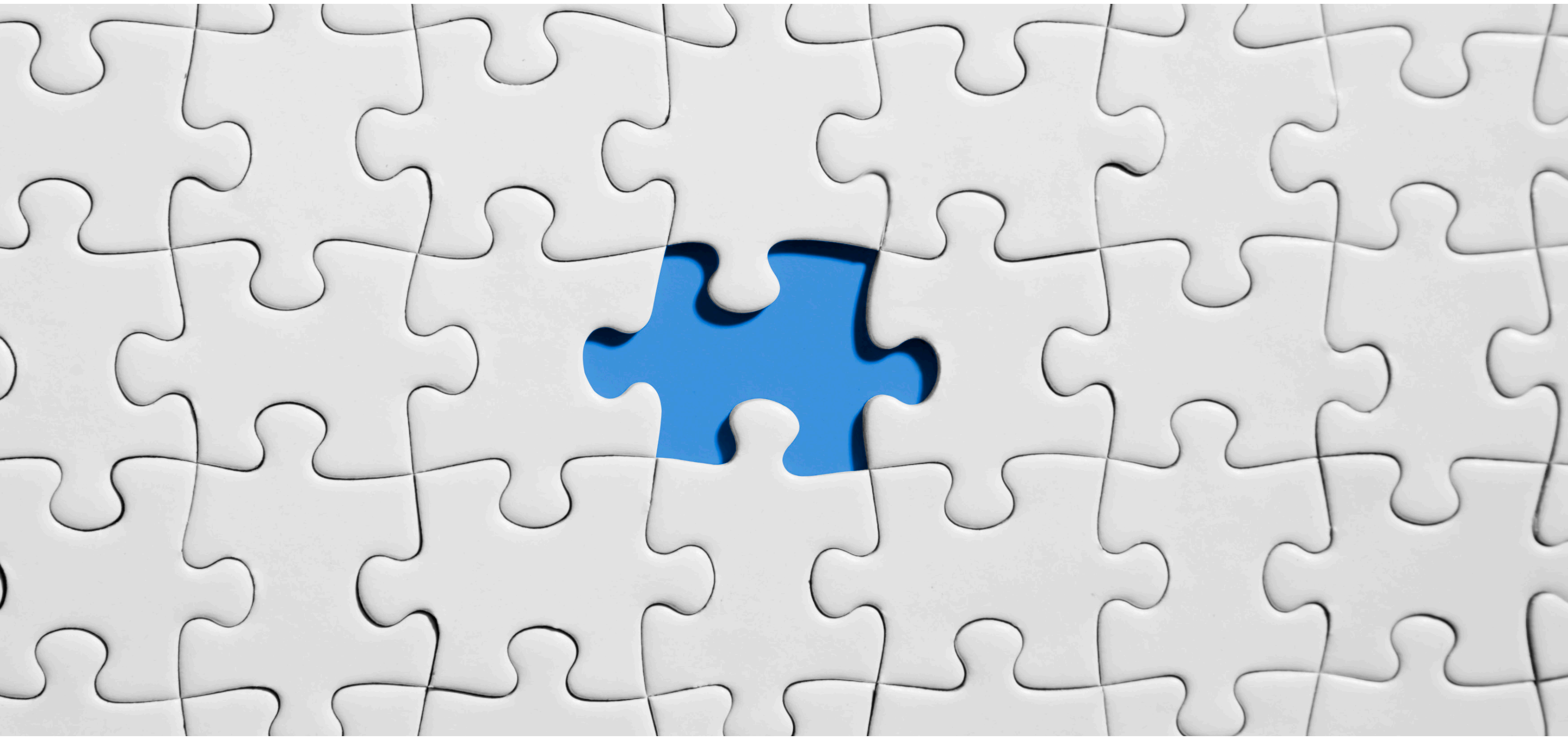
HOW 5G TRANSFORMS CLOUD COMPUTING



Mikhail Gloukhovtsev

Sr. Solutions Architect

Digital Solutions, Cloud & IoT

Orange Business Services

Mikhail.gloukhovtsev@orange.com

In memory of my father

The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

Table of Contents

1. Introduction

5G – the fifth generation of cellular network technology – is sometimes referred to as “the 5G wireless revolution.” Indeed, 5G offers many advantages such as significantly higher data-throughput rates compared with the previous 4G (LTE) technologies, low latency, and greater system capacity. It will affect many industries and services including cloud computing. Opinions on how 5G will transform cloud computing run the gamut from considering 5G as a cloud computing killer to expecting it to be a great enabler of new cloud-based applications and services that were previously impossible to provide.

Some industry researchers think that ultra-fast and high-capacity 5G network makes data processing and transfer directly between autonomous devices possible and, as a result, cloud computing as it exists today will become a thing of the past. According to the opposite views, 5G will enable the use of a wide range of completely new applications and services in cloud computing. The synergy of 5G cellular and cloud technologies will provide the possibility of creating innovative business services.

5G changes the cloud computing landscape by adding new architectures and technologies such as Mobile Cloud Computing (MCC), Multi-access Edge Computing (MEC), Cloud Radio Access Network (C-RAN) and others. There are predictions of developing “all-cloud 5G architecture” that enables an all-cloud digital transformation of networks, Operation Support Systems (OSS)/Business Support Systems (BSS), and services.

In this article I consider the emerging concept of 5G and cloud computing symbiosis and how it transforms cloud computing. This transformation results in new cloud computing architectures, affects various businesses using cloud services, and brings new business opportunities. I also review key challenges in 5G-cloud solutions. I hope my article will help the readers in developing cloud-services strategies in the era of 5G.

2. 5G Technology

2.1 What Is 5G and How Is It Different From 4G (LTE)?

5G is a standard for the next generation of wireless network technology. It is being defined by the International Telecommunications Unit (ITU) that introduced the International Mobile Telecommunication system for 2020 and beyond standard—IMT-2020—in 2017. The framework and overall objectives of developing the 5G standard are defined in IMT-2020. To meet IMT-2020 requirements, the industry association 3GPP (3rd Generation Partnership Project) has introduced the 5G New Radio (5G NR) standard. A system using 5G NR software is defined as a “5G” system.

Table 1 shows the evolution of the main features of 1G-5G technologies. 4G provides significant bandwidth and capability improvements over 3G. The transition to 5G adds more frequency bands, both in and around the existing 4G spectrum, as well as a new millimeter wave (mmWave) spectrum (see Section 2.2.1). 5G supports low latency and mission critical applications, i.e. ultra-reliable and low latency communications (URLLC), enhanced mobile broadband (eMBB), etc. (Section 2.3). As seen from Table 1 and Figure 1, 5G will deliver multi-Gb/s peak rates and ultra-low latency.

	1G	2G	3G	4G	5G
Approximate deployment date	1980s	1990s	2000s	2010s	2020s
Theoretical download speed	2kbit/s	384kbit/s	56Mbit/s	1Gbit/s	10Gbit/s
Latency	N/A	629 ms	212 ms	60-98 ms	< 1 ms

Table 1. The Evolution of the Main Features of 1G-5G Technologies (Ref.1)

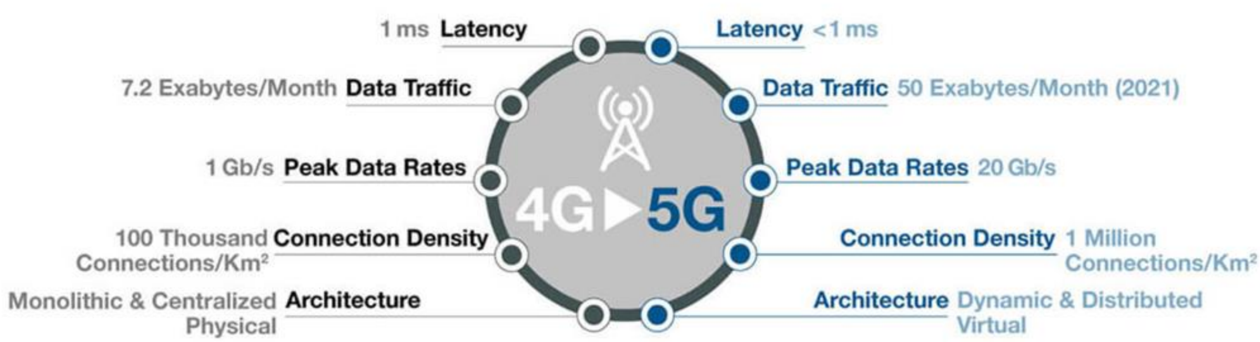


Figure 1. Comparison of Main Features of 4G vs. 5G (Ref.2)

According to IMT-2020, theoretically 5G cell can support speeds of up to 20 Gb/s for downloads and 10 Gb/s for uploads, with latency as low as 4 ms. For example, if downloading an HD video file would take about 10 minutes in 4G, it will take seconds to do so using 5G. However, in most situations, real world download/upload speeds depend on which 5G spectrum band is used (Section 2.2.1).

2.2 5G Features

5G technology features (Fig. 1) and use cases are very different compared with the previous 1G-4G technologies. While the previous wireless technologies were used to connect people to people and to the Internet, 5G connects things to people, to the Internet, and to other things (IoT). 5G networks are part of developing Industry 4.0.

Today we witness an evolution of networks by implementing Software Defined Networking (SDN), Network Function Virtualization (NFV) and cloud-native architectures to enable disaggregation and virtualization of primary functions[1] (see Sections 4.1-4.4). This results in separation of control plane and user plane and introduces capabilities such as network slicing and mobile edge computing (MEC) (Sections 4.3 and 7). 5G initiates the transition to cloud-native networking models enabling a Services-Based Architecture (SBA), which transforms a response-request method of communication into a producer-consumer type model (Section 9).

2.2.1 5G Features: Spectrum and Speeds

Before we consider the speed and bandwidth associated with different bands of the 5G spectrum, let us recall how the signal frequency determines the transmission distance. While all radio waves travel at the speed of light, the frequency used by a 5G tower directly affects the transmission distance. High-frequency waves have shorter ranges. On contract, lower frequency waves are transmitted over longer distances.

The difference between the highest and lowest frequency of the signal defines the bandwidth. Therefore, using higher bands of the radio spectrum results in a broader range of frequencies and in a higher throughput (Fig. 2). For example, millimeter waves in the high-band spectrum are able to carry large amounts of data. However, radio waves in higher bands are also absorbed more easily by gases in the air, trees, and nearby buildings. mmWaves (see below) are therefore useful in densely packed networks, but their transmission distances are not long (Fig. 2). The distance challenge is illustrated by the so-called “three-house rule”: to operate 5G, an antenna is needed for every three houses. Hence there are no good or bad parts of the 5G spectrum and different parts of the spectrum can be used by 5G providers to maximize distance and get as much throughput as possible at the same time.

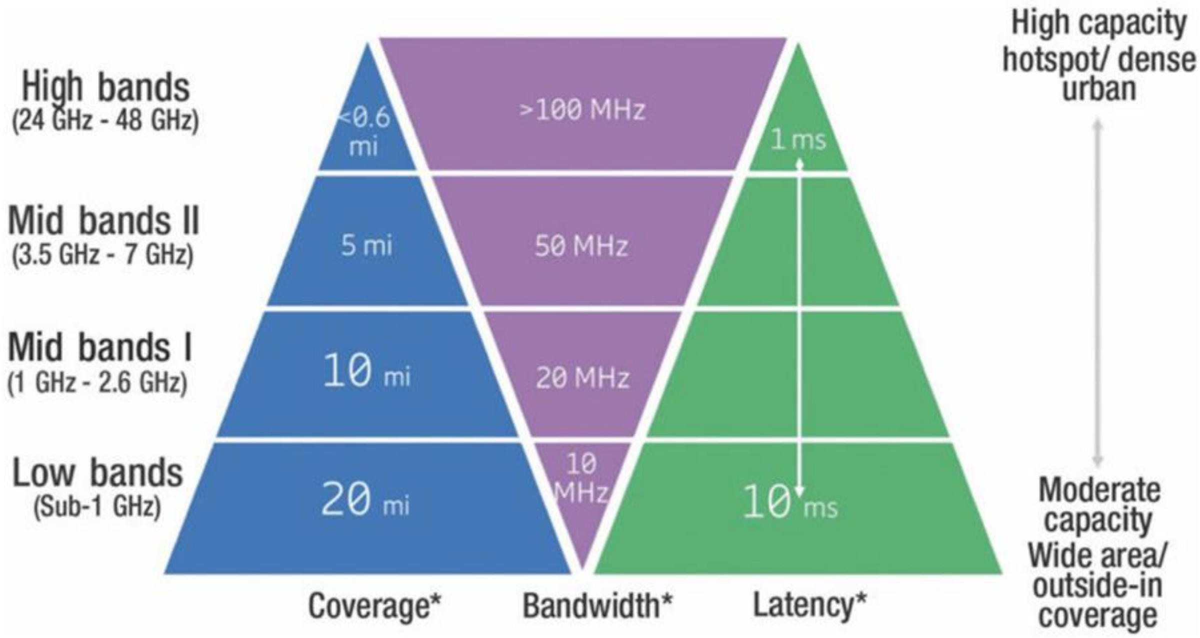


Figure 2. Coverage, Bandwidth, and Latency for 5G Spectrum Bands (Ref.3)

Low-band spectrum. This can be described as "sub" 1GHz spectrum (Fig. 2) that is primarily used today by US carriers for 3G and LTE. The low-band spectrum provides consumers a very wide coverage area with good building penetration, but the peak data speeds are only up to 100 Mb/s. Sub-1 GHz band supports widespread coverage across urban, suburban and rural areas and helps to support Internet of Things (IoT) services.[4]

Mid-band spectrum. This spectrum between 1 and 6 GHz provides faster throughput and lower latency than the low-band spectrum (Fig. 2). Mid-band transmissions are less suitable for good in-building penetration, but peak speeds can reach as high as 1 Gb/s. The coverage, throughput, latency and capacity characteristics of the mid-band spectrum make it an excellent candidate for deployment of 5G URLLC services.[4]

High-band spectrum. Sometimes referred to as mmWave in the industry, this high-band spectrum enables peak data rates up to 20 Gb/s (Fig. 1) with a very low latency. The actual speeds in this band are often 1-2 Gb/s. However, high-band coverage area is not large and the difficulty for these waves to traverse building walls and windows makes indoor coverage limited. Major telecommunication companies are working on solutions for these propagation challenges because mmWave is so fundamental to achieve 5G speed and latency targets.[5]

2.2.2 5G Features: Lower Latency

Real-time applications require low latency that can be reached with 5G. The 5G high-band latency levels are quite low (Table 1, Fig. 2). However it should be noted that frequently cited latencies of 1 ms do not include the time to the server. The latency of 1 ms has been recorded in tests in a laboratory environment.[6] The "air latency" (between a phone and a tower) in 2019 equipment is typically 8-12 ms. The latency to the server and further back in the network increases the average latency to ~30 ms, that is 25%-40% lower than typical 4G deployed. The latency can be reduced to 10-20 ms by adding edge servers (Section 6) close to the towers.

2.2.3 5G Features: Higher Network Capacity

5G network capacity is much higher compared with that of 4G – up to 100x the number of connected devices per unit area. 5G also uses a new technology called Massive MIMO (multiple input multiple output, mMIMO). mMIMO utilizes multiple targeted beams to spotlight and follow users around a cell site for improving coverage, speed, and capacity. As a result, more people can simultaneously connect to the network and maintain high throughput.

2.3 5G Use Cases

5G networks, which are a cornerstone in building up infrastructure to implement new business models, support the following use cases:

1. Enhanced mobile broadband (eMBB): It provides theoretical peak download speeds of up to 20 Gb/s and a reliable 100 Mb/s user experience data rate in urban areas. This will better support increased consumption of video as well as emerging services like virtual and augmented reality.
2. Ultra-reliable and low latency communications (URLLC): Applications used by autonomous vehicles, smart grids, industrial automation, remote patient monitoring, and telehealth require URLLC.
3. Massive machine-type communications (MMC): The ability to support at least one million IoT connections per square kilometer (Fig. 1) with wide coverage including inside buildings.
4. Fixed wireless access (FWA): The ability to offer fiber type speeds to homes and businesses using new frequency bands, mMIMO (Section 2.2.3) and 3D beamforming technologies.

2.4 Transition to 5G

5G will initially operate in combination with existing 4G networks. The transition to 5G will require disruptive changes in the network architecture and infrastructure of telecommunication companies (telco) and service providers (Section 11). It will also result in a more distributed architecture (Section 8) with an increasing role of edge computing providing compute/storage and analytical services closer to the source of data (Section 6). Understanding these technology transformations, Dell Technologies has started executing on a 5G strategy[7] based on the following guiding principles:

- 5G requires a new foundational architecture resulting from the once-in-a-decade re-architecture of cellular networks.
- 5G will be the first end-to-end architecture which is completely software-defined, from the radio access network (RAN) (Section 4.5) to the core. Dell Technologies' experience in software-defined architectures puts the company in the best position to help the service provider partners in 5G architectural transformation.
- The new operational model will be based on APIs and software programmability at a level not yet seen in networking. It will provide the separation of control and user planes (CUPS) taking place not just at the macro level (network), but at the micro level.
- This new paradigm will result in de-centralization of the infrastructure (Section 8).
- 5G Operations will be data-driven and the need to capture, process, and act on network data and events in real-time will increase the role of machine learning.
- Openness and disaggregation of the different architecture layers (Section 4.2) are key design principles in this new 5G world.

3. Making 5G Work: 5G and Cloudification

There are two key aspects in the relationship between 5G technologies and cloud computing. First, further development of cloud computing has to meet the 5G needs. This is reflected by growing roles of edge, mobile edge, and fog computing in the cloud computing realm. The second aspect is that 5G technologies are undergoing “cloudification” through network “softwarization” (Section 4), NFV, SDN, etc. Hence, both technology types influence the developments of each other. 5G deployments bring up discussions about the convergence of computing, cloud, and IoT that takes us to the era of hyper-connectivity. In this article I review both aspects and focus on the interdependencies in the evolution of 5G and cloud technologies.

3.1 Will 5G Technologies Kill Cloud?

Some tech industry experts say that 5G could be a cloud computing killer and the cloud as we know it will become obsolete. There are even references to “post-cloud world.” For example, Jon Markman has predicted in his publication in Forbes[8] that ultra-fast wireless networks may eliminate the cloud as a computing platform. A post-cloud world would comprise billions of autonomous smart devices. The data collected at the network edge will be processed by these devices in real time. This is an extreme view and I will show in the next sections that 5G and cloud computing are in a symbiotic relationship.

3.2 Symbiosis Between 5G and Cloud Computing

As 5G devices only process local data, cloud computing capable of providing elastic compute resource pools becomes critical for 5G technology-based services for managing compute- and data-intensive applications such as Big Data analysis and AI applications. To meet 5G demands, cloud computing transformed by including mobile cloud computing (MCC, Section 5) and multi-access edge cloud computing environment (MEC, see Section 7) plays an important role in developing 5G services. As a result, next generation cloud technologies will require a paradigm shift in their architectures and management frameworks.[9] Furthermore, the cloud is considered as the foundation block for cloud radio access network (C-RAN) (Section 4.5).

Cloud computing and 5G networks are in mutually driven transformations. Indeed, not only are cloud platforms changing for meeting 5G requirements but these changes in the cloud reshape the telecoms architectural landscape by offering flexibility and lower total costs as well as high service availability (see Telco Clouds in Section 11). Cloud capabilities will be extended to the network edge to form an edge cloud like MEC (Sections 6 & 7). This enables network functions like C-RAN to run at the edge of the network. The future network architecture will be not exclusively edge cloud or core cloud but a hybrid of both to form distributed cloud computing (Section 8).

To meet the requirements specified in IMT-2020, 5G network in its turn undergoes a “Cloudification” process including software-defined networking (SDN), network virtualization, network slicing, and self-managed networks. We will consider these technologies in the following sections. Let us start with 5G network “softwarization.”

4. 5G Network Softwarization as Network Cloudification

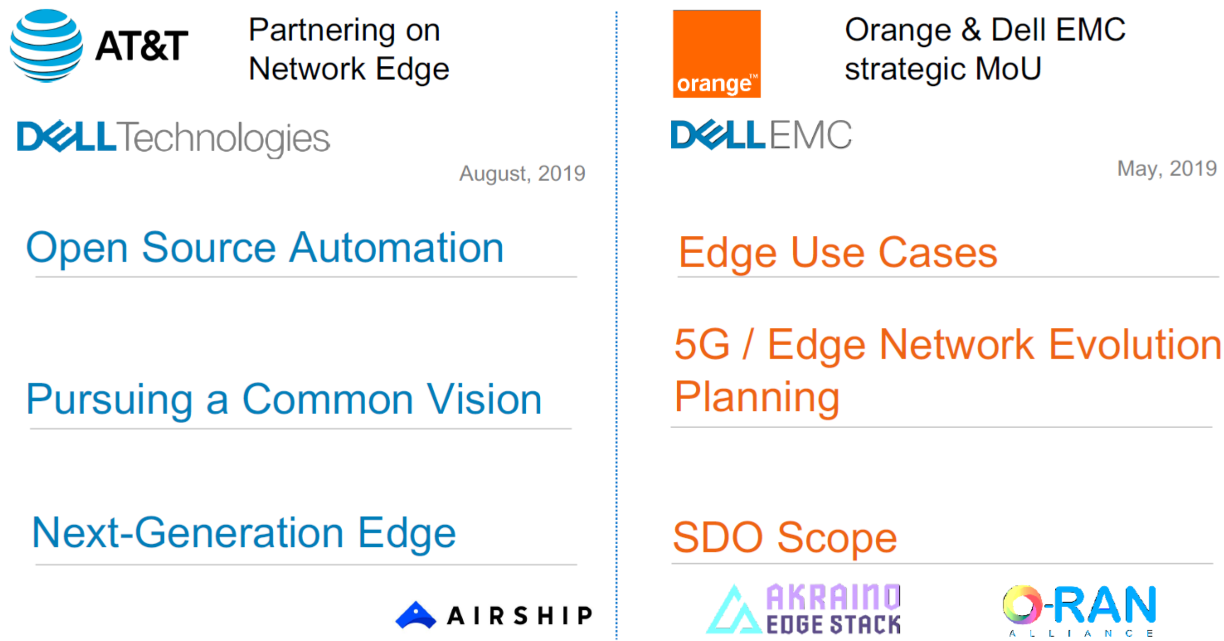


Figure 3. Dell Partnerships with AT&T and Orange (Ref.11)

The emerging term *network softwarization* refers to the overall approach in designing, implementing, managing, and maintaining network equipment and/or network components by software programming. Softwarization takes place at all network layers by leveraging SDN, NFV, edge and cloud computing. It becomes a key feature of future-generation wireless network systems (Section 4). Taking this into consideration,[10] the 5G Infrastructure Public Private Partnership (5G-PPP) has included several key software-defined components in 5G standardization, such as SDN, NFV, and multi-access mobile edge computing (MEC) (Sections 4.4, 4.1, and 7).

5G-driven network cloudification has become a catalyst for establishing partnerships of telecommunication companies with cloud services companies (Fig.3). For example, AT&T recently announced a partnership with VMware (Dell Technologies is the parent company of VMware) for combining their respective expertise in open source software to develop solutions enabling the transition to a cloud-native 5G network.

Dell Technologies Standards and Forums Participation

Figure 4. Participation of Dell Technologies in Standard Development (Ref.11)

To gain advantage of the new business opportunities created by 5G, communication service providers (CoSP) need open, industry-standard architectures including SDN, NFV, cloud native applications, and MEC.[12,13] Dell Technologies actively participates in developing 5G standards (Fig. 4) and this positions it as a key partner for CoSPs; the partnership with Orange is an example (Fig. 3).[12] Having decades of data center transformation experience, Dell is working closely with Orange to help in re-tooling Orange operations to quickly and profitably implement new 5G services. According to the partnership announcement,[12] Dell Technologies and Orange will collaborate on the definition and development of the following:

- Use cases, business models and proof of concepts for Edge Cloud
- Open source consortia and partnerships for the edge ecosystem
- Definition and validation of infrastructure accelerators, such as FPGAs, GPUs, and SmartNICs (Section 11.2), for edge workloads, including cloud/virtual RAN (C-RAN/V-RAN), MEC, and real-time, interactive, latency-sensitive applications
- AI/ML-enabled software to support remote automation of a multi-technology, heterogeneous edge built on virtual machines, containers, and bare metal workloads
- Edge infrastructure platforms supporting Telco environmental, space, operational, and automation requirements

5G-driven cloudification in Telcos is influencing operational transformation as well, leading to a transformation in B2B services to software-defined WAN (SD-WAN) and management of multi-cloud services.

4.1 Network Function Virtualization

Network Function Virtualization (NFV) is gaining momentum quickly as a way to transfer network functions from proprietary network hardware to software-based applications executing on general-purpose commercial off-the-shelf (COTS) hardware. NFV benefits are reductions in CapEx and OpEx (for example, power consumption), increased efficiency (infrastructure shared by multiple tenants), flexibility to scale resources up or down, agility in deployment of new network services by reducing the cost of network changes and upgrades. NFV enables telecommunication companies to transition from operating the network to programming the network.

Dell Technologies and VMware have been working jointly on validated solutions for the network edge to deliver on a vision of a new generation intelligent, programmable, and automated edge platform. The Dell Technologies 5G Ready Solution for VMware NFV Platform[14] combining hardware, software, and Dell Technologies engineering is designed to create a more flexible, scalable, and agile platform for CoSPs. It includes open standards-based Dell Technologies cloud infrastructure hardware (compute, networking) and a choice of a Virtual Infrastructure Manager (vCloud Director or VMware Integrated OpenStack) with vSAN (Fig. 5). The solution supports the latest 14th generation PowerEdge Servers based on Intel Xeon Scalable Processors. The solution software is optimized for Dell Technologies cloud infrastructure.

Dell EMC Networking

2 x S6010-ON

1 x S4048T-ON

PowerEdge Rack Servers

8 or 12 x R640/R740/R740xd

Figure 5. Dell 5G Ready Solution for VMware NFV Platform Configuration (Ref.15)

4.2 Network Function Disaggregation (NFD)

Network function disaggregation (NFD) represents the evolution of switching and routing appliances from proprietary, closed hardware and software solutions to totally decoupled, open components. Employing COTS x86, ASIC or programmable merchant silicon (within generic servers or white box switches), disaggregated devices are horizontally integrated. NFD allows for selecting and integrating the network operating system (NOS), Layers 2 and 3 networking protocols and the element management interfaces for each specific requirement. This results in a bespoke solution that is perfectly tailored for a given 5G application.

4.3 Virtualization in 5G Networks: Network Slicing

One of the fundamental changes in 5G compared to older technologies is that the network architecture will no longer be a structure of monolithic elements. Instead, the architecture is based on a network slicing concept that makes use of network virtualization and softwarization of the different network elements. Network slicing enables creation of multiple virtual networks on top of a common shared physical infrastructure.

The 5G Infrastructure Public-Private Partnership (5G-PPP)[16] defines a ”network slice” as a composition of network functions, applications, and infrastructure bundled together. A network slice can be allocated to specific applications, services, use cases, or business models to meet their requirements. It can be dedicated to a specific tenant (e.g., a service provider) that, in turn, uses it to provide a specific telecommunication service (e.g., eMBB). The decoupling between the virtualized and the physical infrastructure (Section 4.2) allows for efficient scaling of slices. Network slices span the whole protocol stack from the underlying (virtualized) hardware resources up to network services and applications running on top of them. From a business point of view, a network slice provides a combination of all the relevant network resources, network and service functions required to fulfil a specific business case or service, including OSS and BSS.[17]

A generic framework for 5G network slicing consists of several main layers: infrastructure layer, network function and virtualization layer, service layer, and slicing management/orchestration layer.[18]

Infrastructure Layer: The infrastructure layer defines the actual physical network architecture. In this layer, several policies are conducted to deploy, manage, and orchestrate the underlying infrastructure. This layer allocates resources (compute, storage, bandwidth, etc.) to network slices in such way that upper layers can get access to handle them according to the context.

Network Function and Virtualization Layer: The network function and virtualization layer executes all the required operations to manage the virtual resources and network function's life cycle. SDN, NFV and different virtualization techniques operate at this layer.

Service and Application Layer: The services and applications for connected vehicles, virtual reality appliances, mobile devices, etc., function at this layer.

Slicing Management and Orchestration (MANO): This layer is used for monitoring and management of the functionality of the above layers,

Network slicing enables a Network-as-a-Service (NaaS) business model[19] that allows CoSPs for using subscription-based provisioning of cloud-transformed network services that are customized to users' demands. 5G security (Section 14) benefits from network slicing, as it ensures that the security-related events in one slice do not affect the others.

4.4 SDN and Cloud Computing: Working in Unison

4.4.1 SDN as a Way to 5G Network as Code

SDN enables dynamic reconfiguration of network elements in real-time so that 5G networks can be controlled by software rather than hardware.[20,21] The main idea behind SDN is to move the control plane away from network hardware and enable external control of data through a logical software entity called a *controller*. The controller that manages packet-flow control to enable intelligent networking is layered between network devices and applications.

4.4.2 SDN Integration with the Cloud

The SDN concept has been around for years, but its role in the cloud becomes prominent with the introduction of 5G networks. SDNs can be changed quickly and all together, without a need for reconfiguring each hardware device individually. As these networks are programmable through software, they integrate well with cloud platforms.

So what can be done to make software-defined 5G networks more integrated with public clouds? Let us consider Cisco Cloud Application Centric Infrastructure (Cisco Cloud ACI) as an example. Cisco Cloud ACI enables managing private cloud-based SDN and a public cloud-based SDN together, from one console and from one view. The SDN provider, not the cloud provider, controls the relationship. In this case, ACI controller is integrated with an API from a public cloud provider, such as Microsoft, Google, or Amazon. A greater transparency between the SDN and the public cloud is achieved and as now they are aware of each other, they can work together (see also Section 10).

The future of software-defined 5G networking and cloud computing integration will likely to include the following:

- Integration between enterprise SDNs and clouds will continue to deepen
- Integrated network monitoring and management
- Machine learning enablement

4.5 Moving Virtualization From the Core to the RAN. The Role of Cloud RAN in 5G Networks

The main components of radio access network (RAN) include a base station and antennas. RANs connect mobile users and wireless devices to the main core network. A major feature of 5G networks are small cells particularly at the new mmWave frequencies with a very short transmission distance (Section 2.2.1). 5G macro cells use MIMO antennas leveraging active antenna systems (AAS) (Section 2.2.3). This provides more simultaneous connections to the network and maintains high throughput.

5G is transforming RAN into cloud-RAN (C-RAN).[22] The key function of C-RAN is to provide resource "cloudification" by pooling the baseband processing resources to manage and dynamically allocate them on demand. Three primary components of C-RAN architecture are as follows (Fig. 6):

1. Pool of centralized baseband units (BBU): The BBU pool functions as a cloud. The multiple BBU nodes in the pool dynamically allocate resources to remote radio units (RRUs) based on current network needs.
2. RRU networks: RRUs of different cells are connected to the cloud via a high-speed front-haul link. RRUs consume much less energy and require less CapEx and OpEx than traditional base stations.
3. Fronthaul/transport network: The fronthaul links carry data in the Common Public Radio Interface (CPRI) format between a BBU and a set of RRUs.

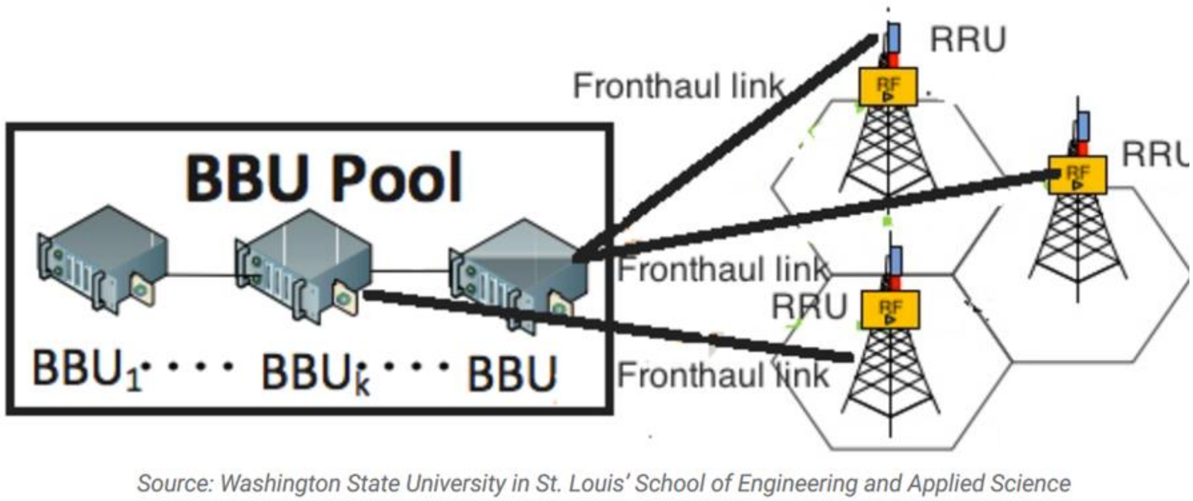


Figure 6. **C-RAN Architecture Components (Ref. 23)**

Splitting the base stations into radio and baseband parts and pooling BBUs from multiple base stations into a centralized and virtualized BBU pool is the main idea of the C-RAN concept. C-RAN, which can be seen as a logical evolution in fiber-to-the-antenna (FTTA) networks or distributed RAN (D-RAN), moves the RAN functionality to the cloud computing platform. The use of centralized cloud processing provides C-RAN with flexibility in terms of signal processing complexity and coordination among cells and networks so that the resources can be used efficiently.[23-26] The advantages of C-RAN architecture include lower total-cost-of-ownership (TCO), enhanced spectral efficiency, the ability to reuse infrastructure, simplified network management, and simplified support of multi-standards.

For delivering high-speed, low-latency connectivity between C-RANs and ecosystems of network carriers, cloud providers, and business users, 5G infrastructures can be deployed on global interconnection and colocation platforms, such as Platform Equinix using Equinix Cloud Exchange Fabric (ECX).[27]

For the best support of the 5G high volume traffic, SDN and NFV technologies are integrated by carriers into their C-RANs. As a result, C-RAN operators are able to create virtual RANs (V-RANs) by enabling baseband BBU functions on virtual machines at a centralized cloud location. For example, in November 2019, Dell Technologies and Wind River announced a V-RAN solution that integrates Wind River Cloud Platform – cloud-native Kubernetes and containerbased architecture for automated management of thousands of nodes in an edge cloud infrastructure – with Dell PowerEdge server hardware.[28,29] The high compute power, performance, and memory of Dell Technologies PowerEdge servers make them capable of supporting low-latency applications. This shows that V-RAN-based solutions can cost effectively provide carrier grade performance and massive scalability.

5. 5G and Mobile Cloud Computing (MCC)

5.1 What Is MCC?

MCC is a cloud computing system including mobile devices and delivering applications to the mobile devices (Fig.7).[30] Key features of MCC for 5G networks include sharing resources for mobile applications and improved reliability as data is backed up and stored in the cloud. As data processing is offloaded by MCC from the devices to the cloud, fewer device resources are consumed by applications.[31-33]

5.2 MCC Architecture

The general architecture of MCC is presented in Figure 7. Compute-intensive processing of mobile users' requests is off-loaded from mobile networks to the cloud. Mobile devices are connected to mobile networks via base stations (e.g., base transceiver station, access point, or satellite).

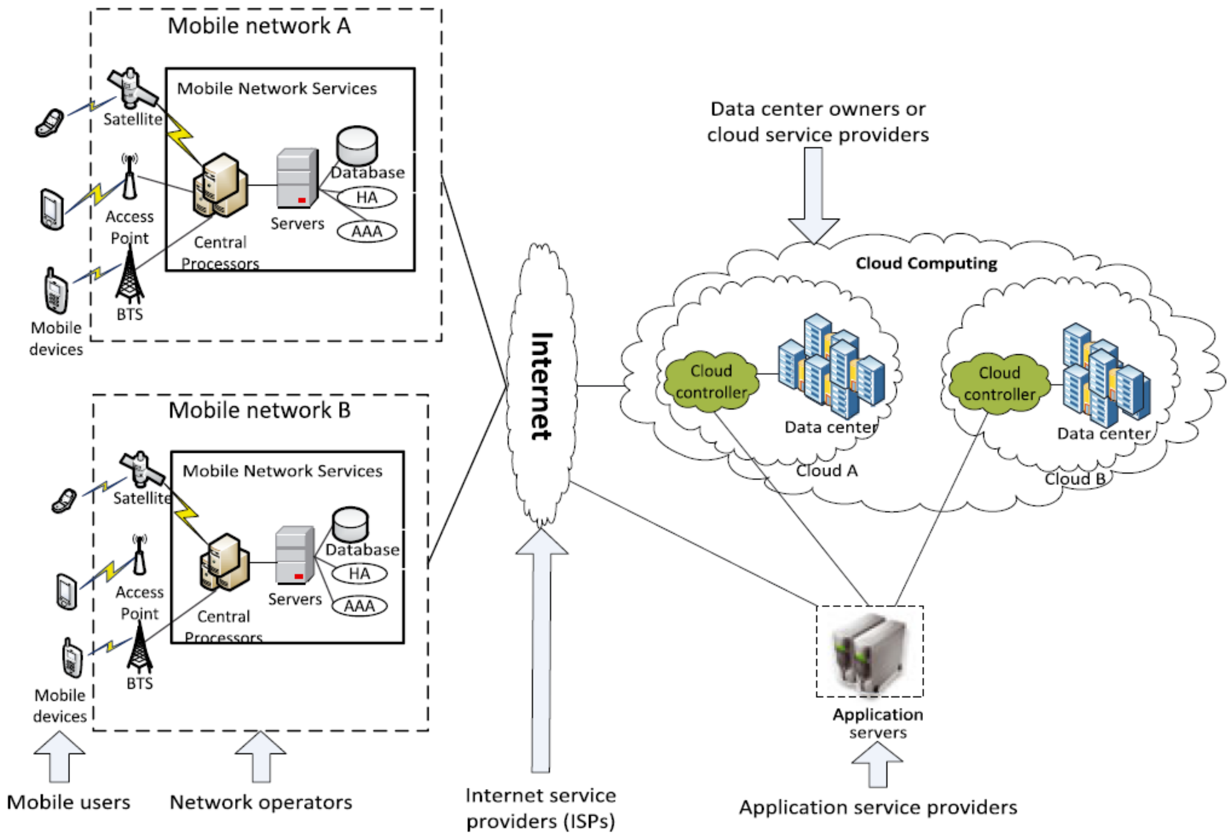


Figure 7. Mobile Cloud Computing Architecture for 5G Networks (Ref.32).

6. 5G Extends the Cloud to the Edge

6.1 Edge Computing

Edge computing refers to infrastructure that enables data processing as close to the source as possible, allowing for faster processing of data, reducing latency and backhaul/core network traffic. This type of computing will require data centers located closer as well as adding micro data centers (Section 13) into or near 5G towers as a means of creating this access.

Edge computing operates in a heterogeneous environment and its standardization is not an easy task. The “architect once, scale forever” approach used in public clouds does not apply to the edge. The broad variety of workloads and compute and storage resources at the edge creates a requirement for a bespoke design. Hence, the need for a “Telco Cloud” (Section 11) has never been more apparent.[34] In many cases, a converged infrastructure is used at the edge. As they have the components needed for a converged infrastructure, Dell Technologies solutions offer the depth of technology and products for building new capabilities for the edge use cases.[35]

Dell Technologies expects to see the following key trends in developing 5G distributed edge architectures:[36]

- Development of ruggedized and purpose-built server platforms for deployment in harsh environments at edge locations (Section 13.3).
- Creating micro data centers for deployments at the edge (Section 13.2).
- Capability for remote life cycle management of hardware and software at the edge.
- Use of field-programmable gate arrays (FPGAs) and SMART-NICs in edge servers to free up the host CPU for customer applications. SMART-NICs combining standard NIC functionality with FPGAs and smaller CPU cores will deliver a programmable network hardware data plane in the server. 5G deployments requiring additional network services such as C-RAN, network slicing, and NFV, will make these FPGAs and SMART-NICs more important in edge architectures.
- Distributed security architectures will emerge to monitor various network traffic patterns and resource consumption at the edge for anomaly detection and threat management.

Edge computing evolution has led to development of smaller form factors that are capable of performing enterprise-level tasks. Smaller form factors have lower requirements for power, cooling, and space for deployments at the edge.

6.2 Fog Computing

Edge computing is considered a subset of the "fog computing" concept that defines the functional characteristics of edge computing. Fog computing’s function is to improve the efficiency of data transport to and from the edge device to the cloud for further data processing, analysis, and long-term retention. We can visualize the relationship among these terms with the conceptual diagram shown in Figure 8. A comparison of the key features of cloud, fog, and edge computing is given in Table 2.

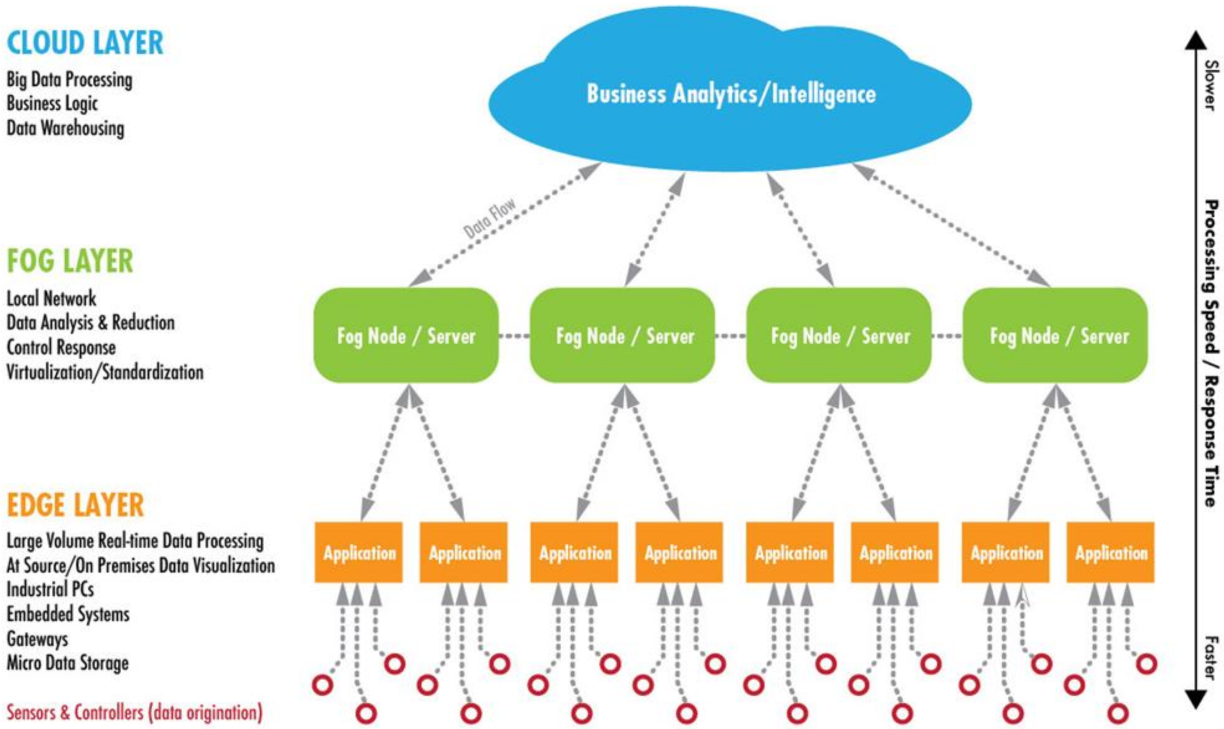


Figure 8. Relationship between Edge, Fog and Cloud Computing (Ref.37).

Computing Concept	**Location of Data Processing**	**Processing Power and Storage Capabilities**	**Security for data transmitted from mobile devices**
Cloud computing	Cloud data center located far from the data source	High performance processing capabilities	Data is vulnerable to cyberattacks during the transmission from the source.
Fog computing	Fog nodes are closer to the data sources than cloud data center but they physically are more distant from the data-generating devices like IoT sensors and actuators.	Less data storage and limited processing power compared with the core cloud.	Less data transmission. Data distribution on fog nodes enhances data integrity.
Edge computing	The data is processed on the data-generating sensor or device itself.	The processing power and storage capabilities are small since both are performed on the devices/IoT sensor	The attack surface is smaller as the data remains on the device itself.

Table 2. Comparison of the Key Features of Cloud, Fog, and Edge Computing.

In a fog environment, data processing is closer to the data-generation location than in the cloud (Table 2). A gateway receiving data from the end points transmits data for processing to the cloud returning processed data back. Fog computing makes traditional networking devices such as routers, switches, set-top boxes, and proxy servers along with dedicated nano-servers and micro-data centers function as fog nodes and create wide-area cloud-like services.[38] Mobile edge servers or cloudlets (Section 13) can also operate as fog nodes. The concept of fog radio access network (F-RAN) where fog resources are used to create a BBU pool for the base stations is also getting attention in the industry.

6.3 How Are Edge and 5G Intertwined?

5G and edge computing are two inextricably intertwined technologies.[39] Indeed, 5G increases speeds of data transfer by up to ten times that of 4G (Section 2.2.1), whereas MEC reduces latency. There are several reasons why 5G needs edge cloud computing (Table 3). Considering the 5G standards suggests that MEC is the main way to meet the IMT-2020 latency targets (Section 2.2.2).

Demand Driver	Edge Capability in 5G
Application Latency	Latency can be reduced to 1 ms if the applications are deployed closer to the users and 5G RAN.
Application Exposure	The new 5G core will offer application exposure for edge deployments.
Transport Offload	Service delivery from edge will minimize the backhaul traffic.
Processing Offload	Application processing at the edge will be offloaded to the core cloud datacenter while preserving the user experience.

Table 3. Interplay of 5G and Edge Computing (Adapted from Ref.40).

6.4 Network Edge Reference Architecture Overview

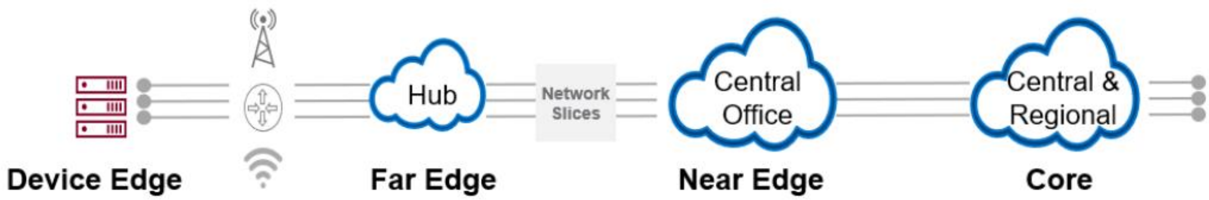
Dell Technologies network edge architecture consists of Core, Near Edge, Far Edge, and Device Edge (Fig. 9) with its focus on the Near Edge. This reference architecture has been created with 5G in mind.[41] It is designed to be consistent throughout the network so local branches can be easily integrated to the core network. Everything is managed as a cloud environment based on software-defined technologies from VMware.

Core. The core is the heart of communication service provider (CoSP) network services. It is a centralized location for the CoSP's management plane and all federated control planes for the Near Edge regions. VNFs such as EPC, IMS, PCRF, MANO, Analytics, and OSS/BSS, are part of this 4G/5G core architecture.

Near Edge. Near Edge is defined in the Dell Technologies network edge architecture as network infrastructure that is close to the network core but far enough away that it cannot rely only on the core network's services. It has a limited amount of server resources and storage space. Near Edge aggregates all types of traffic from fronthaul to backhaul. Each Near Edge site manages multiple Far-Edge sites.

Far Edge. Far Edge is related to a micro data center (Section 13) located at a cell tower or close to a customer's premises. This location, sometimes referred to as the "Last Mile" to the subscribers, is the closest to the user's access. Its function is to provide key services where latency and reliability are the most critical factors to the users. Network services deployed and utilized by Far Edge include V-RAN (Section 4.5), industrial IoT, connected cars, AR/VR, etc. Most of these services require ultra-low latency, high network bandwidth, and real-time synchronization.

Device Edge. Device Edge refers to customer or enterprise premises where thousands of devices are connected to the network and can act as sources and consumers of data. These include mobile devices, IoT sensors, and connected industrial equipment.



Metric	Device Edge	Far Edge	Near Edge	Core
Deployment / location	Enterprise / residential	Cell-sites, street cabinets	Central office / local exchange	Core datacenter
# of instances	Millions	1000s	100s	10s
Latency	40-80 microseconds	0.5 ms (for V-RAN)	5-10 ms	20-50 ms
# of servers / racks	1-3 servers	1-5 servers elastic	1-5 racks elastic	1 rack or more
Other considerations	Thermal / security	Environmental conditions	Environmental conditions	Standard DC
Use cases	Managed enterprise medical applications, IoT applications	Video streaming, AR/VR, gaming, autonomous vehicles	5G enablement, video surveillance, video transcoding, IoT	5G enablement (core)

Figure 9. Network Edge Requirements (Ref.41)

6.5 Dell Technologies Edge Portfolio

Optimal compute infrastructure plays a critical role in transforming the CoSP network to make it highly efficient, flexible, and scalable as required for 5G deployment. Dell Technologies offers a wide range of compute options which gives CoSPs the flexibility to design their disaggregated network architecture (Section 4.2) based on the appropriate workloads and data center location requirements.[42]

Figure 10. Dell Technologies Edge Portfolio (Ref.41)

Dell Technologies and VMware's reference architecture uses Dell Technologies hardware and VMware software to provide a consistent infrastructure from core to edge that addresses the requirements of a software-defined 5G network.[43]

7. Multi-Access Edge Computing (MEC)

7.1 What Is MEC?

The European Telecommunications Standards Institute (ETSI) defines Multi-access Edge Computing (MEC), formerly known as mobile edge computing, as a network architecture providing IT services and cloud-computing capabilities at the edge of the mobile network, within the RAN and closer to the customers.[44] As a result, MEC reduces latency and enables real-time performance to high-bandwidth applications.[45]

Depending on location of the MEC nodes, there are two types of MEC. The first case is when the MEC node resides inside the CoSP network infrastructure. In the second case the MEC node is located on the customer's premises.

1. **Telco Mobile Core Network:** This is the case for which the MEC concept was originally established. This includes use cases from CS/passive optical network (PON)/mobile telephone switching office (MTSO), virtual central offices (VCO), video hub office (VHO), soft handover (SHO) and others.

2. **Telco Customer-Premises Service Providers:** This concept refers to an architecture using smart edge devices providing the traditional terminations (i.e., NTU, NTD, ONT, STBs, etc.). These new edge devices are also capable of running value added or third-party specialized services. Cloud-native workloads (Section 9) and IoT gateway services are examples of these third-party specialized services.

Modern mobile applications are limited by CPU and memory resources as well as battery capacity since processing and data demands drain the battery quickly. MEC enhances the capabilities of these resource-limited mobile devices by offloading executing data/computation-intensive applications. In essence, MEC creates a new framework to support tight integration between wireless networking and cloud computing as required for implementing 5G network technologies.

7.2 MEC Reference Architecture

A conceptual view of the MEC reference architecture proposed by ETSI[44] is shown in Figure 11. The MEC system resides between the user equipment (UE) and the mobile core networks and includes management and functional blocks at the mobile edge host level and the mobile edge system level. As shown in Figure 11, mobile edge services are hosted by the mobile edge platform, which interacts with mobile edge applications so that they can advertise, discover, offer, and consume mobile edge services. The mobile edge platform provides element management functions and administers application essentials such as the life cycle, service requirements, operational rules, and security.

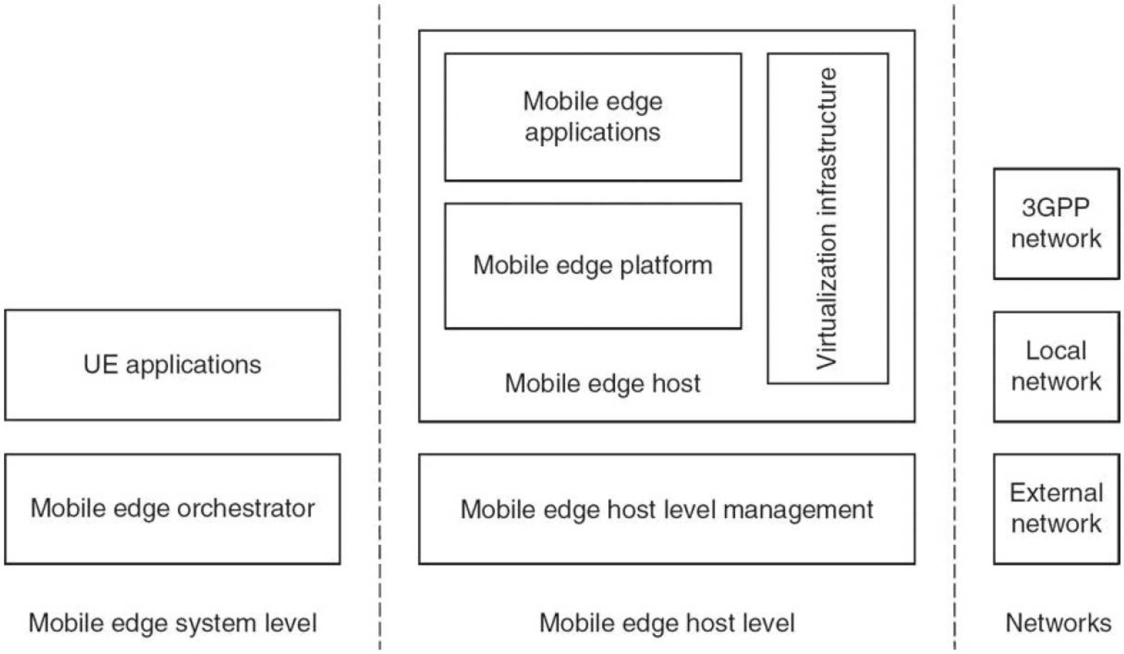


Figure 11. Conceptual ETSI MEC reference architecture (Ref.46)

7.3 Operator-Oriented Use Cases of MEC

Local Content Caching at the Mobile Edge. Shared large files such as high-definition videos are typical on social media, and their transmission stresses the backhaul network capacity. Such content can be cached locally at the mobile edge hosts (Fig. 11) to significantly reduce traffic demand on the backhaul.

Data Aggregation and Analytics. MEC applications running on multiple MEC hosts can process the data collected by end-points to extract the metadata of interest before forwarding the metadata to core servers.

Mobile Media Streaming with Bandwidth Feedback. A radio analytics application running on the MEC host can monitor the available wireless bandwidth and forward it to an MEC-capable back-end video server. Such information can then be used by TCP at the video server to adjust its sending rate properly.

Mobile Backhaul Optimization. Traffic and performance at the RAN level can be monitored and processed by MEC applications. This allows for optimizing the backhaul by using techniques such as application traffic shaping, traffic routing, and capacity provisioning.

Computation Offloading. Offloading computation-intensive and resource-intensive applications to resource-rich servers can augment the capabilities of mobile devices.[47]

7.4 Integration Between MEC and C-RAN

Integration between MEC and C-RAN is challenging since C-RAN uses remote radio head (RRH) architectures. The MEC hosts may reside at the remote radio heads or the baseband units of the C-RAN system. In either case, using network virtualization and MEC platform results in integration with C-RAN.

8. The Role of Distributed Cloud Computing in 5G Networks

In the traditional model of cloud computing, remote sites use the cloud as a hyper-scalable centralized virtual datacenter. 5G transforms the cloud computing concept in the opposite direction – to distributed cloud. Distributed cloud is a cloud execution environment that is geographically distributed across multiple sites, including required connectivity in between, managed as one entity and perceived as such by applications.[48]

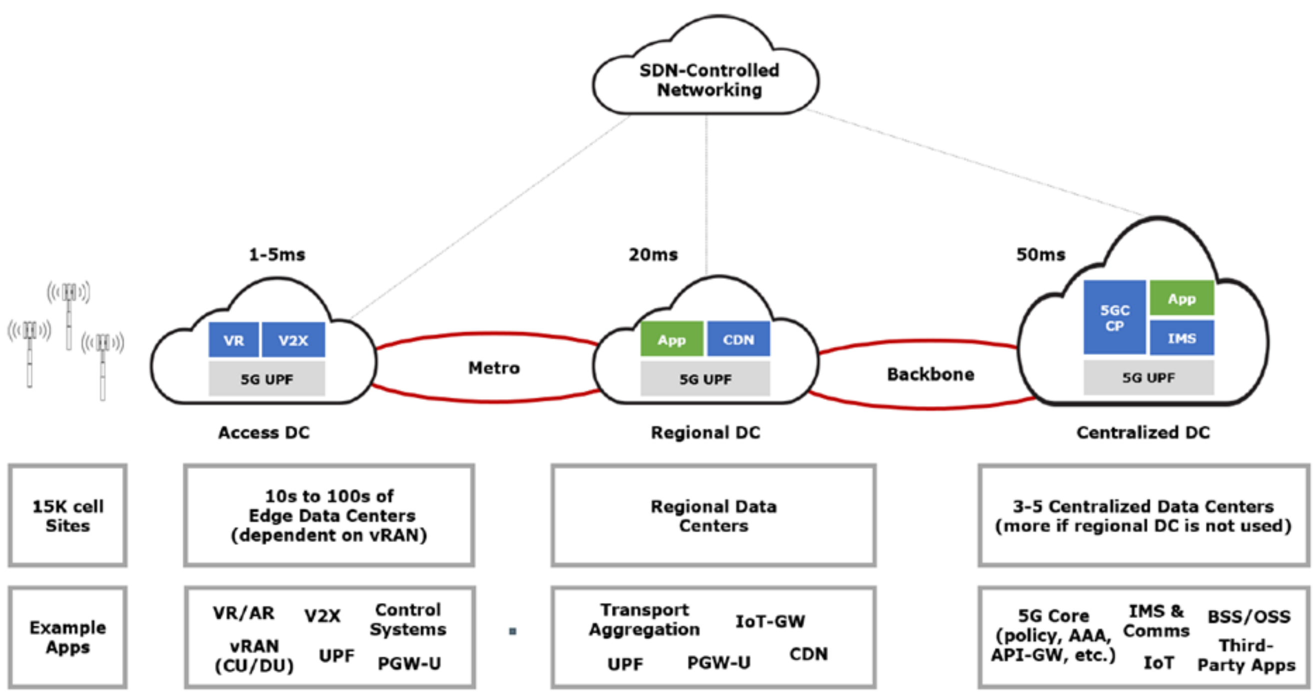


Figure 12. Distributed Cloud Concept for 5G Networks (Ref.50)

Distributed cloud composed of central cloud and edge cloud is based on SDN (Section 4.4), NFV (Section 4.1) and 3GPP (Section 2.1) technologies to enable multi-access and multi-cloud capabilities (Fig.12).[49,50] It can have many edges and the placement of workloads will be determined by the location where the application's requirements can be met at the lowest cost. The key characteristic of distributed cloud is abstraction of cloud infrastructure resources, where the complexity of resource allocation is hidden to an application.

9. Cloud-Native 5G Platform and 5G Service-Based Architecture

The key business objectives for 5G, which can only be delivered by a cloud-native environment, include better bandwidth, latency, and density. Cloud-native network architecture (Fig.13) for 5G has the following characteristics:

1. Cloud data center-based architecture and logically independent network slicing on the network infrastructure.
2. It uses C-RAN to build RAN and to implement 5G required for on-demand deployments of RAN functions.
3. It provides simpler core network architecture and on-demand configuration of network functions via control and user plane separation (CUPS, Section 2.4).
4. It implements automated network slicing service (Section 4.3) to reduce operating expenses.

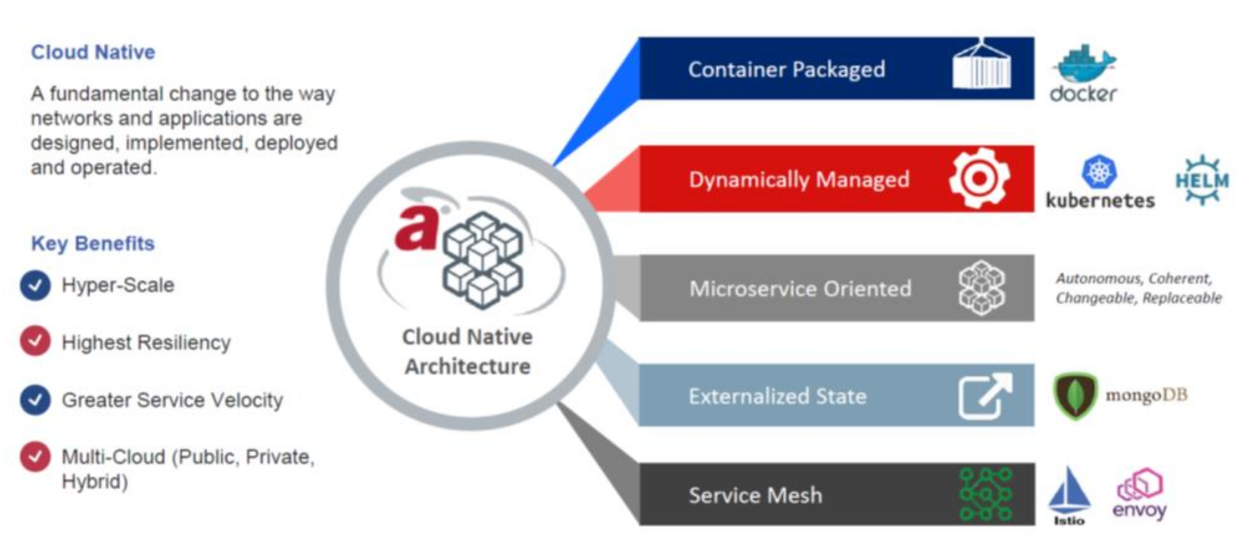


Figure 13. Cloud-Native Architecture Features (Ref.51)

Using cloud-native architecture for 5G enables new services that go beyond traditional broadband, voice, and messaging. The core network of 5G must be dynamic, efficient, and scalable to connect new devices, deliver high performance, meet traffic demands, and extend services over any access technology. A cloud-native architecture is considered to be the only way to achieve these goals.[52] For example, one of the limitations of NFV is the use of monolithic virtual network functions (VNFs) running in virtual machines. Cloud-native implementations of VNFs in containers as containerized network functions (CNF, also known as Cloud-native Network Functions) result in the paradigm shift from a monolithic architecture to a microservices architecture. CNFs providing more portability and scalability than VNFs are gaining popularity in Telco clouds (Section 11) running VNFs in Docker containers or deploying them on Kubernetes (Fig. 13).

The 3GPP has recognized that 5G requires increased service agility. The 3GPP working group has designed a services-based architecture (SBA)[53] that mirrors cloud-native architecture. Just like containers and micro-services architectures did for cloud computing, SBA will make 5G services more agile.

10. 5G Edge and Cloud Hyper-Scalers

The cloud hyper-scalers – AWS, Microsoft Azure, and Google Cloud – recognize the growing role of edge computing as 5G transforms the traditional cloud computing architecture (Section 6). As a result, they have come to the market with edge computing products. For example, AWS Outposts, Local Zones, and AWS Wavelength, announced in 2019, are designed to move data processing closer to the customer and end users. AWS Outposts are fully integrated configurations with built-in top-of-rack switches. The AWS-developed software connects AWS-designed hardware to the AWS cloud. This allows customers to run sensitive applications on-premises and use AWS cloud for more generic workloads. While Outposts target on-premises workloads and resources, Local Zones solution focuses on end users and resources located not on premises but in a particular geographic area. Local Zones enable AWS to build mini data centers (Section 13) by using Outpost server racks.[54]

With AWS Wavelength, AWS services can be used at the edge of the 5G network to minimize the latency for connecting a mobile device to an application.[55] Application traffic does not need to leave the mobile provider’s network in order to reach application servers running in Wavelength Zones. Wavelength combines AWS compute and storage services with the high bandwidth and low latency of 5G networks and offers AWS benefits, like elasticity, availability, and low pay-as-you-go pricing. Wavelength will be available on networks of telecommunication providers such as Verizon, Vodafone, KDDI, and SK Telecom in the United States, Europe, Japan, and Korea.

In November 2019, Microsoft and AT&T announced opening select preview availability for Network Edge Compute (NEC) technology.[56] NEC enables the use of Microsoft Azure cloud services at AT&T network edge locations closer to customers. As a result, AT&T's software-defined and virtualized 5G core – which AT&T calls the Network Cloud – becomes capable of delivering Azure services. As MEC enables a low-latency connected Azure platform in a customer's location, NEC provides a similar platform in a network carrier's central office to make Azure cloud services available. Since deploying Azure data centers in every major metropolitan city throughout the world is not cost effective, the NEC platform offers a solution for specific low-latency application requirements that cannot be satisfied in the cloud.

11. Telco Cloud

11.1 What is Telco Cloud?

As this article shows, 5G is not just technology. In fact, it is a telecom transformation modernizing the network into a telco cloud using a software-defined, automated, virtualized infrastructure built on NFV.[57] Telco cloud, NFV and SDN are changing the telecommunications industry landscape (Table 4). To meet the demand fueled by 5G for cloud-like agility, telecommunication companies need to configure their networks to use cloud software, rather than purpose-built appliances. A telco cloud virtualizes network infrastructure and provides the cloud data center resources needed by telecommunication companies operating at scale to deploy and manage 4G/5G mobile networks with data transfer capabilities. Recent announcements by AT&T on partnerships with Dell Technologies, Microsoft, and IBM demonstrate the importance of cloud computing to the telecom industry.[58] Partnerships with the global leaders in cloud computing such as this partnership enable telecommunication companies to position themselves as trusted cloud service providers to enterprises.

NFV Infrastructure	OpenStack for Telco	NFV Ecosystem
Accelerating time to market by using a fully integrated NFV platform optimized for both 4G and 5G networks. It reduces infrastructure costs and increases operational efficiency.	Providing the fastest path to deployment on OpenStack with a production-proven carrier-grade OpenStack solution.	Benefit from the most comprehensive NFV ecosystem certification program, with over 120 VNFs certified on VMware vCloud NFV.

Table 4. Telco Cloud Components (Ref.59)

The cloud offered by communications service providers (CoSPs) delivers the flexibility, scalability, and responsiveness to the entire CoSP value chain and thereby makes it easier to develop, launch, and provision new services supporting 5G networking. Cloud architectures are agile and enable CoSPs to flexibly deliver 5G and other networking services where and when they need them. Depending on the use case and workload type, VNFs can operate in the core cloud, at the near edge, far edge or in-between the core and the edge, depending on the use case, workload needs, or customer value.

Cloudification of the telco networks is an emerging trend in the telco industry (Section 3). It is a radical shift from the traditional siloed single-purpose hardware appliance model to virtualized network infrastructure. Telco cloud can be seen as an umbrella term that covers three different application types. The first one is the NFV cloud, which is a private cloud deployed within a telco that hosts VNFs. The Enterprise service cloud, which is a combination of in-house and third-party cloud capabilities that CoSPs offer to their enterprise customers, is the second type. The third category is a cloud deployment that supports telco-specific IT applications such as OSS and BSS along with other regular enterprise applications.

In telco cloud, each application performing a particular network function benefits from the capabilities of cloud computing such as elasticity, high availability and manageability. VNFs running on SDNs increase deployment agility. Containerized network functions (CNFs)[60] (Section 9) have further increased the density of functions per host (for example, see Glasgow Network Functions).[61] Advantages of CNFs over VNFs include cost savings due to resource efficiency and development velocity. The use of serverless CNFs is the next step in providing fault resiliency.[60]

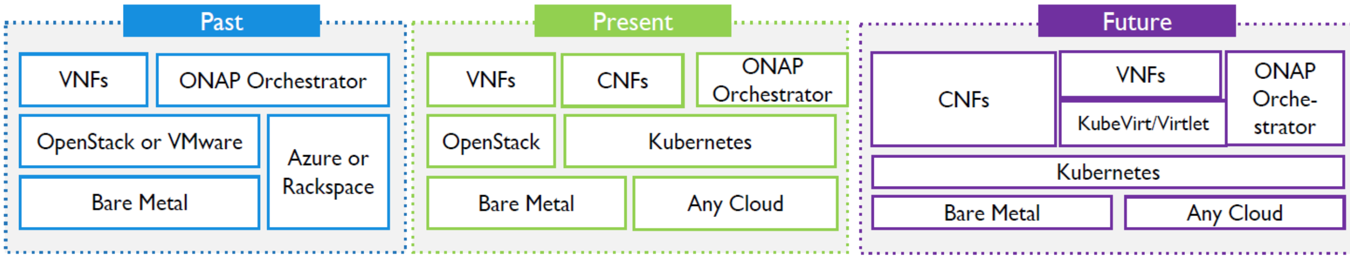


Figure 14. Evolution from VNF to CNF (Ref.60)

The 5G-driven cloudification of network architecture is a transition from separate physical boxes for each component (e.g., routers, switches, firewalls) to VNFs on VMs running on VMware or OpenStack in telco clouds[62] and to CNFs running on Kubernetes (containers) (Fig.14) in telco clouds.[60,62] In cloud transformation of telco infrastructures, unified management of heterogeneous resources provided by different vendors becomes a top priority. That is why most telcos choose OpenStack-based clouds to ensure successful cloud evolution.[63]

11.2 Collaboration of Dell and VMware with CoSPs on Telco Clouds

CoSPs need more flexibility in their data centers to handle the new services enabled by 5G. VMware recently announced expanded relationships with AT&T and Vodafone to provide a cloud-based architecture for 4G and 5G networks.[64] VMware is also working with Ericsson and T-Systems to collaborate on telco cloud transformation. Dell Technologies has joined Airship, an AT&T-led initiative to automate provisioning telco cloud infrastructure for 5G. Dell Technologies' participation in the Airship project will contribute to extending automation from cloud infrastructure software to hardware as well.

Another example of CoSP collaboration with Dell Technologies on telco cloud transformation is a memo of understanding signed by Orange and Dell Technologies to collaborate on a distributed cloud architecture to virtualize and transform the telco's network for 5G.[12,65] Orange is looking to get its network 5G ready by using NFV.

The agreement scope includes Orange Business Services in 196 countries and territories around the world. The carrier and Dell Technologies will collaborate on both technology and exploring business cases. Implementing standardized IT and cloud technologies as an alternative to fixed-function hardware will allow Orange to provide more dynamic, agile edge compute, storage, and network solutions. Technologies used will include FPGAs, GPUs, and intelligent network adapters (SmartNICs) based on ASIC, FPGA, and System-on-a-Chip (SoC) for edge workloads including cloud and virtual RANs, MEC and real-time interactive, latency-sensitive applications. A part of the agreement will be a joint effort to identify which technologies Orange needs in the future.[12]

11.3 VMware Telco Cloud

VMware Telco Cloud is the industry's only comprehensive platform optimized for both 4G and 5G.[62] It has a single architecture supporting both and offers integrated and open NFV infrastructure, SD-WAN and Service Assurance solutions. The fastest time to revenue and the lowest cost of ownership are advantages of VMware Telco Cloud solutions. VMware vCloud NFV with VMware Smart Assurance provides the platform that CoSPs need for delivering 5G-enabled services to their customers.

VMware's Telco Cloud platform is also a part of the recently announced expanded partnership between Nokia and VMware supporting the Telco Cloud transformation journey.[66] VMware also announced as a technology preview its new solution for Telco Cloud orchestration and automation: Project Maestro.[62] This new purpose-built solution complements the growing VMware Telco Cloud portfolio by providing capabilities for streamlining network service orchestration and automation for 5G networks.

12. All Cloud Strategy For 5G

12.1 All Cloud 5G Architecture

Some telecommunication companies develop technology strategy on the assumption that eventually the 5G architecture will incorporate a cloud-based digital transformation of networks, operation systems, and services – the so-called "All Cloud" architecture.[67,68] Indeed, the diversified business requirements of 5G-based services transform networks from traditional hardware-centric infrastructures towards distributed data centers and cloud-based networks. According to 5GPPP, a central cloud will be connected via a backhaul network to many edge computing clouds that are 20 kilometers from the user at most. If services can be executed in the network edge, that will reroute traffic away from the C-RAN.

At the core of the "All Cloud" strategy is the full transformation of infrastructure networks in their four domains: equipment, network, services, and operations. The strategy goal is to leverage capabilities of pooled hardware resources, fully distributed software architecture, and full automation. According to the All Cloud strategy, the entire network will shift to cloud-centric architecture, and all network functions, services, and applications will run in the cloud data centers. The next step forward is to use the cloudification concept to make network software fully distributed and automated and implement network functions cloudification (NFC).[69,70]

12.2 5G and Multi-Cloud Connectivity

The use of 5G, IoT, and edge computing requires a multi-cloud strategy.[71,72] To enable the new 5G era of connectivity, compute and storage resources provided by multiple clouds will be distributed throughout the network for realizing the low latency of 5G. New types of data centers – micro data centers – are emerging (Section 13). CoSPs are building out edge clouds and implementing VNF and CNF. A multi-cloud environment gives CoSPs service agility at greater scale and with a lower price point. Multi-cloud connectivity solutions become critical. For example, the Equinix Cloud Exchange Fabric (ECX Fabric) on Platform Equinix allows CoSPs to take advantage of 5G mobile networks when accessing multi-cloud services over low-latency connections.[73]

To attain cloud redundancy for Nokia's applications, Nokia and VMware have worked together on developing the Nokia Service Management Platform for running on the VMware Telco Cloud platform (Section 11.3). This application mobility platform has been designed for application portability, workload rebalancing, and optimized disaster recovery across data centers and clouds.[74] It enables CoSPs to embrace a multi-cloud strategy to unify network and IT environments and connect them to various clouds.

13. Data Centers for the 5G Era

13.1 Requirements for 5G Cloud Data Centers

5G-initiated transformation of cloud architecture results in changes of requirements for cloud data centers. The main trends in cloud data centers are: moving to the network edge (Section 6), the emergent concept of distributed data centers, use of micro data centers/cloudlets and ability to provide low latency for network-intensive 5G applications. The 5G era cloud data center should be a distributed, open architecture data center that is efficient, flexible, and intelligent.[75]

Open architecture of cloud data centers enables adoption of open-source northbound APIs. Delivering high throughput and ultra-low latency required by 5G is a challenge for data centers that have network-intensive workloads and storage-intensive workloads such as CDN and 4K/8K video on demand. This requires using heterogeneous computing architecture components such as GPU/FPGA, programmable intelligent network adapters (SmartNICs) based on ASIC, FPGA, and System-on-a-Chip (SoC) (Section 4), and neural-network processing units (NPUs) to ensure the data centers are energy efficient and cost effective. Offloading computing capabilities to GPU/FPGA clusters or NPUs improves cost-effectiveness of high-density computing. For ultra-high performance IOPS, next generation storage-class memory is used as the default flash storage. The important role of network slicing in 5G networks (Section 4.3) requires data center networks to flexibly orchestrate and reassemble network slices.

13.2 5G Cloudlets, Edge and Micro Data Centers

As telco infrastructure undergoes softwarization to meet 5G requirements (Section 4), it becomes a cloud with the central offices morphing into cloud edge nodes. In many cases cloud edge nodes are large enough to be classified as pod-type data centers or so called “cloudlets.” “Micro data centers” that are either integrated into or located adjacent to 5G towers[76] are an example of cloudlets.

Sizes of micro data centers can be just 6RU of rack space or include one or two server cabinets. For example, edge computing sites offered by EdgeMicro are the size of a large bedroom. EdgePoint data centers provided by Compass have a 12-rack capacity and are housed in a hardened shell.[77] Such sizes of micro data centers allow for their deployment in constrained environments such as rooftops and easements. Hence, micro data centers cost-effectively make edge computing possible in locations where traditional IT infrastructure deployment modes are unavailable and impractical or cost prohibitive to retrofit. It is critical for micro data center size that it should fit to the load size of an 18-wheeler. As micro data centers are geographically dispersed, the ability to manage them remotely using a single management framework becomes critical.

With the increasing deployments of MEC for the future 5G, the micro data center market is expected to grow to USD 14.5 billion by 2025 (a compound annual growth rate of about 25% from 2019 to 2025).[78,79] While startups are attempting to gain a share in the market of edge modular micro data centers, already established leaders in data center technologies such as Dell Technologies are developing edge cloud solutions.[79] We review the Dell Technologies solutions in the next two subsections (see also Section 6.5).

13.3 Dell Rack Servers in Telco Cloud Data Centers

The mainstream servers in the PowerEdge portfolio from Dell Technologies (Section 6.5) can be used to implement C-RANs and edge data centers.[41] For example, the R640 is a 1U two-socket rack server having NEBS3-compliant hardware and providing a high degree of computing density. NEBS (network equipment-building practice) compliance is a key feature for hardware that must endure varying and tough physical conditions. The carrier-grade server also offers extended temperature operation. The R740 2U server has also been NEBS3 certified and offers a high degree of expandability for meeting growth in networking bandwidth. The XR2 platform is a smaller 1U ruggedized server and can be used in locations with additional environmental constraints.

13.4 Dell Edge Micro Modular Data Centers

Server products alone are not enough for creating network edge installations. The ESI (Extreme Scale Infrastructure) team at Dell Technologies has developed micro modular data centers (MDC) well-suited for deployment in locations where preexisting server facilities may not be available.[80,81] One of the compelling attributes of these MDCs, featuring built-in cooling and security, is that they can house standard servers in these undeveloped locations.

MDCs provide a self-contained design with cooling, power, and dedicated enclosure to secure deployments in remote locations, some of which may be unprotected. MDCs are designed to be flexible and can host a variety of Dell Technologies servers, networking, and storage systems to accommodate diverse 5G edge workloads. Micro-MDC is a pre-integrated data center that can quickly deploy at Near and Far Edge sites (Section 6.4).

Figure 15. Dell Technologies Edge Modular Data Centers (Ref.80)

.

14. 5G Security and Cloud Platforms

The size limitation of this article does not allow for an overview of 5G security in MCC, MEC and other 5G cloud platforms. As mentioned in Section 4.3, 5G security benefits from network virtualization features such as network slicing. However, 5G cloud platforms, for example MCC and MEC, are a complex combination of diverse technologies including wireless networking, distributed computing, and the virtualized network and server environments. This increases the cyber-attack surface with various cyber-attack vectors. Even making each technology component individually secured, there is no guarantee that the entire MEC system is secure because of the complex dependencies among them. For example, security mechanisms in network slicing need to ensure strict isolation between different network slices running on shared infrastructure. This is necessary to prevent VMs in one slice from impacting those in other slices. Another example is a need for developing SDN-specific security mechanisms as the connectivity within and between 5G data centers is expected to be managed via SDN (Fig. 12). Therefore, automated holistic security management will become crucial in cloud-based 5G mobile network deployments.

Collaboration between the 5G network security product vendors and data center solution providers is becoming more common. For example, Dell Technologies recently executed a global reseller agreement with A10 Networks providing AI-driven network and application security.[82] This agreement enables Dell Technologies to resell the A10 full product line of multi-cloud and 5G security solutions.

For a review of efficient 5G security strategies for MCC and MEC, the readers are referred to Refs. 83 and 84.

15. Conclusion

To meet requirements of emerging 5G services, telecommunication companies need to dramatically change their network infrastructure. The key directions of this transformation are network virtualization, SDN, network slicing, C-RANs, V-RANs, network function disaggregation, developing cloud-native applications (e.g., containerized network functions), all of which fall under the umbrella of so-called *network softwarization*, a term used interchangeably with *network cloudification*. This is exemplified by cloud-native networking models providing cloud-type agility to the network services. These new features lead to an interplay of two different trends influencing the developments of 5G networking and cloud computing as they become intertwined. The 5G advent results in emphasizing the role of edge cloud computing (MEC), moving data processing from the central data center closer to the edge/cloudlets, where data is generated and used by low-latency applications.

However, 5G technologies also transform network architectures in the opposite direction to using centralized compute/data-intensive processing of Big Data and AI applications at the core cloud because of the vast amounts of data generated by 5G-based applications. Both core and edge data centers/cloudlets are amalgamated into the same distributed cloud computing system. Indeed, while an edge data center processes data generated at the edge, it offloads compute-intensive processing to the core data center. Hence, 5G technologies make a paradigm shift in the cloud computing landscape by transforming it from the traditional model of cloud computing with remote sites using cloud as a hyper-scalable centralized virtual data center into a complex diverse system of distributed cloud computing composed of core cloud, near edge, far edge, device edge subsystems, and C-RANs. The use of multi-cloud services also grows in 5G networking deployments.

What trends in cloud computing transformation driven by 5G will dominate? It is a difficult question because the answer depends on new application types that 5G will enable. The 5G use case list is long: IOT, ultra-reliable and low latency communications, self-driving vehicles, industrial automation, facial recognition, machine-to-machine communications (Section 2.3), and development of new types of applications is in the pipeline. The diversity of these applications and use cases result in a wide variety of requirements for latency, reliability and processing the volume/type of data traffic generated. This makes the cloud computing transformation initiated by 5G development a complex multidirectional and multifaceted process. I hope my article painting a picture of how 5G technologies shape the cloud computing evolution will help the readers in transforming their cloud computing services to enable transition to 5G technologies.

16. References

1. Setting the Scene for 5G: Opportunities & Challenges. International Telecommunication Union, 2018
2. Securing 4G, 5G and Beyond with Fortinet: https://www.fortinet.com/solutions/mobile-carrier.html
3. http://broadbandlibrary.com/5g-low-latency-requirements/
4. Spectrum for 4G and 5G. Qualcomm Technologies, December 2017.
5. http://www.arrow.com/en/research-and-events/articles/what-frequency-spectrum-will-5g-technology-use-and-how-does-this-compare-to-4g
6. http://www.telekom.com/en/company/details/5g-speed-is-data-transmission-in-real-time-544498
7. DELL EMC and 5G. Analysis and strategy to capture the 5G mobile opportunity, White Paper, Dell, 2018.
8. http://www.forbes.com/sites/jonmarkman/2019/01/27/supersonic-5g-wireless-could-kill-the-cloud/#2b3eae442207
9. R. Buyya, R. N. Calheiros, J. Son, A. V. Dastjerdi and Y. Yoon, Software-Defined Cloud Computing: Architectural elements and open challenges, 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014, pp. 1-12.
10. Vincent W. S. Wong, Robert Schober, Derrick Wing Kwan Ng and Li-Chun Wang. Key Technologies for 5G Wireless Systems. Cambridge University Press, 2017.
11. Partnering on The Journey to 5G, Dell EMC Webinar, SDxCentral, 2019,
12. http://corporate.delltechnologies.com/en-us/newsroom/dell-emc-and-orange-collaborate-for-telco-multi-access-edge-transformation.htm
13. http://www.rcrwireless.com/20190430/5g/5g-multi-cloud-digital-transformation
14. DELL EMC Ready Solution for VMware NFV Platform. Dell White Paper. H17238.2, 2018.
15. Dell EMC Ready Solution for VMware vCloud NFV 3.0 OpenStack Edition Platform. Dell White Paper. April 2019.

16. http://5g-ppp.eu
17. E. J. Oughton, Z. Frias, S. van der Gaast, and R. van der Bergc. Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of the Netherland. Telematics and Informatics. Volume 37, April 2019, pp. 50-69.
18. A. N. Toosi, R. Mahmud, Q. Chi and R. Buyya. Management and Orchestration of Network Slices in 5G, Fog, Edge and Clouds. Fog and Edge Computing: Principles and Paradigms. Wiley Telecom, 2019.
19. http://www.cisco.com/c/en/us/solutions/enterprise-networks/network-as-service-naas.html
20. http://www.cisco.com/c/en/us/solutions/enterprise-networks/sdn-cloud-computing.html
21. http://www.sdxcentral.com/5g/definitions/5g-sdn
22. Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi. Wireless network cloud: architecture and system requirements. IBM Journal of Research and Development 54.1 2010, p. 4:1–4:12.
23. http://www.sdxcentral.com/5g/definitions/cran/
24. M. Peng, Y. Sun, X. Li, Z. Mao, and C, Wang. Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues. IEEE Communications Survey & Tutorials. 2016.
25. A. Maurya. 5G Technology and Its Key Drivers: Cloud Computing and Cloud Radio Access Networks. International Journal of Advanced Computational Engineering and Networking, 2015, vol. 3, pp.89-94.
26. http://www.intel.com/content/www/us/en/communications/products/programmable/applications/baseband.html
27. http://blog.equinix.com/blog/2019/06/19/5g-and-the-future-of-service-delivery/
28. http://www.windriver.com/news/press/pr.html?ID=22531
29. http://www.computerweekly.com/news/252474367/Wind-River-teams-with-Dell-for-5G-distributed-edge-system
30. https://www.ibm.com/cloud/learn/what-is-mobile-cloud-computing
31. Advances in Mobile Cloud Computing and Big Data in the 5G Era. C, Mavromoustakis, G. Mastorakis, C. Dobre (Eds.). Springer, 2017.
32. H. T. Dinh, C. Lee, D.t Niyato and P. Wang. A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. 2013; 13:1587–1611.

33. . M. Umar, J. Ferzund, T. Sajjadand, and S. M. Owais. Dealing with Issues of Mobile Cloud Computing using 5G Technology. International Journal of Computer Science and Network Security, 2017, V.17 No.8, p.246.
34. http://blog.dellemc.com/en-za/dell-technologies-vmware-5g-rocks-the-network-edge
35. http://www.sdxcentral.com/articles/sponsored/cto-ihab-tarazi-dell-technologies-edge-changing-networking/2019/10
36. http://blog.dellemc.com/en-us/edge-computing-driving-force-new-architecture-innovation
37. http://www.winsystems.com/cloud-fog-and-edge-computing-whats-the-difference
38. http://dzone.com/articles/cloud-vs-fog-vs-edge-computing-3-differences-that-matter
39. http://stlpartners.com/edge-computing/5g-edge-computing
40. http://www.cisco.com/c/en/us/solutions/enterprise-networks/edge-computing-architecture-5g.html
41. Dell EMC Network Edge Reference Architecture for VMware v Cloud NFV OpenStack Edition 3.1. Dell EMC Service Provider Solutions. 2019.
42. http://www.dellemc.com/en-us/service-providers/edge-computing.htm
43. http://blog.dellemc.com/en-us/expanding-edge-portfolio-software-defined-networking/
44. http://www.etsi.org/technologies/multi-access-edge-computing
45. http://www.juniper.net/us/en/products-services/what-is/multi-access-edge-computing/
46. Derrick Wing Kwan Ng, Li-Chun Wang. Key Technologies for 5G Wireless Systems. Cambridge University Press, 2017.
47. Xu Chen, Lei Jiao, Wenzhong Li, and Xiaoming Fu, Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing, Oct. 2015.
48. http://www.rcrwireless.com/20190614/5g/5g-oriented-edge-computing-has-arisen-as-the-time-requires
49. The Network Fabric for Distributed Cloud and 5G. Solution Overview. Pluribus Networks, 2019.
50. G. Brown. 5G Service Velocity & the Distirubuted Cloud. Heavy Reading White paper. 2019.
51. S. Sathyanarayan. How cloud-native core network helps fulfill the demands of 5G services. Affirmed, 2019.
52. http://www.sdxcentral.com/articles/news/nokia-argues-cloud-native-is-essential-to-5g-core/2019/11
53. http://www.metaswitch.com/knowledge-center/reference/what-is-the-5g-service-based-architecture-sba

54. http://aws.amazon.com/outposts/
55. http://aws.amazon.com/wavelength/
56. http://news.microsoft.com/2019/11/26/att-integrating-5g-with-microsoft-cloud-to-enable-next-generation-solutions-on-the-edge
57. Partnering on the Journey to 5G, Dell EMC Webinar, SDxCentral, 2019. https://www.sdxcentral.com/resources/sponsored/webinars/dell-emc-5g-telco-service-providers
58. http://www.dellemc.com/en-us/service-providers/telco-cloud-solutions.htm#
59. Building the New Cloud Economy with VMware Telco and Edge Cloud. VMware, 2019.
60. http://events19.linuxfoundation.org/wp-content/uploads/2018/09/Evolving-Cloud-Native-Landscape-Dec-2018-ContainerDays-Japan.pdf
61. http://netlab.dcs.gla.ac.uk/projects/glasgow-network-functions
62. http://telco.vmware.com/
63. http://www.lightreading.com/cloud/network/leading-lights-2019-finalists-telco-cloud-strategy-/d/d-id/750952
64. http://about.att.com/story/2019/att_dell_opensource5g.html
65. http://virtualizationreview.com/articles/2019/05/14/dell-tech-orange.aspx
66. http://ir.vmware.com/overview/press-releases/press-release-details/2019/Nokia-and-VMware-expand-partnership-to-ease-large-scale-multi-cloud-operations/default.aspx
67. http://www.telcotransformation.com/author.asp?section_id=390&doc_id=737952
68. http://carrier.huawei.com/en/industry-insights/all-cloud-journey-to-digital
69. http://atos.net/en/blog/network-function-cloudification
70. http://blog.mellanox.com/2015/03/from-network-function-virtualization-to-network-function-cloudification-secrets-to-vnf-elasticity
71. http://www.rcrwireless.com/20190311/5g/the-journey-to-5g-living-in-a-multi-cloud-world
72. http://www.rcrwireless.com/20190430/5g/5g-multi-cloud-digital-transformation
73. http://blog.equinix.com/blog/2018/07/17/the-role-of-5g-networking-and-cloud/
74. http://futurumresearch.com/nokia-and-vmware-strengthen-ties-to-smooth-multi-cloud-operations
75. http://www.actualtech.io/5g-will-change-data-centers
76. http://www.thegreengrid.org/ja/newsroom/blog/getting-speed-5g-strategy-and-micro-data-centers-steven-carlini-schneider-electric
77. http://www.networkworld.com/article/3445382/10-hot-micro-data-center-startups-to-watch.html

78. http://www.marketwatch.com/press-release/micro-data-center-market-comprehensive-analysis-growth-forecast-from-2018-to-2025-2019-06-07
79. http://www.gminsights.com/industry-analysis/micro-data-center-market
80. Dell EMC Network Edge Reference Architecture for VMware v Cloud NFV OpenStack Edition 3.1. Dell EMC Service Provider Solutions, Dell 2019.
81. A. Butcher and J. Boccuzzi, Exploring Computing Platforms for Radio Access Networks. Dell EMC White Paper, 2019.
82. http://www.a10networks.com/blog/a10-networks-and-dell-emc-partner-to-deliver-5g-security-and-multi-cloud-solutions-to-customers
83. D. Chandramouli, R. Liebhart and J. Pirskanen. 5G for the Connected World. Wiley, 2019.
84. P. Marsch, Ö. Bulakçi, O. Queseth and M. Boldi. 5G System Design: Architectural and Functional Considerations and Long Term Research. Wiley, 2018.