance and Analytics, Resource orchestration and VNF Management together with Domain Managers, provides the fundamental capabilities for orchestration and automation, enabled by model driven orchestration and observability, necessary to reap the business benefits of Network Slicing, i.e. TTM/TTC as well as OPEX. It also has a crucial role in service assurance.
- Connectivity, EPC and 5G Core are key modules/building blocks of network slices deployed to suit individual business/service needs. They also provide key capabilities for steering traffic and sessions (into Slices) and for controlling access to Slices. In LTE and NG-RAN, a NW Slice provides additional information on the policy for handling traffic in addition to the regular QoS for radio bearers. LTE and NG-RAN may serve several NW Slices and each NW Slice may (or may not) contain users with different QoS requirements. RAN may choose to map a single NW Slice to a unique RAN Partition or map several NW Slices to the same RAN Partition. The decision on which resources to isolate and which to share between the RAN Partitions impacts how the RAN service requirement for the NW Slice can be fulfilled.
- Cloud Infrastructure and Transport together with VIMs and SDNs are key enablers providing the fundamental capabilities for infrastructure sharing and programmability.

Different functions, which also gives the selection mechanism between different network slices, can be deployed and combined to support required use case/s as exemplified in Figure 4. Several hybrid cases will exist based on specifics of the use-case to be supported. The upper show slice supporting the massive machine

type communication (MTC) 4G case viable for foreseeable future. The slice supporting local critical MTC using 5GS will certainly be expanded in many deployments with LTE, EPC.

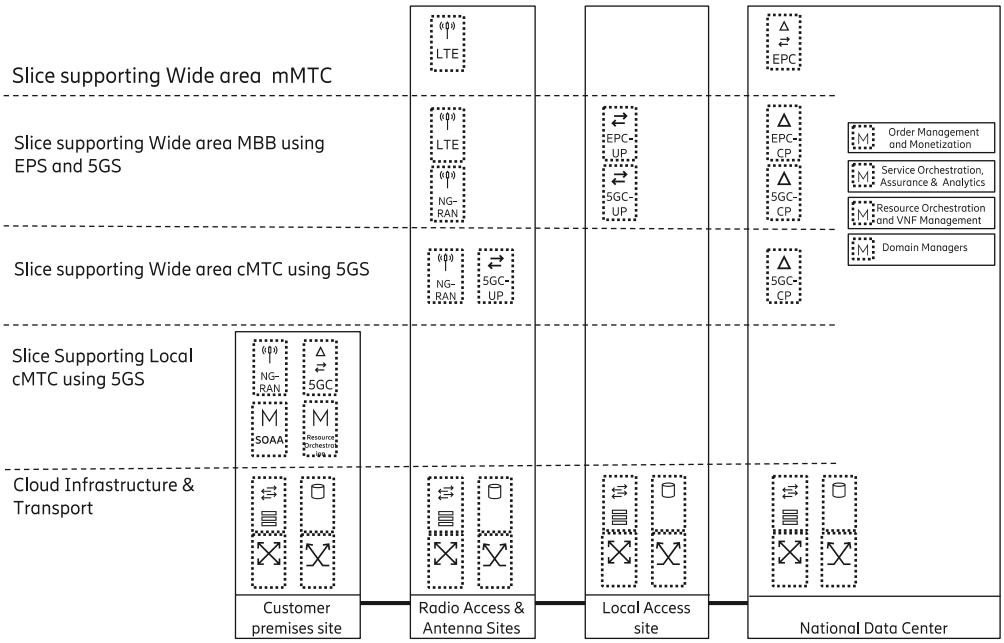


Figure 4 Network Slicing Deployment view examples

See also [2]Flexibility in 5G transport networks: the key to meeting the demand for connectivity.

## 3.2 Security

The networked society will be built on unprecedented connectivity and the ability to access cloud services from anywhere in the world. This connectivity brings with it an evolving threat landscape including an increased risk from a myriad of low tech IoT (Internet of Things) devices which can threaten the network integrity.
Security can be divided in to six different general functions to be used to protect a systems asset.

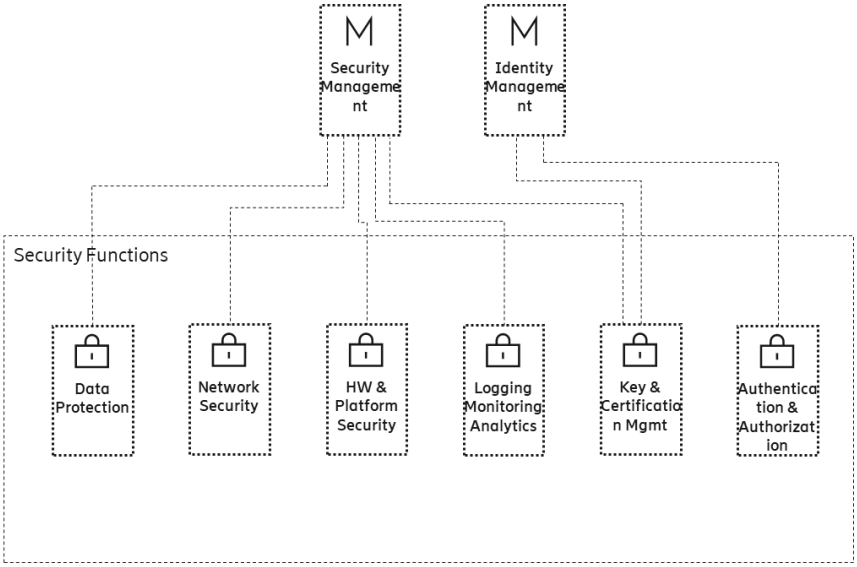


Figure 5 Security functions

The six security functions are managed by either the security management function and/or the identity management function. The functions contain the following functionality:

- Data protection
  Confidentiality and Integrity protection of data in Transit, in Process and at Rest
- Network Security
  Firewall, Security Gateway, Traffic separation, Intrusion Detection
- Hardware and Platform Security
  Trusted execution and storage
- Logging, Monitoring and Analytics
  Generation and collection of security event logs used for monitoring and analytics
- Key and Certificate Management
  Secure key and certificate generation, provisioning and deletion
- Authentication and Authorization
  Identification, authentication and authorization of humans and machines (services, containers, resources, …)

Security is needed in all network domains and use case areas. Below are the illustrations describing how the security functions are used in the 5G RAN deployment architecture and in the 5G core Service Based Architecture.

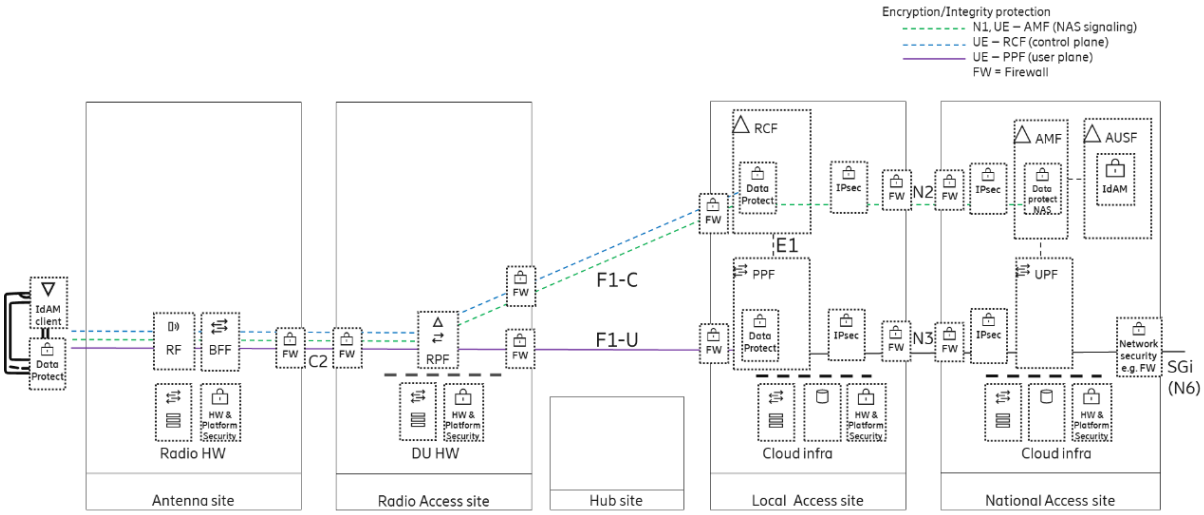


Figure 6 Security in 5G RAN

The 5G RAN, in the above deployment view, is divided between radio and local access sites where the F1 interfaces are deployed between sites. The security functions are deployed based on what the standards require and based on the threats and risks there are on the RAN deployment.

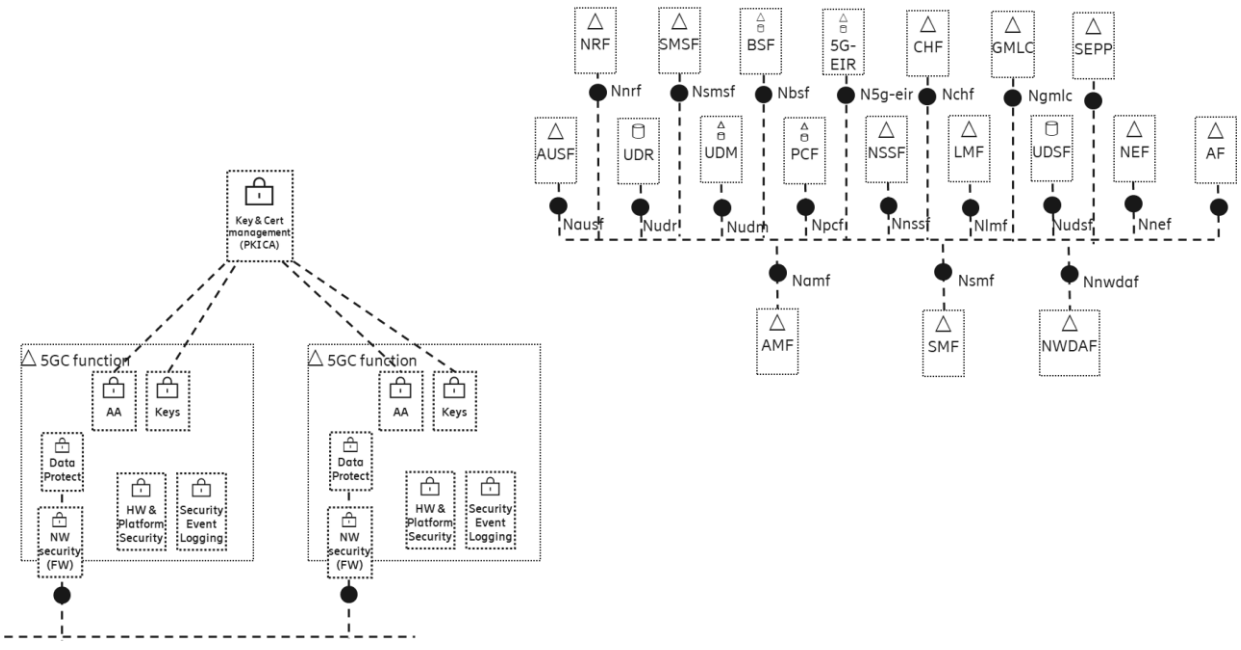


Figure 7 Security in 5G Core

The security functions are shown in two of the 5G core functions for simplicity. Some of the security functions are implemented per node (from bare metal to virtualized like VM/container).

The cloud security architecture addresses security challenges derived from multi-tenancy, divided responsibility and the dynamic environment. The security management & orchestration layer dynamically deploys and adjusts security

functions, policies and related configurations of customers and providers, e.g. in IT cloud and telecom cloud deployments.

Historically (and current) way of implementing security is to protect against the identified threats on the systems. Still a risk-based approach must be taken and mitigate the risks that can't be accepted.

There is architecture aspect on the National Institute of Standards and Technology (NIST) cybersecurity framework - where and how the different mechanisms are implemented.


Figure 8 NIST Cybersecurity Framework

Instead of just only trying to protect against the known threats, systems shall also have the possibility to detect if the system is under attack. Based on the attack, it shall respond and recover to normal state after the mitigation.

The detection if a system is under attack, can be done in different places of the architecture, depending on what information is needed to detect the attack. A node itself can have functionality to identify an attack, but if the node is under attack the detection mechanism may not work properly. The advantage of being able to detect in the node itself is that logs and other data don't have to be sent to an external system for MI/analytics.

In some cases, an attack can't be detected based on information from one source only and an external system will have to collect logs and other data feed from different sources to perform MI/analytics. Below architecture describes a system/function internal and external detect/respond loop.

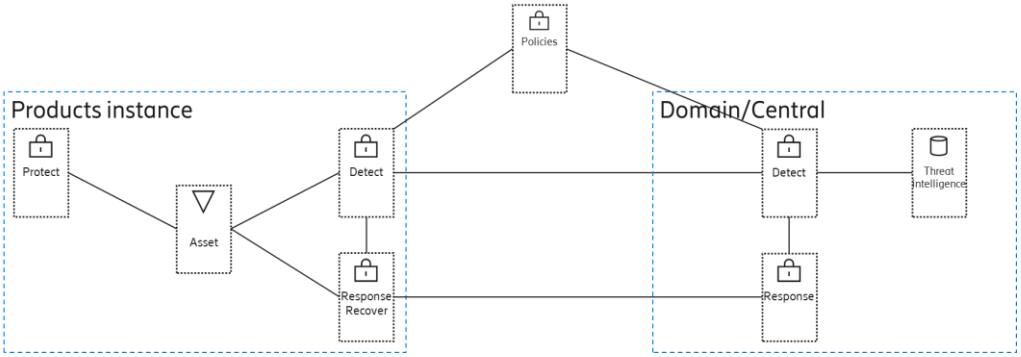


Figure 9 Detect, Response and Recover

To optimize the network, one must have detection functions on different levels from a single node/function/product to domain level and the central Security

Management level. Logs and other data feeds to be used to detect attacks always must be sent to a collection system, but it can be pre-processed to limit the amount of data to be transferred and stored.

See also [5] 5G security - scenarios and solutions, [4] Signaling security, [3] End-to-end Security Management for the IoT.

## 3.3 Distributed cloud

Many of the use cases that can be seen from different industries are driving a need to deploy more and more software functions close to the device or the data source, thereby enabling computing at the edge of the networks. Those software functions are both operator network functions that are part of the network transformation but also non-operator applications that implement those use cases. The location depends on the use case and the business value it provides. The main drivers of edge computing are low latency, high bandwidth, self-contained private or dedicated deployments and resilient and data governed computing and storage.

The distributed cloud provides an execution environment for cloud application optimization across multiple sites, managed as one solution and perceived as such by the applications. It is based on SDN, NFV and 3GPP edge computing. The execution environment may be formed by multiple heterogeneous instances of software with end-to-end service orchestration and external exposure and management interfaces. This enables multi-access and multi-cloud capabilities and unlocks networks as open platforms for application innovations.

A layered network architecture is applied inside and between data centers utilizing software defined networking capabilities of the solution.

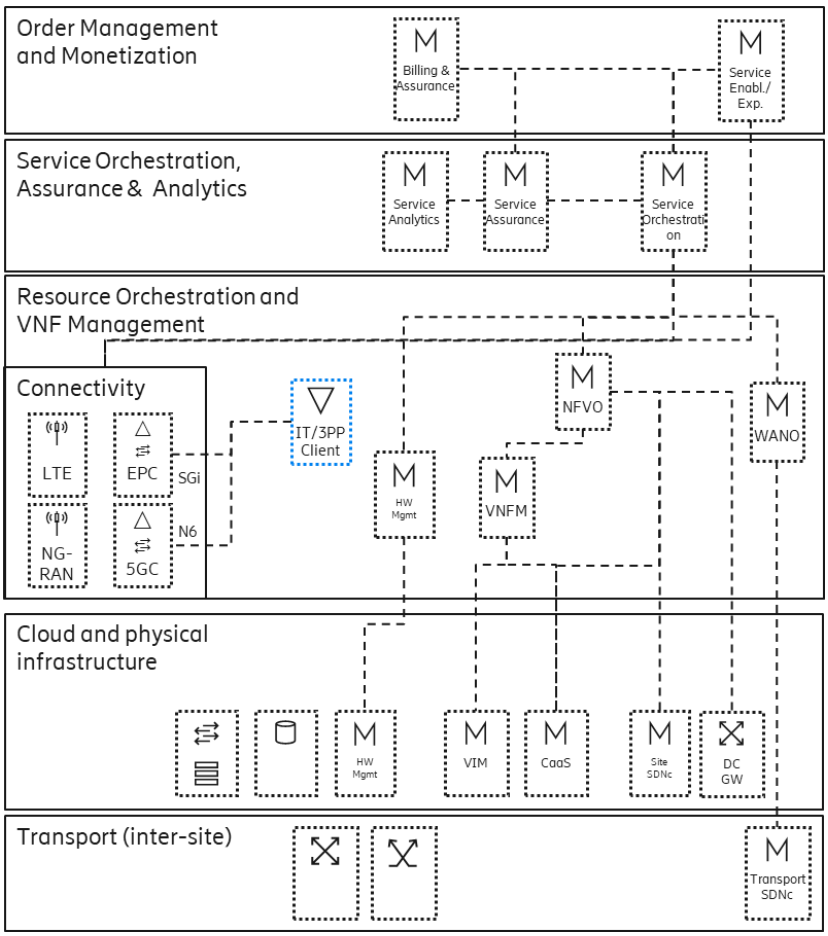


Figure 10: Distributed Cloud functional view

Management include orchestration and placement support as well as monitoring and automation. The same orchestration functions, supporting information model and descriptors are used for operator network functions and non-operator applications. Orchestration cover the request and allocation of resources (placement support), compute and networking, the deployment of the SW components and the establishment of the connectivity according to the constrains and requirements of the applications. A smart resource inventory is required to enable proper application deployment in a distributed cloud. It is the common topology awareness at OSS level, shared with assurance, that enable closed loop automation and smart placement functions.

Application connectivity is supported for both EPS and 5GS. When the applications require to be placed below existing SGi/N6 anchor points, local breakout (LBO) of traffic flows related to the application is supported. This can be achieved by distributing SGi/N6, deploying core network user plane function supporting LBO.

Transport network across sites is set as part of the planning process while connectivity between the required instances will be created as a dynamic overlay on top of that networking infrastructure at instantiation time, depending on the service definition.

To make assets consumable to the applications, three APIs are exposed (and made available through an SDK for application development):

- infrastructure management APIs responsible of requesting resources for runtime (it may include network slices), tenant mgmt., etc. that will support different infrastructure stacks.
- application LCM APIs, including connectivity interaction (LBO, UPF anchoring point, etc.). Applications shall not interact directly with the LBO APIs but instead will request certain requirements (in the application descriptor or at execution time) that the service orchestrator will translate into placement and network connectivity, including LBO configuration.
- and network/application APIs based on network exposure functions.

A distributed cloud comprehends cloud instances with various level of independency, that are managed efficiently, being perceived by the user as one cloud instance although internally it consists of a multitude of virtual infrastructure managers.

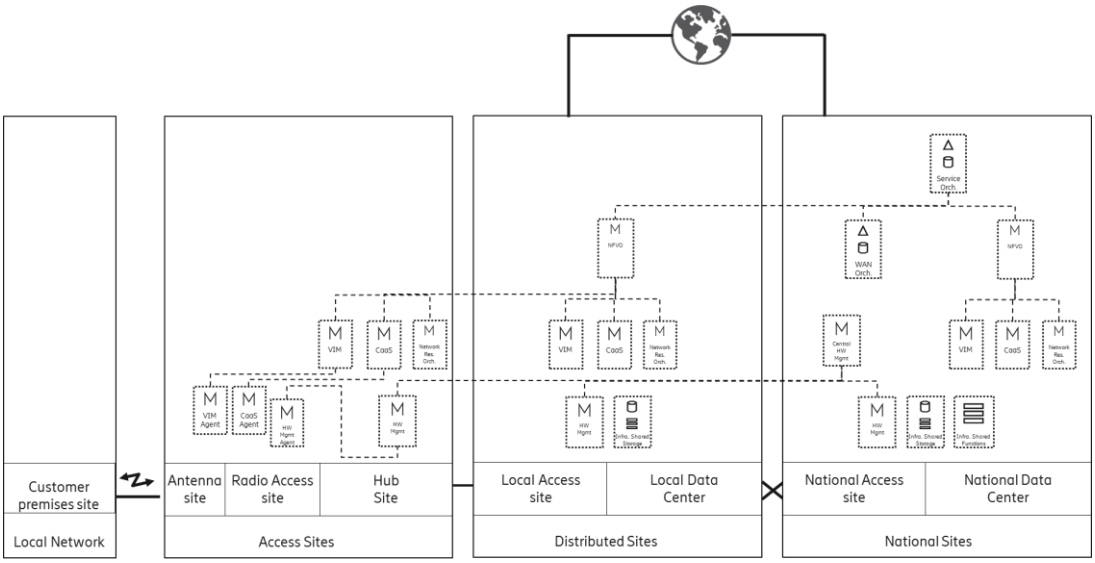


Figure 11 Distributed cloud infrastructure deployment view

Support for the virtualization layers (containers and VMs) and the multitude of deployment options for them impose a rationalization of combinations and options. VMs and containers will be deployed side by side with support for strict tenant separation. Interworking between the virtualization solutions will be supported and will take place on the boundaries between the different instances.

Most sites will host at least one full stack of a virtual infrastructure manager providing IaaS (Infrastructure as a Service) or CaaS (Container as a Service). Very small sites with strict resource limitations may not contain the full management stack but kind of management agents to support remote control, for example using federation, distribution, or other approaches.

See also [6] Virtualizing network services - the telecom cloud, [19] Distributed cloud: A key enabler of automotive and industry 4.0 use cases, [21] Edge Computing and 5G .

## 3.4 Service exposure

Service exposure will be critical to the success of solutions that rely on edge computing, network slicing and distributed cloud. Without it, the growing number of functions, nodes, configurations and individual offerings that those solutions entail represents a significant risk of increased operational expenditure. The key benefit of service exposure in this respect is that it makes it possible to use application programming interfaces (APIs) to connect automation flows and artificial intelligence (AI) processes across organizational, technology, business-to-business (B2B) and other borders, thereby avoiding costly manual handling. AI and analytics-based services are particularly good candidates for exposure and external monetization.

Service exposure is also a crucial part of automation making it possible to automate processes that runs over borders, e.g. processes that runs over B2B interfaces, over organization borders in an enterprise or over different technology domains.

There are three main types of service exposure in a telecom environment:

- Network monitoring
  Examples include network publishing information as real-time statuses, event streams, reports, statistics, analytic insights and so on
- Network control and configuration
  Involves requesting control services that directly interact with the network traffic or request configuration changes.
- Payload interfaces
  Examples include messaging and local breakout to interact with the data streams through local breakout for optimization

The functional architecture for service exposure is built around four customer scenarios. In the case of internal consumers, applications for monitoring, optimization and internal information sharing operate under the control and ownership of the enterprise itself. In the case of B2C, consumers directly use services via web or app support. B2C examples include call control and self-service management of preferences and subscriptions. The B2B scenario consists of partners that use services such as messaging and IoT communication to support their business. The B2B2X scenario is made up of more complex value chains such as mobile virtual network operators, web scale, gaming, automotive and telco cloud through web-scale APIs.

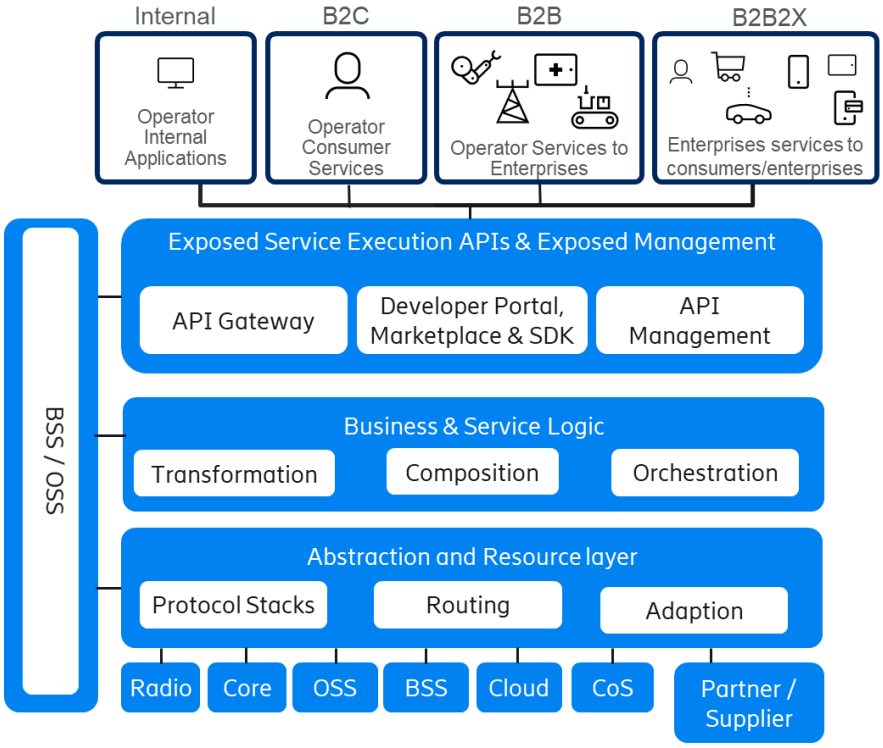


Figure 12 Network exposure functional view

The functional architecture for service exposure is divided into three layers that each act as a framework for the realization. Domain-specific functionality and knowledge are applied and added to the framework as configurations, scripts, plug-ins, models and so on. For example, the access control framework delivers the building blocks for specializing the access controls for a specific area.

The abstraction and resource layer are responsible for communicating with the assets. If some assets are located outside the enterprise – at a supplier or partner facility in a federation scenario, for example – B2B functionality will also be included in this layer.

The business and service logic layer are responsible for transformation and composition – that is, when there is a need to raise the abstraction level of a service to create combined services.

The exposed service execution APIs and exposed management layer are responsible for making the service discoverable and reachable for the consumer. This is done through the API gateway, with the support of portal, SDK and API management.

Business support systems (BSS) and operations support systems (OSS) play a double role in this architecture. Firstly, they serve as resources that can expose their values – OSS can provide analytics insights, for example, and BSS can provide "charging on behalf of" functionality. At the same time, OSS are responsible for managing service exposure in all assurance, configuration, accounting, performance, security and LCM aspects, such as the discovery, ordering and charging of a service.

One of the key characteristics of the architecture presented is that the service exposure framework life cycle is decoupled from the exposed services, which makes it possible to support both short- and long-tail exposed services. This is realized through the inclusion and exposure of new services through configuration, plug-ins and the possibility to extend the framework.

Another key characteristic to note is that it is possible to deploy common exposure functions both in a distributed way and individually – in combination with other microservices for efficiency reasons, for example. Typical cases are distributed cloud with edge computing and web-scale scenarios such as download/upload/streaming where the edge site and terminal are involved in the optimization.

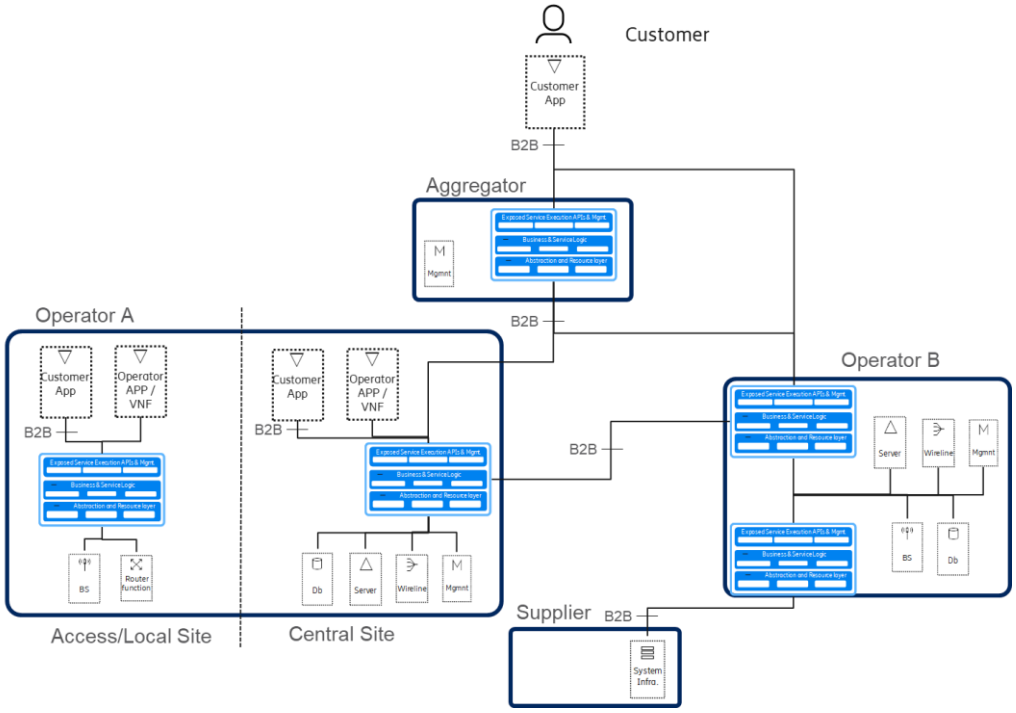


Figure 13 Network exposure deployment

Service Exposure can be deployed in multitude of locations, each one with different set of requirements which drives the modularity and configurability requirements.

In the Operator B case in the picture is Service Exposure used to expose services in a full B2B context. Here is the BSS integration and support required to handle all commercial aspects of the exposure and LCM of Customer, Contracts, Orders, Services etc. and Charging and Billing. Operator B also acquires services from a Supplier, also here are B2B commercial support deployed.

Operator A deploys the Service Exposure both at the Central Site and at the Edge Site e.g. for latency or payload requirements. Services are exposed to operator A own APP/VNFs, limiting the need to B2B support. However, some of the APP are owned by an external partner in a hosting scenario, both centrally and at the edge, this drives the full B2B support deployed for the externally owned Apps.

Aggregator in the picture deploys the Service Exposure for creating services as a combination more than one supplier. UDN and Web-Scale integration both falls into this category. This is also a B2B interface with specific requirements as exposure to the customer is done through the aggregator, e.g. advertising or discovery of services are done through the Web-Scalers portals.

Operator A and Operator B also have a federation relationship handled by the Service Exposure where both parties are on the same level in the value chain of the ecosystem. Here is a subset of the B2B support deployed.

See also [7] Service exposure: a critical capability in a 5G world, [8] 5G network programmability for mission-critical applications .

## 3.5 Machine learning and AI

AI/ML technologies open new possibilities to enable the network to provide machine created intelligence and to use this intelligence for system automation and optimization as well as insight creation both on node and network level as well as on higher OSS/BSS levels. New use cases based on AI/ML will have an impact on the network functionality and on the network architecture. The two main areas of AI/ML that concern the network are Machine Learning and Machine Reasoning, both with different purpose and system requirements.

Machine learning and machine reasoning can both be used to build intelligent logic, but they have different approaches. Machine learning is typically used for learning a complex function from vast amounts of data – for example system anomaly detection, KPI trend forecast, user behavior clustering, alarm/event correlation using supervised and unsupervised learning, as well as policy optimization using reinforcement learning.

Machine reasoning, on the other hand, implements abstract thinking as a computational system. Such a system contains a knowledge base storing declarative and procedural knowledge and a reasoning engine employing logical techniques such as deduction and induction to generate conclusions.

The architecture and logical placement of functionality in the network is influenced by several factors such as who owns the equipment and where the corresponding equipment needs to be placed due to bandwidth or regulatory requirements. It may also be influenced by bandwidth limitations for transfer of data. Different deployment options will impact the architecture like centralized or decentralized inference, model training and update mechanisms.

Data management from data collection, ingest, transport, storage to parsing, cleaning and feature selection is an essential part for the development and training and update of ML models and therefore an important aspect in the network architecture. 3GPP is standardizing new functions in the Service Based Architecture, the MDAF (SA5) and NWDAF (SA2).

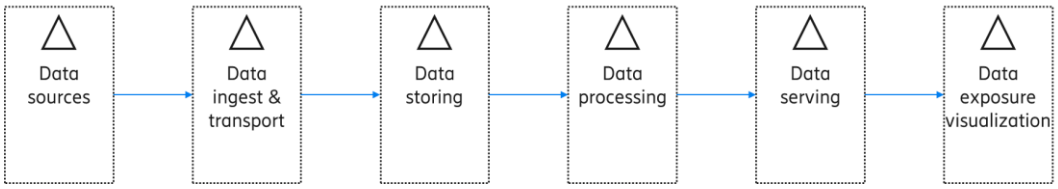


Figure 14 Components for managing data transport and its processing

In general, there are many data sources. Management systems and functions such as service exposure are examples of data sources.

The data sources generate data or insights which shall be brought in for further analysis or presentation. To ensure that data is only collected once, a data routing and distribution mechanism which can feed data from one source to multiple consumers is needed. The ingest and transport function is followed by the functions needed for storage, processing and presentation/visualization.

Data in transit always needs to be encrypted and may also need to be encrypted at rest depending on the type of data. The integrity of the data needs to be ensured, and data will need to be transferred through gateways and firewalls. Access to data pipeline data and functions will be protected by secure keys and certificates, and subject to authentication and authorization.

In a large distributed system, decisions are made at different places and different granular levels. Some decisions are based on local data and governed by tight control loops with low latency. Other decisions are more strategic, affect the system globally and are made based on data collected from many different sources. Decisions made at a higher global level may also need real time response in critical cases such as power-grid failures, cascading node failures, and so on. An intelligent system that automates such large and complex systems must necessarily reflect the distributed nature and support the management topology.

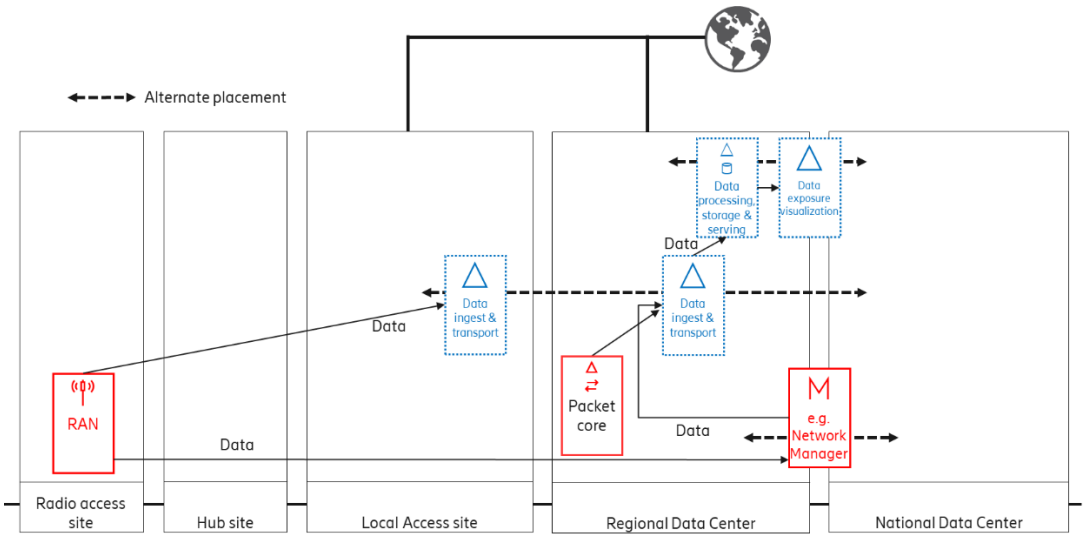


Figure 15 Typical deployment of logical functions in Machine learning and AI

Data generated at the edge, in a device or network edge node, will at times need to be processed in place. Data may not economically be transferred to a centralized cloud; there may be laws governing where data can reside and there can be privacy or security implications of data transfer. The scale of decisions in these cases is restricted to a small domain, so the algorithms and computing power necessary are usually fast and light.

However, local models can be based on incomplete and biased statistics which may lead to loss of performance. There is a need to leverage the scale of distribution, make appropriate abstractions of local models and transfer the insight to other local models. Learning about global data patterns from multiple networked devices or nodes without access to the actual data is also possible. A recent approach is so called federated learning: learn local models based on local data patterns, send the local models to centralized cloud, average them and send back the average model to all devices.

Global decision making, on the other hand, relies on knowledge of the global state and global data patterns – that is, patterns of events occurring across local nodes. The global state is built by combining the knowledge from multiple components and using it to make decisions.

A common distributed and decentralized paradigm is required to make the best use of local and global data and models and determine how to distribute learning and reasoning across nodes to fulfill extreme latency requirements. Such paradigms themselves may be built using artificial intelligence to incorporate features of self- configuration, self-optimization and self-healing to support network planning, provisioning and assurance.

See also [9] Artificial intelligence and machine learning in next-generation systems, [10] Cognitive technologies in network and business automation.

## 3.6 Automation

Our vision is a network that is easily controlled through higher level business intents, has a high degree of automation and self-optimization, and continuously learns to improve over time. The network is robust and resilient to unforeseen events and provides the highest level of security to all its users. We call this the zero-touch network.

This changes the approach to management, from managing networks manually to managing automation.

A key part of managing automation is the control loop paradigm, as shown in Figure 16. This is achieved by designing the insights and policies required for a use case, and the decisions that need to be made. Insights are the outcome when data is processed by analytics. Policies represent the rules governing the decisions to be made by the control loop. The decisions are actuated by COM (Control, Orchestration, Management), which interacts with production domains by requesting actions such as update, configure or heal. The control loops are

implemented by multiple OSS functions, can act in a hierarchical way and even delegating to the cloud infrastructure that is becoming more capable.

By leveraging analytics, AI and policy, control loops become adaptive and are central to assuring and optimizing the deployed services and resources.

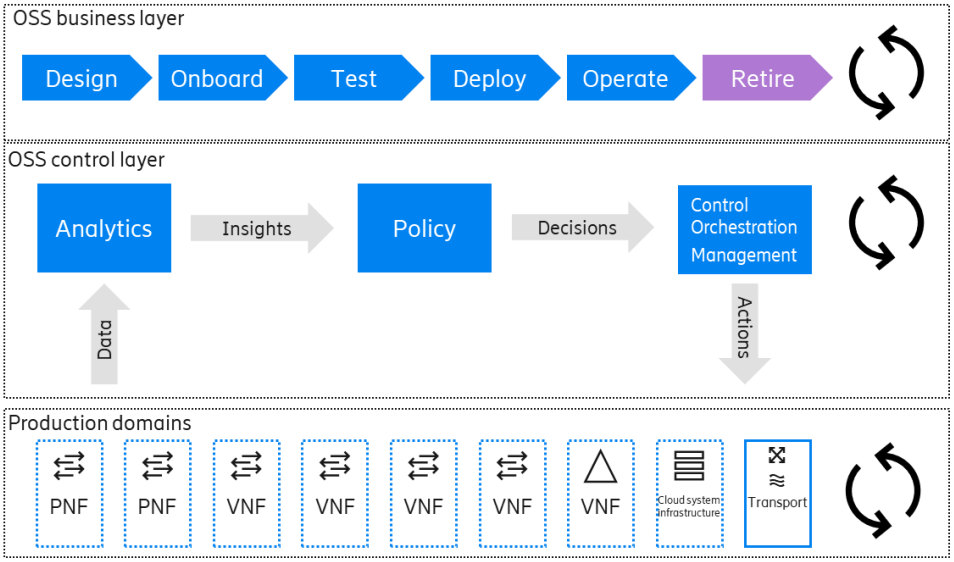


Figure 16 Analytics and policy-driven automation: closed control loop

Figure 17 describes the major functional components, that shall be used to realize automation. This view captures a design-time set of functions (studio-suite) and a run-time set of functions (all others) that will interwork to realize the automated life-cycle management of network functions and services. Machine learning capabilities are present in the assurance and analytics domain.

The drive for automation will address several aspects of the business processes spanning from automating parts of or complete flows to removal of complete steps by providing automation in the network domains. Machine learning will be an essential part as well as advanced analytics.

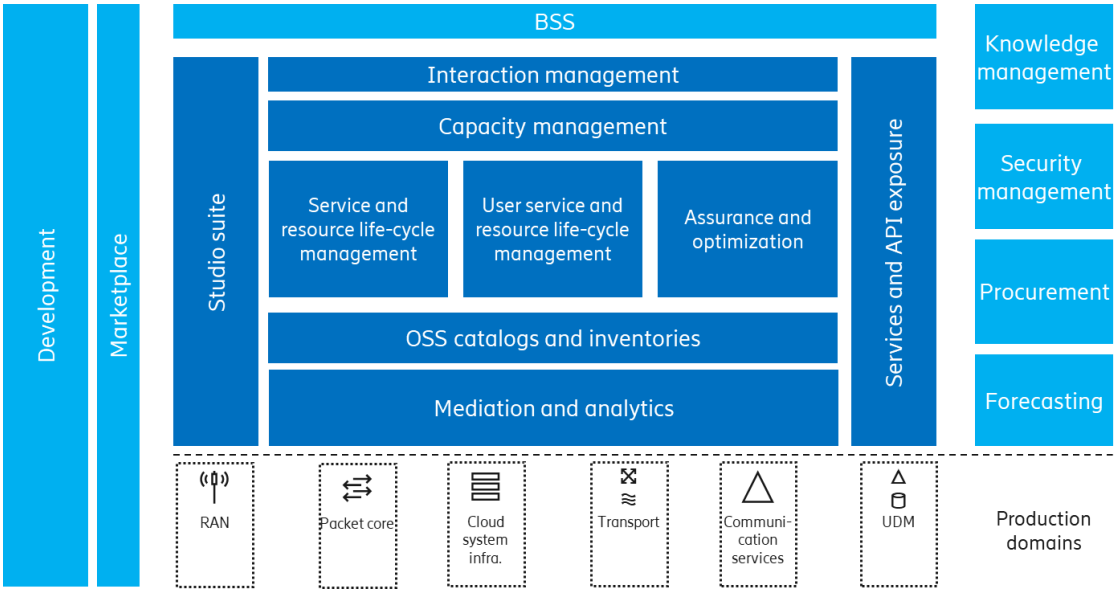


Figure 17 Automation functional view

Automated management in Figure 18 is addressed in a pattern which is described through the application of the tools of Control, Orchestration, Management, Analytics and Policy to closed loops in:

- Infrastructure incl. at least transport, compute and storage
- VNF and cloud runtime environment
- E2e service and network slicing
- User experience

The three key values are:

- Rapid service introduction and lifecycle management
- Dynamic lifecycle management of networks
- Automation of business and operational processes.

These values are achieved through:

- An architectural pattern for advanced automated management.
- A wide policy framework for policy definition, deployment and execution of service and resource policies
- Analytics based on a common architecture for policy and AI/ML enabled assurance.

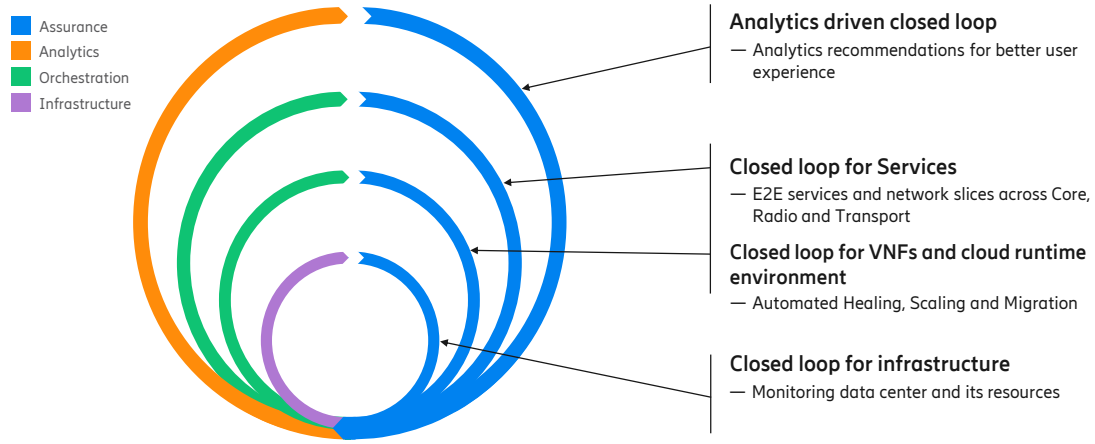


Figure 18 Automation use-cases

See also [11] Open, intelligent and model-driven: evolving OSS, [12] Architecture evolution for automation and network programmability.

# 4 Network architecture domains

The following figure shows a high-level overview of the future network divided up into a few different domains:

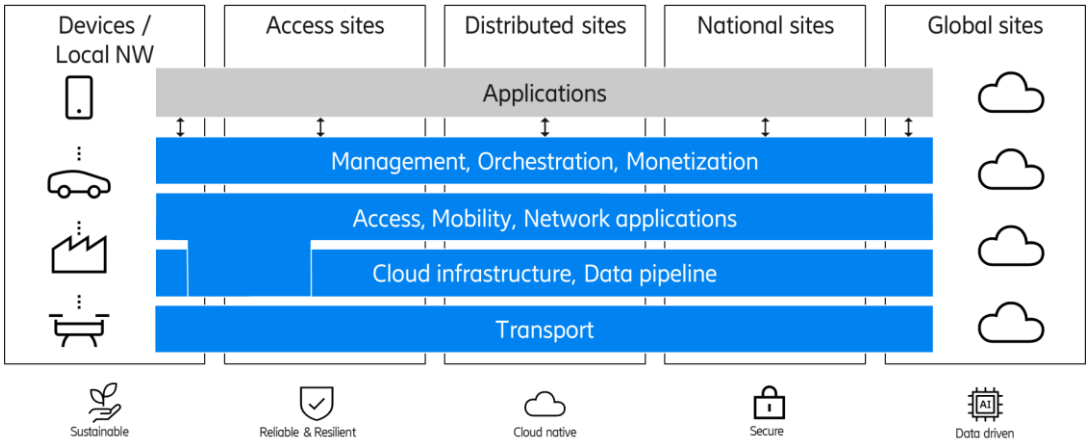


Figure 19 High-Level Network Architecture

The Future Network Architecture needs to provide for a lot of various types of functionality and at the same time it is required to be able to deploy that functionality in different physical locations. Therefore, the architecture is separated into functional domains and topological domains.

Horizontal domains:

- "Transport" contains functionality for transmission and transport primarily between sites but also within sites

- "Cloud infrastructure, Data pipeline" contains functionality for secure processing and storage of both network functionality as well as application functionality. The data pipeline supports all network domains with collection, storage, distribution and processing of data.
- "Access - Mobility - Network applications" contains functionality securing all types of access as well as network integrated applications
- "Management, Orchestration, Monetization" contains functionality to manage and control the network as well as running the business management of customers to the network. Further it contains the exposure of network functionality to external applications
- "Applications" contains network external applications and is utilizing the network for communication, execution and storage

Vertical domains:

- "Devices / Local networks" – The actual device used by a user or a network set-up by a user or enterprise outside the control of the service providers
- "Access sites" – Local sites which are as close as possible to the users
- "Distributed sites" – Sites which are distributed for reasons of execution or transport efficiency or for local breakout
- "National sites" – National sites which are typically centralized within a service providers' network
- "Global sites" – Centralized sites which are publicly accessible from anywhere, typically a large data center

Listed below the picture are some important attributes that the network is providing:

- Sustainable – it is used for many, if not all, communication purposes with little or no damage to the environment
- Reliable & Resilient – it is always available
- Cloud native – the network itself is implemented being cloud native as well as supporting cloud native implementation of applications
- Secure – it supports the trust needed in being the communication platform for the entire society
- Data driven - data is generated and used to manage and orchestrate the network as well as supporting the applications with relevant information

Global connectivity and services have by tradition been deployed in a federated model, where the interfaces are well standardized and offered by one service providers. The complexity with multiple networks has been hidden through interoperability and inter service providers exchange models. However, the rapid deployment of new features makes the traditional standardized federated model hard to use. New methods of enabling exposure of assets from multiple networks is needed, like network asset facilitation and exchange or, on service providers request, aggregation into a single offer

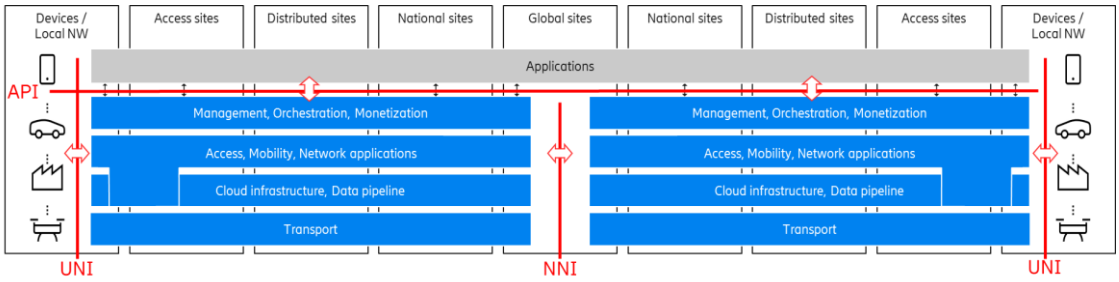


Figure 20 Network architecture business interfaces

The architecture supports various types of business interfaces:

- Service exposure of network capabilities to applications & a 2-sided business model irrespective if the application resides in or outside a service providers domain
- Network access and transport for devices and local networks
- Roaming and interconnect between service providers

## 4.1 Access, Mobility, Network applications

Evolved virtualization, network programmability, and 5G use cases will change everything about network design, from planning and construction through deployment. Network functions will no longer be located according to traditional vertical groupings in single network nodes but will instead be distributed to provide connectivity where it is needed. The design of the 4G/5G split architecture focuses on increased spectrum efficiency, full deployment flexibility, and elasticity; processing is carried out where resources are available and needed. The network can be deployed either in classical pre-integrated nodes or in fully virtualized environments as VNFs or any combination thereof.

### 4.1.1 Access

The external interfaces of the Radio Access Network (RAN) domain are standardized under 3GPP, as is the functional behavior of the RAN domain. Below the high-level specification, 3GPP leaves room for innovation to enhance the network with RAN-internal value-add features — a flexibility that has over many years resulted in continuous improvement in many areas, including spectrum efficiency, energy efficiency, and enhancements to service characteristics. To determine the optimal architectural split, however, the RAN architecture needs to be examined with a finer level of granularity than that offered by 3GPP. Based on function and interface characteristics, preferred execution environment, and spectrum efficiency, the target functional composition includes the following logical ran nodes,

The RAN consists of a set of functions; mobility, radio-link control, beamforming, scheduling, etc. realized by software executing in infrastructure located in various radio sites. The radio unit (RF) with its antenna system, the distributed unit (RPF) and the central unit (RCF/PPF) build up the RAN and may be placed

at different locations, connected through a transport network to each other and by that covering the functions of a RAN.

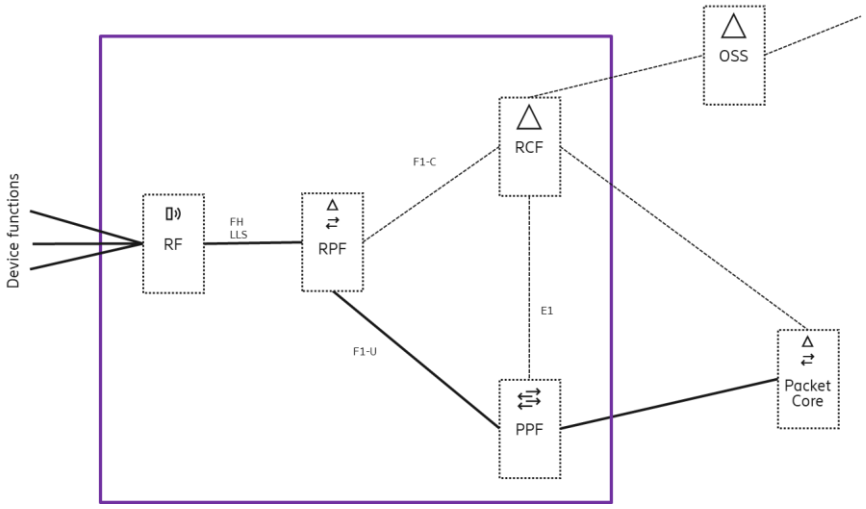


Figure 21 RAN Architecture

For RAN there are two horizontal interfaces relevant for the ability to geographically distribute functions across the RAN SW stack:

- The interface denoted F1 or Higher Layer Split (HLS) that separates the Distributed Unit from the Central Unit as specified in 3GPP.
- The Fronthaul (FH) or Lower Layer Split (LLS) between the RF and the RPF that separates the radio from the digital processing as partly specified in O-RAN Alliance.

From a functional perspective, cloud RAN is a system (of HW and SW) where the RAN functions can dynamically be allocated in time and location. This means that the SW realizing the function can appear where and when it is needed. In this way the functions can be dynamically allocated across different hardware and sites.

- Radio function — RF
  The RF requires special radio hardware and includes functions such as modulation, D/A conversion, filtering, and signal amplification. A beamforming function is introduced and may be co-located with either the RF or the RPF.

- Radio processing function — RPF
  Given the stringent requirements for spectrum efficiency, the RPF benefits from being placed on a special purpose processor. The RPF includes user-plane functions that are synchronous to the Hybrid automatic repeat request (HARQ) loop and it is also the anchor point for carrier aggregation and soft combining. The RPF contains the fast radio scheduler, and is also responsible for the coordinated multi point, for the selection of the MIMO scheme, and for beam and antenna elements.

- Packet processing function — PPF
  The PPF, which is suitable for virtualization, contains user-plane functions that are asynchronous to the HARQ loop, and includes the PDCP layer — such as encryption — and the multipath handling function for the dual connectivity anchor point and data scheduling.

- Radio control function — RCF
  The RCF handles load sharing among system areas and different radio technologies, as well as the use of policies to control the schedulers in the RPFs and PPFs. At the user and bearer level, the RCF negotiates QoS and other policies with other domains and is responsible for the associated service level agreement (SLA) enforcement in the RAN. The RCF controls the overall RAN performance relative to the service requirement, creates and manages analytics data, and is responsible for the RAN SON functions. Like the PPF, the RCF is suitable for virtualization.

In a deployment, each function (RF, RPF, PPF, and RCF) is instantiated. An instance of the radio functions will be associated with many antenna elements at an antenna site, and a set of RF instances are connected to one instance of the RPF. Each antenna element (RF) is associated with one RPF. Hence, an RPF instance handles the cells corresponding to the RF antenna elements it is associated with. Local mobility is hidden under the RPF and not visible to the PPF or the Packet Core.

Each instance of the RCF can handle a small or a large set of PPFs — and all the associated RPFs. In this way, the RCF can keep a holistic view of an area that is just a single cell, up to an area consisting of thousands of cells. With this architecture, RRM coordination and spectrum efficiency within a system area can be maximized, using the full suite of RRM features available.

### 4.1.2 Packet core

The core will exist in an environment that is cloud-based, applying cloud native design principles for scalability, dynamic orchestration of network resources, and a modular and highly resilient base architecture. Moving to 5G there will exist different migration paths, either based on Evolved Packet Core (EPC) or based on a newly defined core network architecture, 5G Core. EPC with functional additions to support the NR radio access though Multi-RAT Dual connectivity together with LTE provides a relatively light weight effort to migrate to 5G. 5G Core is built on a new architecture paradigm, Service Based architecture, with new reference points and services used between network functions in the control plane. While 5G Core has advantages over EPC, it is worth highlighting that it is a new architecture and network paradigm that shall be deployed in the service provider's network.  5G Core is for NR stand-alone deployments and interworking with EPC will remain important to maintain service continuity with EPC, especially in the initial phases and until NR is available on lower bands.

## EPC

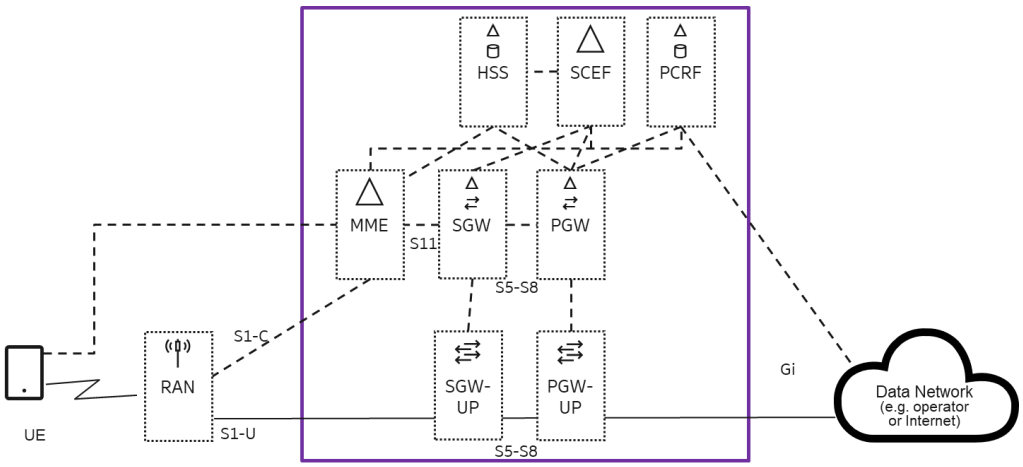


Figure 22 EPC architecture

The EPC consists of the following functions including sub-functions:

- Mobility Management Entity (MME)
  is the key control function in the EPC network. The main functionality of the MME is attach and detach of UE, authentication, choosing SGW and PGW for the UE, and management of PDN connections and EPC bearers. It also handles mobility procedures, UE tracking, and paging.
- PDN Gateway (PGW)
  is the gateway between the internal EPC network and external PDNs, for example, the Internet or a corporate LAN. The PGW provides IP connectivity towards external PDNs, policy and admission control, and packet filtering per user. The PGW can also be used for charging.
- Serving Gateway (SGW)
  routes and forwards the user packet data from the UE to the PGW or from the PGW to the UE. The SGW acts as a local mobility anchor for the user plane during inter-eNodeB handovers and provides charging functionality.
- Policy & Charging Rules Functions (PCRF)
  handles policy control decisions and flow-based charging control functionality. The main functionality is to evaluate operator policies that are triggered by events received from various functions and to provide rules for application and service data flow detection, gating, QoS and flow-based charging
- Home Subscriber Server (HSS)
  is a central database that contains user-related and subscription-related information. The functions of the HSS include functionalities such as storage of the long-term security credentials, subscriber profiles, access authorization, mobility management support
- Service Capability Exposure Function (SCEF)
  is used to securely expose the services and capabilities provided by 3GPP network interfaces. The functionality is brought to 3GPP for standardization through a function called Application Network Interaction Function (ANIF).
- User Plane (UP) includes e.g. functionality for mobility anchor point, external PDU session point of interconnect, packet routing & forwarding, QoS

handling for user plane, packet inspection and policy enforcement lawful intercept

### 5G Core

The key principles of the 5G Core architecture are as follows:

- A flexible and extendable service-based network architecture
- Allows for different core network configurations in different network slices and allow for resource isolation between network slices
- Allow for a user equipment to be simultaneously connected to multiple network slices (more advanced than multiple APN)
- Support subscriber identification and authentication based on IMSI as well as non-IMSI identities a unified EAP based authentication framework
- Separation of Control plane (CP) and User plane (UP) to:
  - Allow scalability of UP and CP functions independently
  - Allow for a flexible deployment of UP separate from the CP, i.e. central location or distributed (remote) location (i.e. with no restriction in the location compared to the CP).
- Support a generic user-plane function that enables both centralized and distributed deployments in a network, and the possibility to have different instances of the UP function centralized and distributed at the same time
- Unified Policy framework with extensions from the policy framework in EPC
- Abstract the transport domain from 3GPP network functions to allow for independent evolution and to enable service providers to use different transport technologies (e.g. Ethernet, MPLS, SDN-based transport, etc.).
- Support wireline and wireless access convergence. An architecture supporting both legacy and 5G Core capable residential gateway is needed. User planes support multiple accesses and is optimized for high throughput

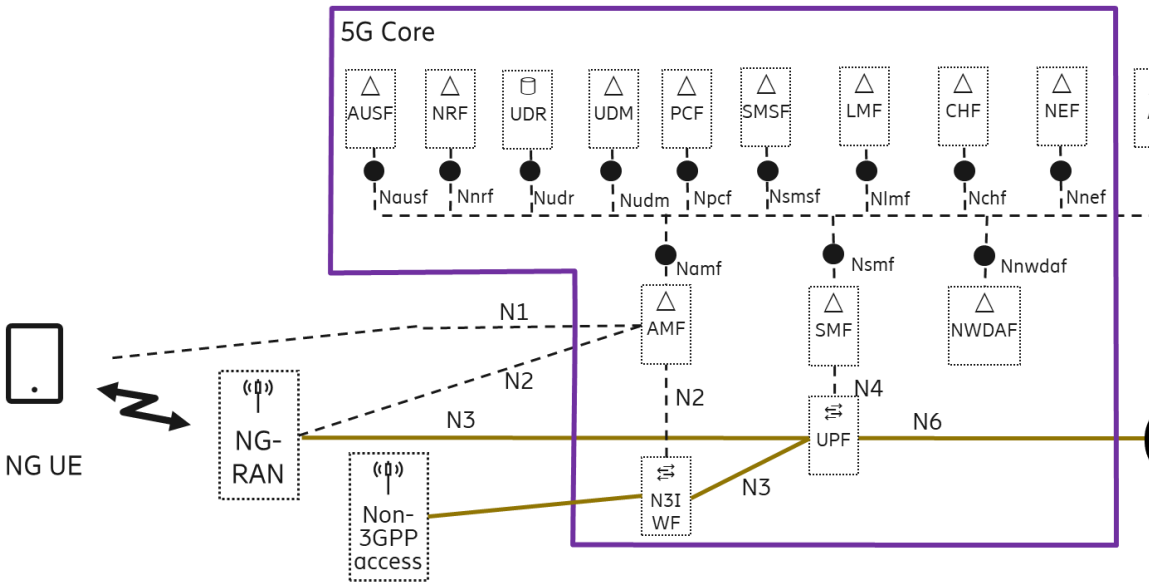


Figure 23 5G Core architecture

The 5G Core consists of the following network functions:

- Access & Mobility management Function (AMF) includes e.g. the following functionality:
  - Termination of RAN CP interface (N2)
  - Access Authentication and Authorization support
  - Mobility management
  - AMF selection for handovers with AMF change
  - Handling of policies for mobility management
  - Lawful intercept (for MM events and interface to LI System)
- Application Function (AF)
- Authentication Server Function (AUSF) includes e.g. the following functionality:
  - Interacts with the UDM for retrieval of the corresponding security credentials for the user.
  - Terminates the request to select and trigger the execution of the authentication of the user.
- Charging Handling Function (CHF) allows charging services to be offered to authorized network functions
- Location Management Function (LMF) supports location determination for a UE
- Network Data Analytics Function (NDAF) provides network analytics information to a network function on a network slice instance level
- Network Exposure Function (NEF)
  provides a means to securely expose the network services and capabilities, Similar functionality as ANIF/SCEF.
- Network Repository Function (NRF)
  includes functionality to find network functions and services to support e.g. the establishment of a PDU Session. Network function service discovery enables network functions to discover instance(s) that provide the expected service(s).
- Policy Control Function (PCF) includes e.g. the following functionality:
  - Supports unified policy framework to govern network behavior.
  - Evaluates operator policies that are triggered by events received from various functions.
  - Provides rules for application and service data flow detection, gating, QoS and flow-based charging.
- Session Management Function (SMF) includes e.g. the following functionality:
  - Selection and control of user plane function
  - Session management including session authorization
  - UE IP address allocation & management
  - Policy & Charging rules handling
  - Lawful intercept
  - Roaming functionality
- SMS Function (SMSF) activates/deactivates SMS service for a service user and send SMS payload
- Unified Data Repository (UDR) is a repository of subscriber information and can be used to service several network functions

- Unified Data Management (UDM) is an evolution of the HSS/AuC and includes e.g. the following functions:
  - Storage of the long-term security credentials.
  - Information storage of subscriber profiles and related management.
  - Access authorization.
  - Location/mobility management support.
- User Plane Function (UPF) includes e.g. the following functionality:
  - Anchor point for mobility
  - External PDU session point of interconnect
  - Packet routing & forwarding
  - QoS handling for user plane
  - Packet inspection and policy enforcement
  - Lawful intercept
  - Traffic accounting and reporting
- Non-3GPP Interworking Function (N3IWF) is used to connect e.g. untrusted WLAN

### 4.1.3 Communication services

Communication services has been a fundamental part of telecom networks since the networks appeared. Most networks were originally created for a single service, voice telephony. Communication services can be divided into two categories;

- Communication services that are interoperable between communication service providers / operators and supported in most devices and networks.
- Communication services provided by a single service provider with co-designed client/devices.

Interoperable communication services of today must be backward compatible to the legacy and support interoperability to older generations of the services still in use.

The fundamental service definitions and network architecture for interoperable communication services such as telephony, messaging, etc. are already in place. The services and their IMS based architecture will be reused for all 3GPP access also in the 5G era, in order to provide services with Quality of Service (QoS) also in challenging radio and mobility conditions. This means that the 5G Core and NG RAN must include similar capabilities as in 4G VoLTE to ensure voice KPIs. In addition, seamless mobility between EPC and 5G Core meeting voice KPI's are introduced to support migration, since extremely few networks will have full NR SA coverage the day 5G core is introduced.

New communication service opportunities will be enabled by 5G engagements into non-traditional segments/customers and their communication need. In most cases communication services are not the focus of the initial engagements. Communication services are however expected either as a utility, or as a value add leveraging the initial investments and deployments. It is in these engagements and new segments we expect new innovative services and communication means like XR, holographic conferencing to emerge. The

offerings and the supporting architecture must fit the emerging business models and more complex value chains which will arise in these new segments. The network architectural impacts come from both the service and its delivery model as well as the need for flexibility and automation.

The IMS architecture is still being evolved in 3GPP to support service-based interfaces to the 5G packet core (5G Core). IMS will be able to use 5G Core via new service-based interfaces (SBI) as an alternative to legacy Diameter interfaces.

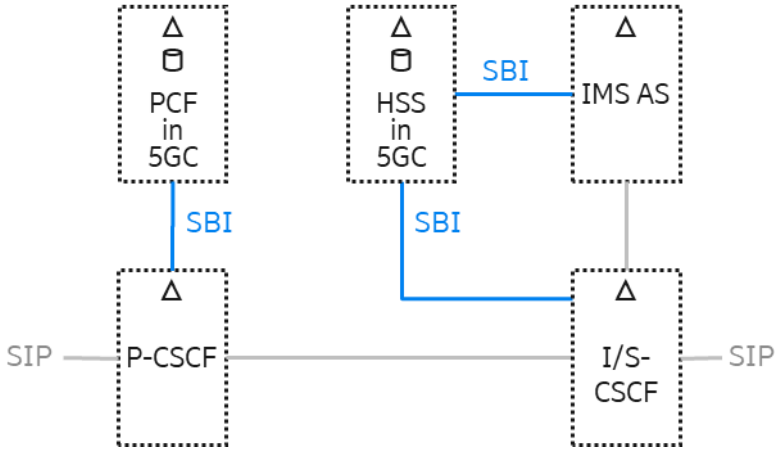


Figure 24 New service-based interfaces between IMS and 5G Core

Evolved management architectures in cloud environments will fundamentally change how IMS applications are deployed and managed. This does not mean that the traffic view of IMS network architecture must or will change. The cloud mechanisms and functions are complementary to the mechanisms and traffic functions in the IMS Network architecture. The network scaling functionality in IMS is required for geographic redundancy purposes.

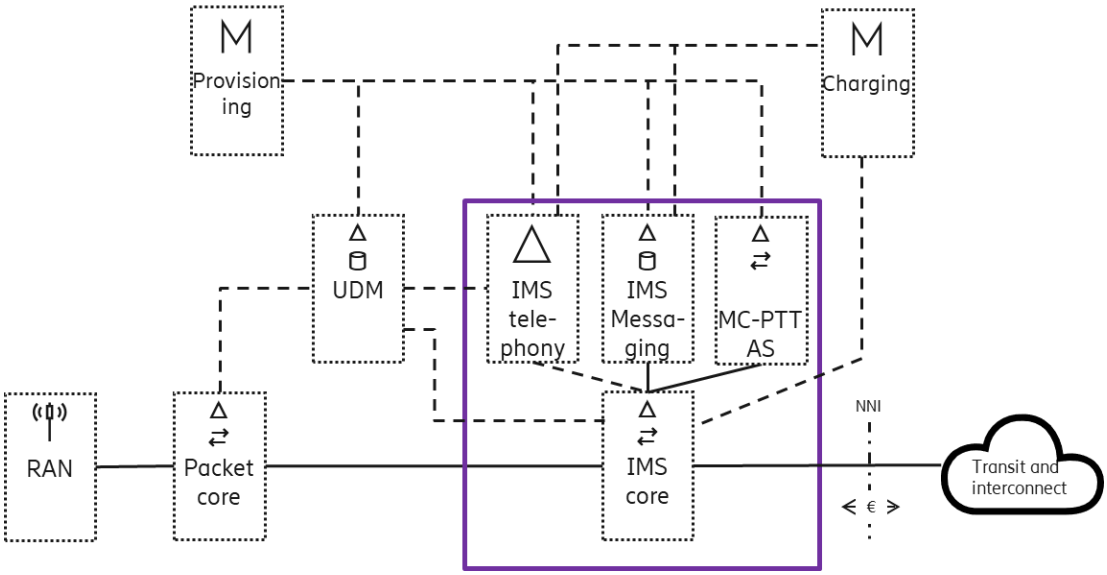


Figure 25 Communication services functions

The Figure 25 above is a conceptual illustration of the heart of the Communication Services architecture; the IMS parts. The IMS architecture supports IP communication services over any access technology (e.g. NR, LTE, WiFi and Fixed) and interworks with 2G/3G networks. It consists of an IMS core providing a multitude of functions such as: SIP session handling (e.g. registration, authentication, routing and service invocation), emergency calls, Interconnect and Roaming, NNI enforcement, charging, accounting, number portability, restoration procedures. It also consists of IMS application functions for, IMS telephony and IMS Messaging. Communication service evolution may enrich existing application functions or introduce new ones. An example of a new application function is Mission Critical Push-To-Talk (MC-PTT) which recently has been standardized in 3GPP.

## 4.2 Cloud infrastructure

The cloud infrastructure ensures the robustness, performance, security and interoperability needed for modern applications. This requires a system that beside traditional cloud capabilities provide telco apps with a real-time network support, while providing new sets of enablers that act as a bridge to cloud-empowered telco applications and solutions.

Application scalability and the introduction of new technologies are both facilitated by the independent life cycles of application components and the services they use. Container as a Service (CaaS) increases the portability of the applications to several IaaS solutions, and thus helps reduce the number of cloud execution platforms that need to be supported.

The cloud native application is composed from set of independent services each with different capabilities and/or functionalities. Services are grouped into functional areas, that is, each functional area offers a set of different building blocks to be used by application architects. A functional area is defined for application services with unique capabilities which are characterizing that application.

An application, with its different instances, can be deployed into a single container execution environment. It is managed by a single container orchestrator entity. The APIs is used both by network applications and by network external applications.

To make cloud ready to support telco and mission critical applications, availability requirements are to be met as well as several others:

- Automation of management operations
- Monitoring support
- Logging support
- Security
- Tenant isolation
- Upgrade support
- Backup and restore
- Quick restart time

- Independent restart
- Network protocol support
- Alarms
- Performance counters
- Trace support
- Soft real-time

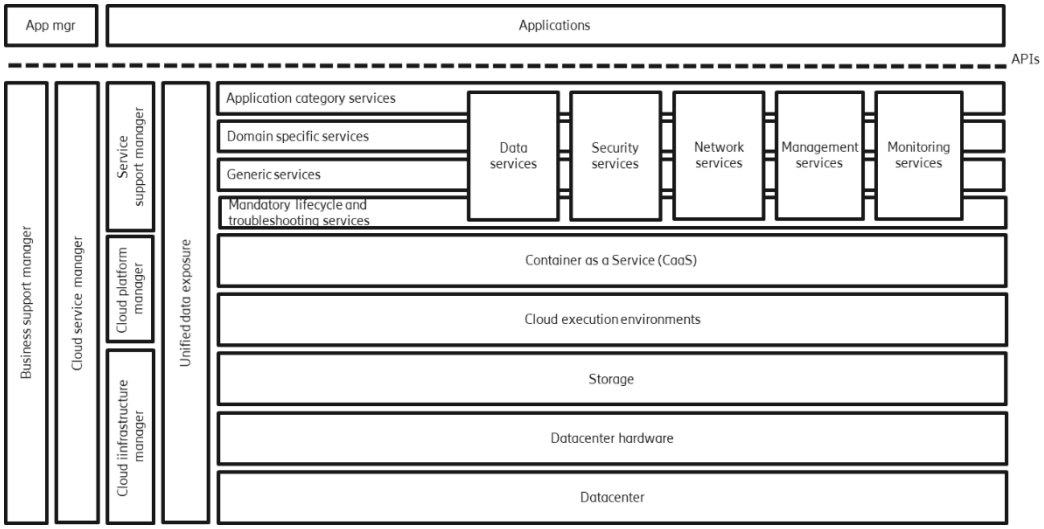


Figure 26 Cloud architecture

Cloud execution environment provides basic cloud services typically provided by an IaaS environment or bare metal.

CaaS contains functionality that is common to all services running on top of the CaaS e.g. policy management, container orchestration, networking, container runtime, deployment support etc.

Generic services contain services that are common to all type of applications, also many cases for the rest of the industry e.g. data management, key/value store, relational data service, data warehouse, malware protection, firewall etc. As part of the generic services we can find specialized platform services, providing very specialized services for horizontal use cases that would span across several application categories.

Core services contains services that are common to all type of services (whether they are application services or belongs to any of the layers in the platform). Examples are trace, log, alarm, performance management, backup & restore, license management etc.

Domain Specific Services contains services that are common to a specific domain, for example IoT, media, OSS/BSS.

Application Category Services contains services that are common to an application category within a vertical.

## 4.3 Management, Orchestration, Monetization

Management of the virtualized environment and new services becomes even more important as all services has a need to be managed in real-time. This dynamic and competitive environment requires management of the networks and their supporting systems to be:

- Less expensive to manage and maintain
- Self-provisioned to drive down costs in an instant-access, cloud, and application-driven world
- Flexible and modular to support "network slices" and "micro services" for new use cases driven by market needs
- Deployable at speed to roll out new services with "zero touch" fulfillment
- Scalable, with an agile IT operational model.
- Bridge the physical and virtual environment

The virtualization/NFV promises to bring cost efficiencies, time-to-market improvements and innovation to the telecommunication industry infrastructure and applications. The new environment will achieve these functionalities through disaggregation of the traditional roles and technology involved in telecommunications applications.

The architecture will support creation of a management system that will provide an easy adaptation of business processes to the ever-changing business landscape which allow for a fast introduction of new products and services helping the service providers to keep be ahead of competition.

The architecture will enable a highly automated operation striving towards zero touch operations. This will lead to drastically reduced OPEX for the service providers.

There is also a drive for openness to with the notion of reducing cost for integration and lower barrier for innovation. From an overall level, there is a clear shift away from standardization, or more specifically standardization only, as the forms of industry alignment. This shift towards an open source approach is visible in three ways:

- Service providers are working more with their own architectures (alone or with selected traditional and non-traditional vendors) and the large service providers are making a subset of their architecture public to influence the community.
- There is a continued increase in open source initiatives in orchestration and management. ONAP being the most prominent but also others like Open Source Mano have impact. Some of the open source initiatives align with standards (or promote certain standards) though this is not the case with all.
- Service providers are producing their own operational support systems using both in-house components and open source components.

The following are the founding principles based on which future reference architecture will be built for an environment possible for real time management of the ecosystem:

1 An open and transparent architecture
2 Micro service-based architecture
3 Data centric applications – where the application logic should have control on how, what, and when the data can be exposed
4 Support and expose open APIs to allow easy access to management data
5 SDK (Development runtime, tooling, documentation, and reference applications) where applicable
6 Repository of reusable, deployable, and runnable applications. (that can be further integrated and/or extended)
7 To a high degree based on open source technologies
8 Platform agnostic execution environment

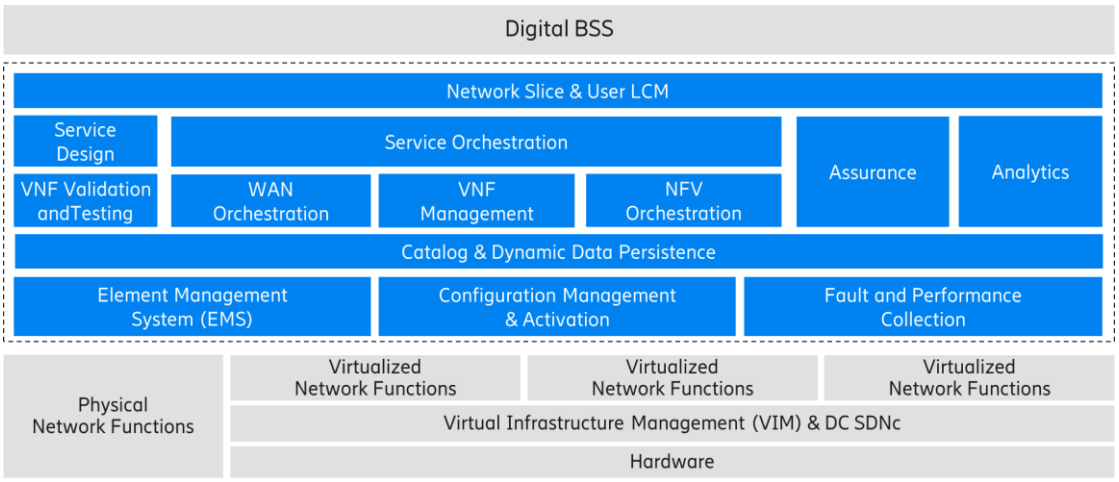


Figure 27 OSS/BSS Architecture

The architecture needs to embrace automation and the movement towards autonomic networks to tackle the complexity increase brought by the technology innovation. This reflects both the possibilities enabled by technology as well as need of service providers to introduce new offerings to the market quickly and reduce their cost of operation. Automation and network governance will become essential for avoiding this spiral of complexity. This automation must be combined with machine learning and in the longer term also artificial intelligence algorithms applied to life cycle management operations, with domain policies to facilitate and improve speed in development of automation functionality.

The domain management functions serve the underlying functional domains, from the basic management point of view, they take a role of element management functions and more advanced domain specific OSS functions

Network Northbound focuses on community innovation; service development across domains via component assembly and flexible business and commercial models to expose the network capabilities to the applications that are used by external developers, consumers and enterprises. Network northbound will

provide an integrated environment where service creation ("the real-time call flow") and management workflow creation (OSS/BSS support) come together.

The network capabilities are made available to the application developers via defined APIs e.g. basic connectivity management services, transport optimization services, identity services, security services etc. There is potentially a need for brokering between the application providers and network service providers. For example:

- The cloud execution can utilize the cloud infrastructure provided by a network service provider, a public cloud provider or be a private cloud.
- The facilitation has the potential to reduce complexity and fragmentation when exposing network assets from many service providers.
- The aggregation aggregate capabilities from many networks and address the many-to-many problem between a network service provider and application providers.

The developer shall not need to consider the integration of each network service provider. Which in practice means that interfaces to multiple network service providers shall be consolidated.

See also [11] Open, intelligent and model-driven: evolving OSS,[12] Architecture evolution for automation and network programmability.

## 4.4 Transport

The latest demands on the transport network come from areas such as increasing RAN and mobile broadband service capacity, new 5G-enabled services and the dynamic deployment flexibility of the 5G RAN split architecture, with its tight transport characteristics. These characteristics are especially manifested in the fronthaul portion of RAN transport where the latency and synchronization requirements are very challenging. Enhanced automation capabilities in the operations and management domain represent a key requirement to meet these challenges.

The strong drive towards virtualization of network functions will have a clear impact on transport flexibility. As soon as a VNF moves, the transport must immediately be reconfigured to support the new topology. The major challenges for transport are programmability, flexibility, and finding the right balance of packet and optical technologies to provide the demanded capacity. This will be a major OPEX driver in the transport network unless it is fully automated. This is especially true in the mobile-centric part of the network.

To achieve this software-defined networking (SDN) will be an important entity that offers a programmatic interface towards the higher layers of transport control. SDN can be used along with an intelligent application, the transport intelligent function (TIF), to design an optimal 5G transport network architecture.

The optimal 5G transport network is built as a self-contained infrastructure underlay with an SDN-controlled overlay for a variety of RAN and user services.

The distributed control plane in the underlay maintains the basic infrastructure and handles redundancy and quick restoration in case of network failures. The service and characteristics aware overlay are handled by the SDN controller with the TIF application, and this creates a dynamically controlled and orchestrated transport network that requires minimum manual interaction. In other words, the underlay network described here handles the infrastructure connectivity and the network overlay handles the services running on top of the underlay.

During a long transition time there will exist both SDN controlled equipment as well as legacy transport equipment, so the higher layers transport control will have to operate with a multitude of protocols. The figure below exemplifies this by illustrating how transport setup can optimized by using information from both RAN and the transport network and using different legacy protocols when reconfiguring the network.

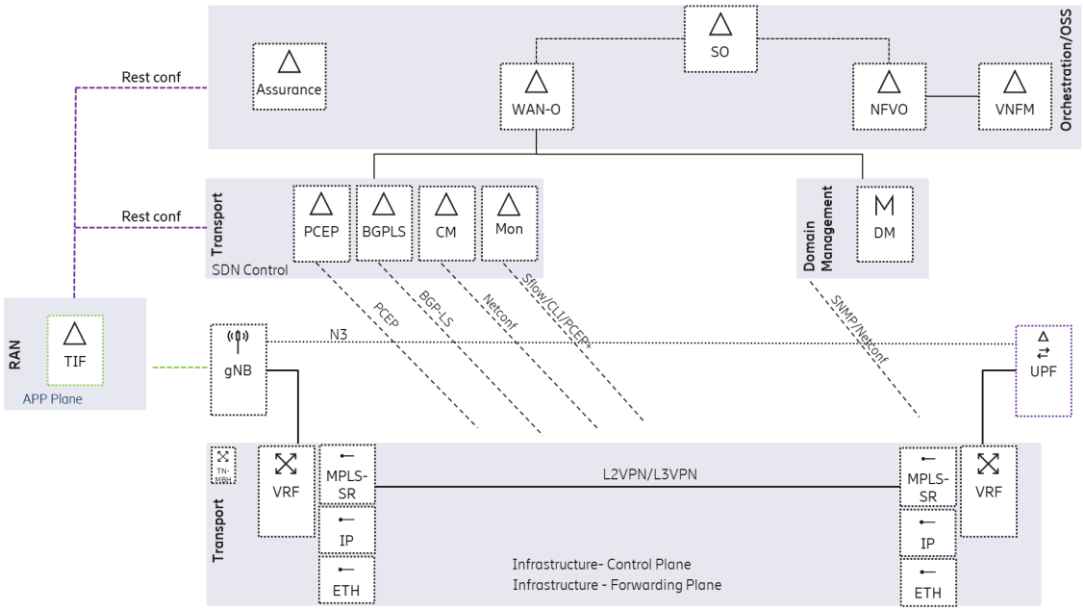


Figure 28: Interaction between RAN and transport

There is a need for Analytics, Policy, Control functions and specific Northbound APIs for service orchestration and SLA reporting. Collection, presentation and correlation of various Transport service level parameters on a per service basis is one of the key capabilities required. Also, necessary level of isolation and prioritization between the various network services needs to be maintained.

The main propositions of the future transport architecture can be summarized as:

- Service agility, programmability, enhanced visibility & cross domain orchestration
- There are 4 major components for this layer to support the mentioned characteristics: Configuration Management, Path Management, Topology Management and Utilization management
- Specific applications for both local Transport domain optimization but also to support cross-domain orchestration functions will be needed.

- The infrastructure layer is built using a stacked underlay/overlay architecture. A service-based overlay providing L2 or L3 VPN service endpoints with MPLS-SR TE on top of a fully routed underlay topology based on IP and IS-IS. IPv6 based forwarding plane is also one of the industry directions for the future
- All management functions that can be centralized will be centralized and only the most needed functions will stay in the infrastructure layer like OAM monitoring and fast reroute solutions.
- Ubiquitous transport service e.g. using Ethernet VPNs (EVPNs) from all the way from the access to termination in a datacenter and between datacenters is envisaged

# 5 Network deployment examples

## 5.1 Wide area public network

Delivering the full 5G experience will involve enhancing many existing use cases and creating new ones that cannot be fulfilled using current technologies, through for example:

- Deploying 5G NR radio access to deliver capabilities far beyond those of previous cellular generations, including massive system capacity, very high data rates, very low latency, ultra-high reliability and availability, low energy consumption, and energy efficient networks – wherever and whenever needed
- Deploying 5G Core for introduction of cloud-native deployment and operational model including automation of network and service management
- Distributed Cloud, to enable workloads to be placed closer to the network edge for better QoS (shorter latency), transport efficiency and higher integrity of data

As 5G will need to coexist and interwork with 4G (and 2G/3G) for many years to come, we're likely to see most of these deployments as non-stand-alone initially, as a way of reducing time to market and ensuring good coverage and mobility. Important aspects for 5G migration include

- Efficient spectrum utilization using spectrum sharing between LTE and NR
- NR carrier aggregation for optimal performance
- Tight interworking between 5G Core and EPC with common functionality for devices connecting over both NR and LTE access.

In order to achieve the best time to market and maximize the reuse of existing network resources, a two-phase approach is needed to deploy 5G to ensure that the full potential of 5G can be achieved as quickly as needed:

1 Start with non-standalone NR (NSA NR, aka UE connectivity option 3). This option uses LTE/EPC as the control plane anchor and uses either LTE or NR or both for user traffic (user plane). This approach is based on existing LTE/EPC and provides time to market benefits for introduction of NR.

Operators can deploy dual connectivity for data (high throughput in NR downlink, best coverage in LTE uplink), while voice traffic is fully on LTE.

2 One of the main drivers for going beyond option 3 is to provide 5G Core-enabled capabilities like enhanced network slicing, edge computing support and operational benefits, even though EPC can also support these services to some extent. Another main driver for going beyond option 3 is to be able to deploy standalone NR and get the radio performance benefits of an NR-only based radio interface. Option 2 (standalone NR) is the 5G Core-based option available in UEs and networks.
Even if general NR coverage may initially be limited, option 2 can initially be deployed for specific use cases in local areas, where devices stay within good NR coverage on a mid or high band. Examples include industrial deployments with ultra-reliable low latency communication requirements, and fixed wireless access (FWA).

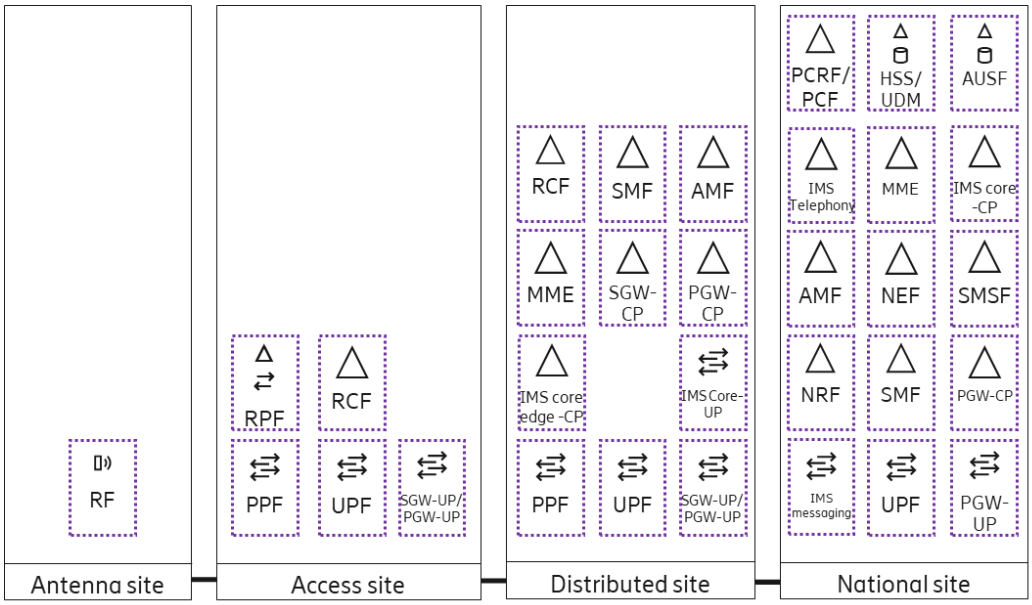


Figure 29 Potential variation in distribution of functions

An operator must consider which functions shall be co-sited, and which functions might be better placed at other sites (separate, or co-sited with other network functions). There are several factors that affect which will be the best solution for each operator:

- Stay with the classic Distributed RAN (DRAN) architecture also when deploying NR. This will provide good performance and has minimum modification of existing network architecture. This architecture is suitable for large parts of an operator's network.
- Evolve the transport network to support very high user data rates. Dark fiber will in many cases be the most cost-efficient alternative, why deploying a Centralized RAN (CRAN) architecture may be beneficial. Centralized RAN is suitable in urban and in some suburban cases where the site distances are not too large. The benefits are improved coordination between sites with Carrier Aggregation, as well as pooling of baseband hardware.

- Consider virtualized RAN (higher-layer-split of CU and DU) deployments added on top of both DRAN and CRAN. However, it is only recommended when cloud infrastructure is already deployed to reuse the cloud investment
- As much centralization as possible for lowering OPEX based on economy of scale. However, distribution of user plane functionality will be needed for supporting distributed cloud.
- The RCF would be more centrally placed depending on availability of fiber as well as other benefits of centralizing the baseband unit. The placement of the PPF will be dependent on the placement of UPF.
- Deployment of voice related network functions is a compromise based on many factors, for example speech path delay, fast call set-up time, short service interruption time at mobility.

See also [13] Communications as a cloud service: a new take on telecoms, [14] 4G/5G RAN architecture: how a split can make the difference,, [15] A vision of the 5G core: flexibility for new business opportunities, [16] Enabling intelligent transport in 5G networks, [23] Communication services over LTE, Wi-Fi and 5G, [20] Simplifying the 5G ecosystem by reducing architecture options, [24] 5G deployment considerations, [27] 5G New Radio RAN and transport choices that minimize TCO.

## 5.2 Private networks

A private LTE/5G network is comprised of 3GPP products but used within an industry (enterprise) for its own data and personal communication needs. There is no public access, only authorized devices/users can use the network resources. In 3GPP terminology a private network is called a "non-public network".

The standard set of functions for voice and data shall be supported by the network infrastructure to leverage the device eco-system. Special interest groups such as 5G Alliance for Connected Industries and Automation (5G- ACIA) develop the eco-system for industries. In many cases business and mission critical capabilities for survivability are combined with specific communication services such as push-to-talk/video.

Deployment cases range from national public safety networks to small campus systems.

- Wide-area private networks covering a region, or a country are similar to regular public networks in how they are deployed: Examples include public safety, mainline rail and utilities.
- Local networks for manufacturing and public/enterprise venues, hospitals, airports, campus, range in 1000s within a country. The network architecture evolution follows the architecture evolution in 3GPP, but deployment size is much smaller in terms of area covered and number of supported devices. Deployment options are needed ranging from local standalone private networks to different levels of integration to external networks, e.g. with and without roaming/interconnect to other PLMNs and internet.

Private networks have stringent technical requirements (latency, reliability, availability, local mobility, positioning etc.) and additional operational requirements: local O&M, local data, local survivability, integration to existing IT/OT (Information Technology/Operational Technology) systems.

The Private network is a system comprised of RAN (E-UTRAN, NG-RAN), Core (Combined Evolved Packed Core – EPC and 5G Core – IMS/VoLTE, Positioning and Mission Critical Push-to-Talk (MC-PTT). Management is done from an Operational Support System – OSS, typically with management functions for local monitoring and self-service by the enterprises. Business Support System (BSS) may be included but it shall be noted that for local stand-alone system charging is typically not included. Various device types from ruggedized phones to IoT gateways (modems) and Nb-IoT/Cat-M devices etc. can be connected via the private network.

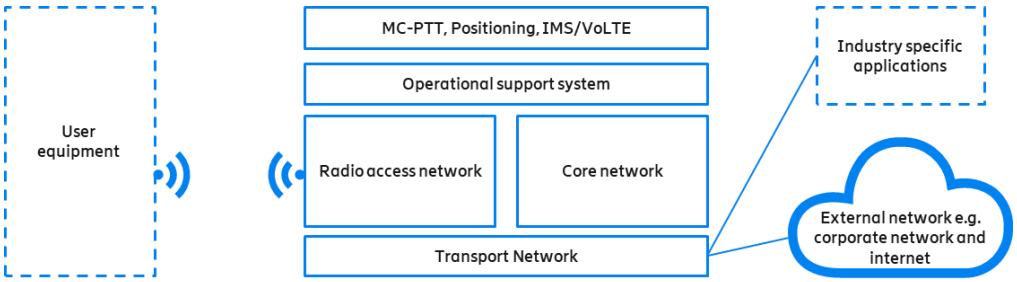


Figure 30 Main building blocks in private network

Deployment vary not only between industry segments, but there are also very different scenarios within each segment. For example, the utility segment ranges from wide-area electricity grid network and meters (electricity, gas, water) to local power generation plants: hydro, wind, nuclear. Resilience and survivability are critical aspects and put demands on redundancy schemes (core, radio) and on the management solution.

- Regional or wide area coverage for non-public use (public safety, mainline rail, utilities):
  A wide-area solution as for regular public network but typically focus on specific (3GPP) communication services such as MC-PTT. High-availability /resilience including power-backup are required. Typically deployed in low-band spectrum.
- Local standalone deployment for the needs of the OT-domain.
  All functionality needed for the specific OT use case is deployed locally to fulfill both technical and non-technical requirements including legal and commercial constraints. Local area spectrum allocation (industry spectrum), indoor coverage and capacity are in focus. The high-availability / ultra-reliable low latency requirements for demanding use-cases will start in local stand-alone deployments.
- Central or shared functions for cases with less stringent requirements on e.g. local survivability and local data. In such cases functionality may be deployed outside the customer premises site at either communication service provider or industry player network sites. This may also happen if industries use cloud-based solutions for data analytics etc.

The figure below shows an example for the case where parts of the functions in the private network are centrally located. The IT domain uses the network infrastructure deployed for the OT domain. In this example, only the RAN infrastructure is shared between IT and OT domains, and all other functionality for the IT domain is deployed outside customer premises. OT communication is isolated entirely from the IT and Internet via a DMZ (De-Militarized Zone) and any interactions towards the OT layer are limited to a very controlled set of data over secure interfaces terminated at the DMZ.

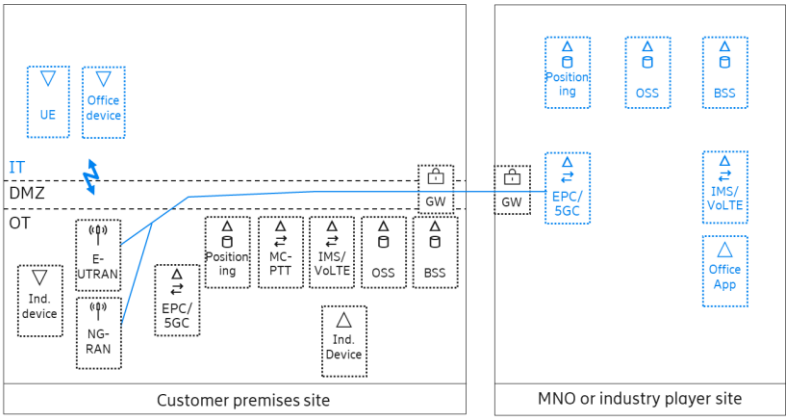


Figure 31 Part of the functions in the private network can be deployed centrally.

See also [17] Boosting smart manufacturing with 5G wireless connectivity, [18] 5G network programmability for mission-critical applications.

## 5.3 Automotive and road transport networks

Connected vehicles and road infrastructure are part of a broader IoT ecosystem that is continuously evolving. To ensure cost efficiency and future-proof support, Communication service providers aim to meet the connectivity demands of multiple industry verticals, including the automotive and transport industry, using common physical network infrastructure, network features and spectrum resources. Cellular connectivity for the automotive and transport services is relevant from four perspectives: massive IoT, broadband IoT, critical IoT and industrial automation IoT.

Figure 32 illustrates how cellular connectivity works for vehicles and roadside equipment. It visualizes vehicles as multipurpose devices in which several connectivity-dependent use cases are executed simultaneously. At the same time, each vehicle also contains an internal network that interconnects in-vehicle sensors, actuators and other devices, including driver and passenger smartphones.

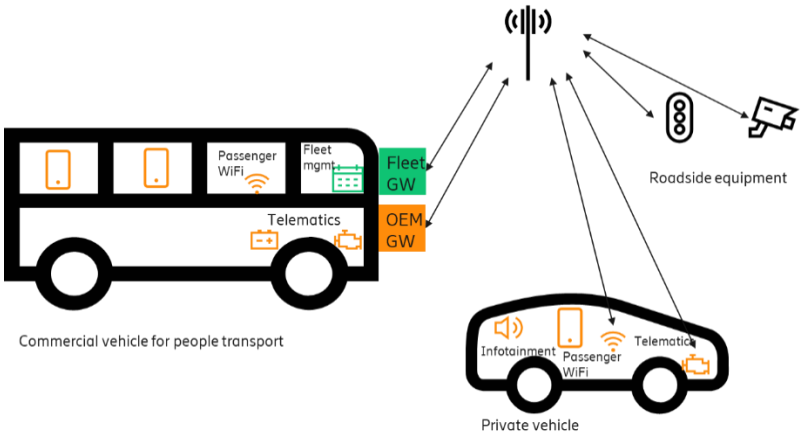


Figure 32 Cellular connectivity for vehicles and roadside equipment

Millions of cars are already connected using 4G cellular access, and cellular broadband IoT connectivity (4G/5G) is expected to grow significantly through 2024 as outlined in [25] Ericsson Mobility Report.. Connected vehicles incorporate applications such as fleet management, in-vehicle entertainment, internet access, roadside assistance, vehicle diagnostics, navigation and advanced driver assistance services. The Automotive Edge Computing Consortium (AECC) estimates that the related data traffic has the potential to exceed 10 exabytes per month by 2025, a volume 1,000 times larger than the present numbers.

Most of the data traffic of connected vehicle services have relaxed latency requirements, which can be leveraged by the service provider for more-efficient usage of existing network resources with minimal cost and interference to existing services. The transfer of such data will be opportunistic, which implies it will be delivered in the background of normal data transfer without seizing network resources from normal traffic.

The policy rules for traffic separation can be provided either statically or dynamically using the Service Capability Exposure Function (SCEF), which is provided by the mobile network towards the OEM. The SCEF is evolving into the Network Exposure Function (NEF) in 5G Core.

The more dynamic approach enables a vehicle application to trigger what policies to be applied by the PGW/UPF, such as requesting a policy with lower priority (lower than best effort) with lower tariffs. By that different policies can be applied for different applications running in the same UE.

To alleviate the pressure of high volume of data, the Cellular Network (both EPS and 5G Core) will support data offloading to designated edge servers. The edge servers are typically connected to a center server as in a distributed computing architecture. The deployment of the edge servers shall be selected at appropriate places in the Cellular Network to meet the service requirements on latency, capacity and cost.

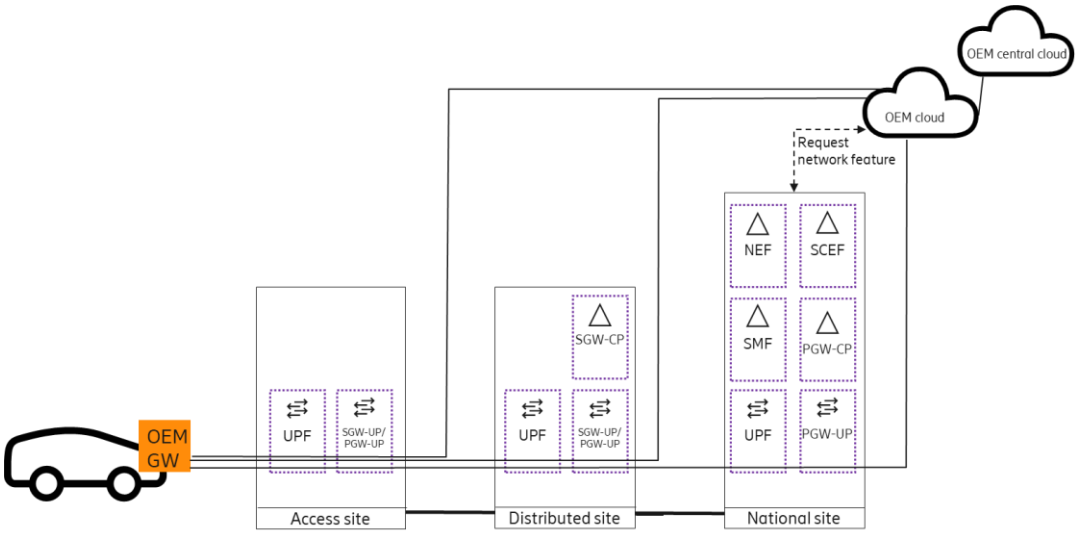


Figure 33 Deployment supporting vehicle OEM cellular subscription

Figure 33 depicts an end-to-end architecture using dedicated bearers for traffic separation, considering distributed computing with edge clouds. The edge cloud servers are shielding the central cloud servers by executing the heavy lifting workloads. The central servers coordinate the heavy workload functions and distribute the load across different edge cloud servers and sites. The central cloud servers steer the vehicle's connection to an appropriate edge, which supports the service and has sufficient computational capacity.

To alleviate the pressure of high volume of data, the Cellular Network (both EPS and 5G Core) will support data offloading to designated edge servers. The edge servers are typically connected to a center server as in a distributed computing architecture reference model. The deployment of the Edge servers shall be selected at appropriate places in the Cellular Network to meet the service requirements on latency, capacity and cost.

See also [22] Driving transformation in the automotive and road transport ecosystem with 5G, [19] Distributed cloud: A key enabler of automotive and industry 4.0 use cases.

# 6 Network evolution journey

The promise of increased agility, improved handling of data growth and lowering cost per bit (TCO), as well as the opportunity to exploit new business models are compelling reasons for service providers to invest in network evolution, leveraging advanced technologies such as 5G, NFV, SDN to name a few.

Notwithstanding the long-term potential that the programmable network platform of the future can deliver to the service providers business, the velocity of transformation will be dictated by the service providers business priorities, market situation and regulatory constraints.

The widespread introduction of advanced automation, toward zero touch systems, and platforms built on shared and public infrastructure will also drive significant changes in organizational design and ways of working. This presents an opportunity to streamline and improve existing business processes to fully exploit the benefits that new technologies can bring to the business. Whilst "non-technical" in nature, these impacts to the service providers business environment are non-trivial and represent a large and influential component of the network evolution journey.

The reality of needing to manage legacy infrastructure, whilst building the next generation network holds true for many service providers. Accordingly, the network evolution will take place in numerous phases of implementation and over many years, therefore requiring courage to stay on track with the journey, given that the longer-term benefits materialize after transformation has taken place in multiple domains.

The table below describes a plausible evolution path for the different architecture domains. The timings are indicative and may vary substantially depending on market maturity and in context with an individual service provider's business situation.

| Area | 2020-2021 | 2022-2023 | 2024-2025 |
|---|---|---|---|
| **Access & Network applications** | Continued 5G non-standalone deployments for eMBB (incl. voice). Initial 5G standalone deployments for eMBB, tight interworking between EPC and 5GC introduced from day one. Initial 5G standalone deployments also for local area private networks. | Increased amount of 5G standalone deployments for eMBB and local area private networks. Additional wide and local area use cases deployed. | 5G standalone deployments dominating. Additional wide and local area use cases deployed. |
| **Cloud Infrastructure** | Cloud infrastructure deployed in national sites. Initial deployments in regional site infrastructure in preparation for distributed cloud. | Cloud infrastructure deployed in access sites. Container based cloud infrastructure being adopted in most sites. | New technologies like function as a service (FaaS) evaluated for live deployments. VM deployments will start to sunset, Containers will be the reference deployment option of cloud infrastructure. Further improvements in platform characteristics delivering flexibility and resilience with optimized cost. |
| **Management, Orchestration, Monetization** | Further deployments for e2e network service orchestration & cross domain orchestration. Limited closed loop | Digital transformation ongoing. Increased analytics, AI & machine learning for | Zero-touch network taking full advantage of AI. Deployment and operation of networks and applications having |

| Area | 2020-2021 | 2022-2023 | 2024-2025 |
|---|---|---|---|
| | automation in daily operations. Starting to leverage deep learning systems. | closed loop automation in daily operations. | minimum human intervention, reaching high performance and zero downtime. |
| **Transport** | 25GbE interfaces in RAN transport start getting deployed. 100GbE optical transponders in long-distance and metro systems. Initial 400GbE capable systems deployed in data centers. Initial deployments of RAN transport interaction. Underlay/overlay transport models. | Initial deployments of more flexible fronthaul based on eCPRI and SDN mobile backhaul. Increased amount of DRAN sites with >1 Gbps backhaul. Possibly 800GbE capable systems deployed in data centers. Automation in transport networks. | 50GbE and possibly 100GbE interfaces in RAN transport start getting deployed. SDN controlled mobile backhaul and fronthaul. Cross-domain orchestration. |

Table 1 Network Evolution Journey - 2020-2025

# 7 Additional reading

Below you find links to some relevant additional information in Ericsson Technology Review articles, White papers and additional sources. This is in addition to the documents referenced in the chapters above:

The central office of the ICT era: agile, smart and autonomous
https://www.ericsson.com/thecompany/our_publications/ericsson_technology_review/archive/central-office-of-the-ict-era

Gearing up support systems for software defined and virtualized networks
https://www.ericsson.com/news/150605-gearing-up-support-systems_244069646_c

Radio access and transport network interaction – a concept for improving QoE and resource utilization
https://www.ericsson.com/news/150703-er-radio-access-and-transport-network-interaction_244069645_c

Make your mobile network ready for 5G voice
https://www.ericsson.com/en/digital-services/trending/5g-voice-evolution-where-to-start

Paving the way to telco-grade PaaS
https://www.ericsson.com/en/ericsson-technology-review/archive/2016/paving-the-way-to-telco-grade-paas

5G security - enabling a trustworthy 5G system
https://www.ericsson.com/en/white-papers/5g-security---enabling-a-trustworthy-5g-system

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

# 8 References


[1] Network Slicing can be a piece of cake
http://pages.digitalservices.ericsson.com/paper-network-slicing-can-be-a-piece-of-cake

[2] Flexibility in 5G transport networks: the key to meeting the demand for connectivity
https://www.ericsson.com/en/ericsson-technology-review/archive/2015/flexibility-in-5g-transport-networks-the-key-to-meeting-the-demand-for-connectivity

[3] End-to-end Security Management for the IoT
https://www.ericsson.com/en/publications/ericsson-technology-review/archive/2017/end-to-end-security-management-for-the-iot

[4] Signaling security
https://www.ericsson.com/en/white-papers/signaling-security

[5] 5G security - scenarios and solutions
https://www.ericsson.com/en/white-papers/5g-security-scenarios-and-solutions

[6] Virtualizing network services - the telecom cloud
https://www.ericsson.com/en/ericsson-technology-review/archive/2014/virtualizing-network-services---the-telecom-cloud

[7] Service exposure: a critical capability in a 5G world
https://www.ericsson.com/en/ericsson-technology-review/archive/2019/service-exposure-a-critical-capability-in-a-5g-world

[8] 5G network programmability for mission-critical applications
https://www.ericsson.com/en/ericsson-technology-review/archive/2018/5g-network-programmability-for-mission-critical-applications

[9] Artificial intelligence and machine learning in next-generation systems
https://www.ericsson.com/en/white-papers/machine-intelligence

[10] Cognitive technologies in network and business automation
https://www.ericsson.com/en/ericsson-technology-

review/archive/2018/cognitive-technologies-in-network-and-business-automation

[11] Open, intelligent and model-driven: evolving OSS, https://www.ericsson.com/en/ericsson-technology-review/archive/2018/open-intelligent-and-model-driven-evolving-oss

[12] Architecture evolution for automation and network programmability https://www.ericsson.com/en/ericsson-technology-review/archive/2014/architecture-evolution-for-automation-and-network-programmability

[13] Communications as a cloud service: a new take on telecoms https://www.ericsson.com/en/ericsson-technology-review/archive/2014/communications-as-a-cloud-service-a-new-take-on-telecoms

[14] 4G/5G RAN architecture: how a split can make the difference, https://www.ericsson.com/thecompany/our_publications/ericsson_technology_review/archive/4g-5g-ran-architecture-how-a-split-makes-a-difference

[15] A vision of the 5G core: flexibility for new business opportunities https://www.ericsson.com/thecompany/our_publications/ericsson_technology_review/archive/5g-core-vision

[16] Enabling intelligent transport in 5G networks https://www.ericsson.com/en/ericsson-technology-review/archive/2018/enabling-intelligent-transport-in-5g-networks

[17] Boosting smart manufacturing with 5G wireless connectivity https://www.ericsson.com/en/ericsson-technology-review/archive/2019/boosting-smart-manufacturing-with-5g-wireless-connectivity

[18] 5G network programmability for mission-critical applications https://www.ericsson.com/en/ericsson-technology-review/archive/2018/5g-network-programmability-for-mission-critical-applications

[19] Distributed cloud: A key enabler of automotive and industry 4.0 use cases https://www.ericsson.com/en/ericsson-technology-review/archive/2018/distributed-cloud

[20] Simplifying the 5G ecosystem by reducing architecture options https://www.ericsson.com/en/ericsson-technology-review/archive/2018/simplifying-the-5g-ecosystem-by-reducing-architecture-options

[21] Edge Computing and 5G
https://www.ericsson.com/en/digital-services/forms/cloud-nfv/ericsson-edge-computing-paper

[22] Driving transformation in the automotive and road transport ecosystem with 5G
https://www.ericsson.com/en/ericsson-technology-review/archive/2019/transforming-transportation-with-5g

[23] Communication services over LTE, Wi-Fi and 5G
https://www.ericsson.com/en/white-papers/voice-and-video-calling-over-lte--securing-high-quality-communication-services-over-ip-networks

[24] 5G deployment considerations
https://www.ericsson.com/assets/local/networks/documents/5g-deployment-considerations.pdf

[25] Ericsson Mobility Report
https://www.ericsson.com/en/mobility-report

[26] Six key trends manifesting the platform for innovation
https://www.ericsson.com/en/ericsson-technology-review/archive/2019/technology-trends-2019

[27] 5G New Radio RAN and transport choices that minimize TCO
https://www.ericsson.com/en/ericsson-technology-review/archive/2019/5g-nr-ran-and-transport-choices-that-minimize-tco