accenture consulting

MULTI-ACCESS EDGE COMPUTING FOR PERVASIVE NETWORKS

This white paper addresses use cases driven by the Multi-access Edge Computing (MEC) architecture, with focus on the ultra-low latency enabled by this 5G technology. It also provides an overview of architecture and standardization activities in this area. Finally, it describes the MEC architecture's key challenges and benefits for communications service providers (CSPs).

1.0 INTRODUCTION TO MULTI-ACCESS EDGE COMPUTING

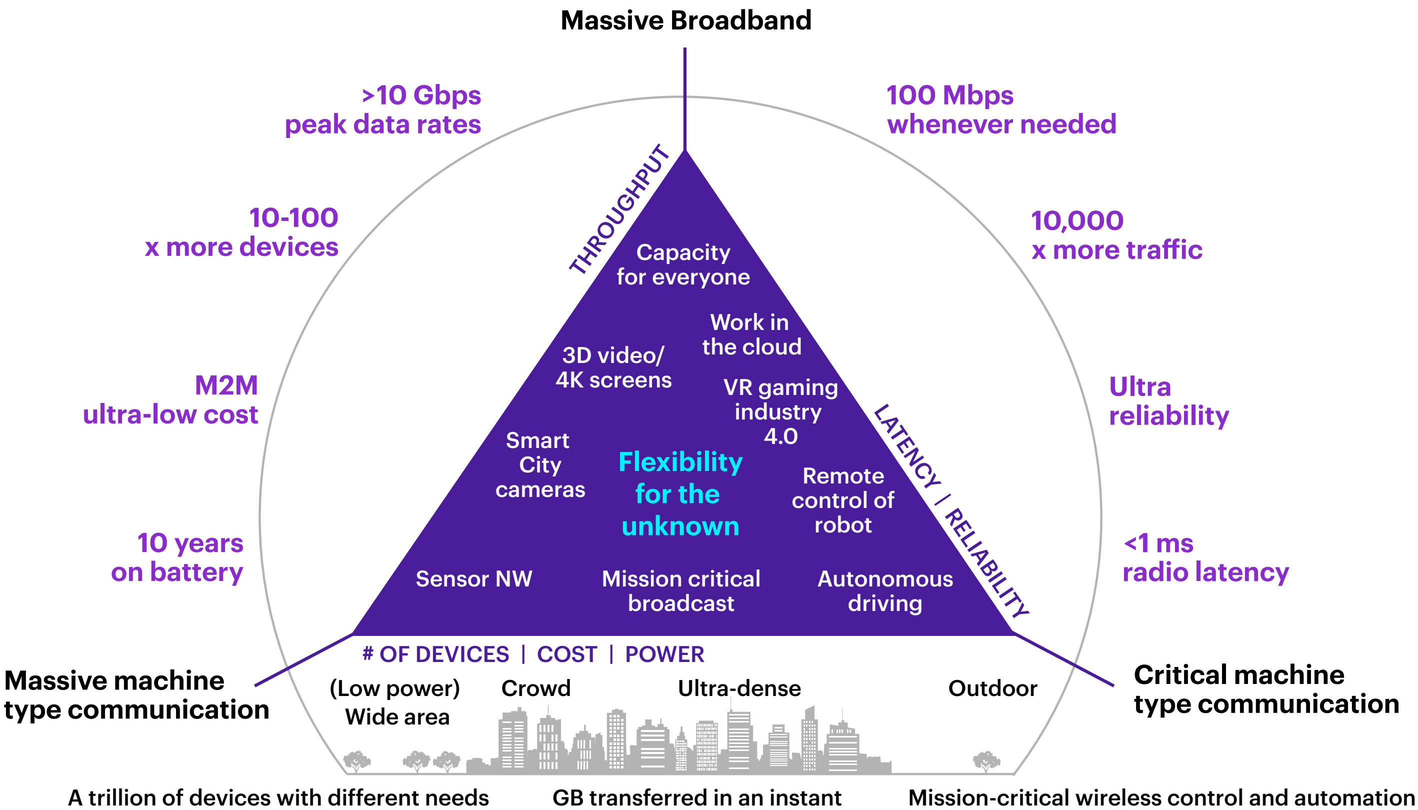
MEC transforms the topology and architecture of mobile networks from pure communication networks for voice and data to an application platform for services. Against this backdrop, the MEC should be seen as complementing and enabling the service environment that will characterize 5G.

While previous generations of mobile communication networks, such as 2G, 3G and 4G, focused primarily on media interaction – connecting people across different places – 5G will enable multiple use cases, including working in the cloud, remote control of robots, VR gaming, AR for predictive maintenance, autonomous driving and automation. Figure 1 illustrates a number of these use cases, including:

I. **V2X and Connected Car.** This concept aims to increase transportation safety, improve travelers' experience, decrease environmental impact and improve transportation efficiency. One of the flagship use cases for MEC, it provides benefits for various players, including CSPs [1].

II. **Industry 4.0.** Industry 4.0, known also as Industry X.0, has the aim to use deep network connectivity features to digitize the industry processes, and embrace ICT and connectivity to increase productivity and efficiency [2]. Robotics for manufacturing, IoT for logistics and deep adoption of sensors for precision agriculture will transform the production model for multiple industries. With 5G MEC, a platform will be provided to remotely connect, control and monitor machines, such as mini robots and drones. In the case of agriculture, it will be possible to harvest fields, with autonomous driving tractors used to transport these smaller devices. As a result, areas of agriculture will become much faster, more efficient and more highly productive. By 2021, the "connected cow" and farm business, currently more than 1 billion EUR in revenue, is expected to grow eightfold, to almost 10 billion EUR [3]

FIGURE 1:
The 5G triangle, which illustrates potential use cases considering the main capabilities of the new generation of mobile networks [2].



III. **e-Health.** MEC can help enable more robust communication, higher-speed data rates and lower latency for such uses as ambient assisted living, elder care, or more sophisticated prosthetics supported by environmental sensor data. In general, health care systems can benefit from preprocessing patients' data in an edge cloud, before medical experts diagnose the patients remotely. Also, the outsourcing of computation facilitates interaction between special-care patients and their surrounding technology. However, as reported by WHO, it is currently difficult to quantify the level of adoption of MEC-driven services [4].

IV. **Enhanced experiences at public venues.** Initial studies and experiments have received very positive feedback from end users. Use cases include locally produced live video streams at stadiums, with replays from different camera angles, and AR and VR experiences at such venues as shopping malls and amusement parks. More than 67% of CSPs consider important, or even critical, the enablement of large public-venue services driven by access edge transformation [5].

Each of these use cases require the proper mix of service qualities that can be classified into throughput, reliability and latency. Within 5G standardization they are referred to as MBB (mobile broadband), MTC (machine type communication) and URLLC (ultra-reliable low-latency communication). For example, for autonomous driving, where vehicles must communicate with each other to predict dangerous scenarios and thus prevent accidents; URLLC is needed to ensure a failure probability of less than 10^-8 and a latency of less than 1ms. In VR use cases, one additional requirement is evident: the need for higher bandwidth. Out of these requirements, we believe that MEC will contribute mainly to URLLC to achieve the necessary service requirements in an affordable and sustainable way.

The fundamental premise for the adoption of MEC in the target network architecture, and its success in that context, is the need to have network-mediated transactions for future innovative services. This assumption has multiple implications and prerequisites:

- **Service.** Requirements for latency, but especially for resiliency and security, will mandate a network-mediated model.
- **Technology.** The network will be better suited to embed some key features, such as session continuity, caching or distributed transaction management, than for the application or device layers.
- **Business.** Telcos will aim to regain a central place in the value chain by offering connectivity services that are instrumental for the implementation of new services – e.g. with the SLAs and open APIs required to enable integration.
- **Regulatory.** The network-mediated model will be allowed by regulators, as it will not be seen as a violation of the network neutrality concept, but rather as a way to monetize the network and fund the business case for the creation of a leading-edge infrastructure.

2.0 MULTI-ACCESS EDGE COMPUTING ARCHITECTURE: FUNDAMENTALS

2.1 MULTI-ACCESS EDGE COMPUTING: WHAT IT IS, AND WHY IT IS REQUIRED FOR 5G

A simple solution to provide URLLC is to bring all the computation, and thus services, into the mobile device. The technology evolution of these devices, and the growth in their computing power, will make possible the requisite performance and storage. What is missing from such implementations, however, are several factors: the resiliency that a central distributed architecture can provide; the interoperable security that an open ecosystem requires; and the agility needed to deliver service upgrades in one single place.

The massive, alternative-architecture, centralized data centers in the backbone of mobile communication networks, known also as cloud services platforms, will perform adequately in providing computation and other services. Unfortunately, when considering some 5G use cases which require 1ms URLLC, this concept has a huge disadvantage. A signal or data packet which needs to travel between a mobile device and the cloud usually travels several hundred kilometers to be processed in the data center, which leads to a round trip time (RTT) of several dozen milliseconds.

To put this in numbers, consider that the propagation delay of a signal or data package within a modern fiber-optic network is limited by the speed of light, which is around 300.000 km/s in vacuum. To achieve a round-trip time (RTT) of less than 1 ms, the maximum distance between a mobile device and a cloud can be no greater than around 100km. Considering that additional time needs to be allowed for sensors to collect data, for embedded or cloud computing to take place, for in-network elements to be processed, and for the actuator to react, then the maximum distance to achieve 1 ms RTT becomes even lower, and should be not more than 10 to 20 km.

From the user's point of view, data centers provide services and are in essence the core or the "heart" of the internet (network). It is not feasible to build big data centers every 10 km to bring the internet to the user more quickly. From the network operator point of view, however, the closest places or "hops" to a user are the base stations at the edge of the network. Those base stations can be utilized to outsource the cloud closer to the user, and therefore, to provide low latency and enable new use cases.

The concept of “edge computing” or the “edge cloud” describes the use of the network’s edge to provide cloud services, as well as one additional aspect. Due to the mobility of end user devices, a possible edge cloud must also be mobile or, more precisely, agile. This means that an application or service running in an edge cloud needs to be migratable without any visible downtime for the user. Because an agile/mobile cloud can now be accessed through different endpoints, it is therefore referred to as a mobile or MEC.

Since URLLC is a key requirement of 5G, the concept of the MEC becomes one of the most important tools of 5G. For example, while an autonomous vehicle is in motion, the location of the controller has to move, too; or at least there is the need for a fast handover to another controller. The concept of the MEC can also be extended to other mobile networks than the Radio Access Network (RAN). In a factory, a MEC could be located near the WiFi access point (AP). In general, it is assumed that only air exists between the device and the MEC. In the failover case, a MEC could also be used to ensure that the downtime of a service becomes very low. Therefore, an MEC not only can provide low latency, but can also help to increase reliability and security.

One of the most important points to consider when creating the MEC is that it is not economically feasible to place huge servers at every base station, which means that the computational capacity at the base station will also be limited. Ultimately, it is impossible to completely replace big data centers with the MEC; but the boundaries between the MEC and big data centers do need to be softened, meaning that not only does computing hardware need to be deployed at the edge of the network, but also that the whole network needs to become more “intelligent.” In addition, a multi-tier hierarchical concept needs to apply, in which moving farther to reach the core network results in greater computational capacity; this can be referred to as fog computing.

Depending on the requirements of the service being demanded, or more precisely, of all services being demanded in a radio cell, the applications need to be more fully distributed in the network. For example, an MEC at the base station could just be used for some simple pre-computation to reduce the bandwidth required in the backbone network, with the actual processing being done deeper in the network.

2.2 STANDARDIZATION AND ECOSYSTEM

Numerous institutions have become involved in standardizing the components of the MEC, starting in 2014 with the European Telecommunications Standards Institute (ETSI), which evolved into its own Industry Specification Group (ETSI MEC ISG) [6]. To date, it has published specifications about technical requirements, framework and reference architecture, several APIs, and other aspects. Although not directly focused on technology, the Internet Engineering Taskforce (IETF) [7] established groups for discussing the MEC in the context of internet technologies.

For the North American market, the Institute of Electrical and Electronic Engineers (IEEE) is involved. On an international level, the 3rd Generation Partnership Project (3GPP) streamlines proposals, illustrated in Figure 2, from different standardization organizations and interest groups (with ETSI, ATIS and GSMA among them) to create a generally accepted standard. The 3GPP Release 15 will contain the first set of 5G technologies.

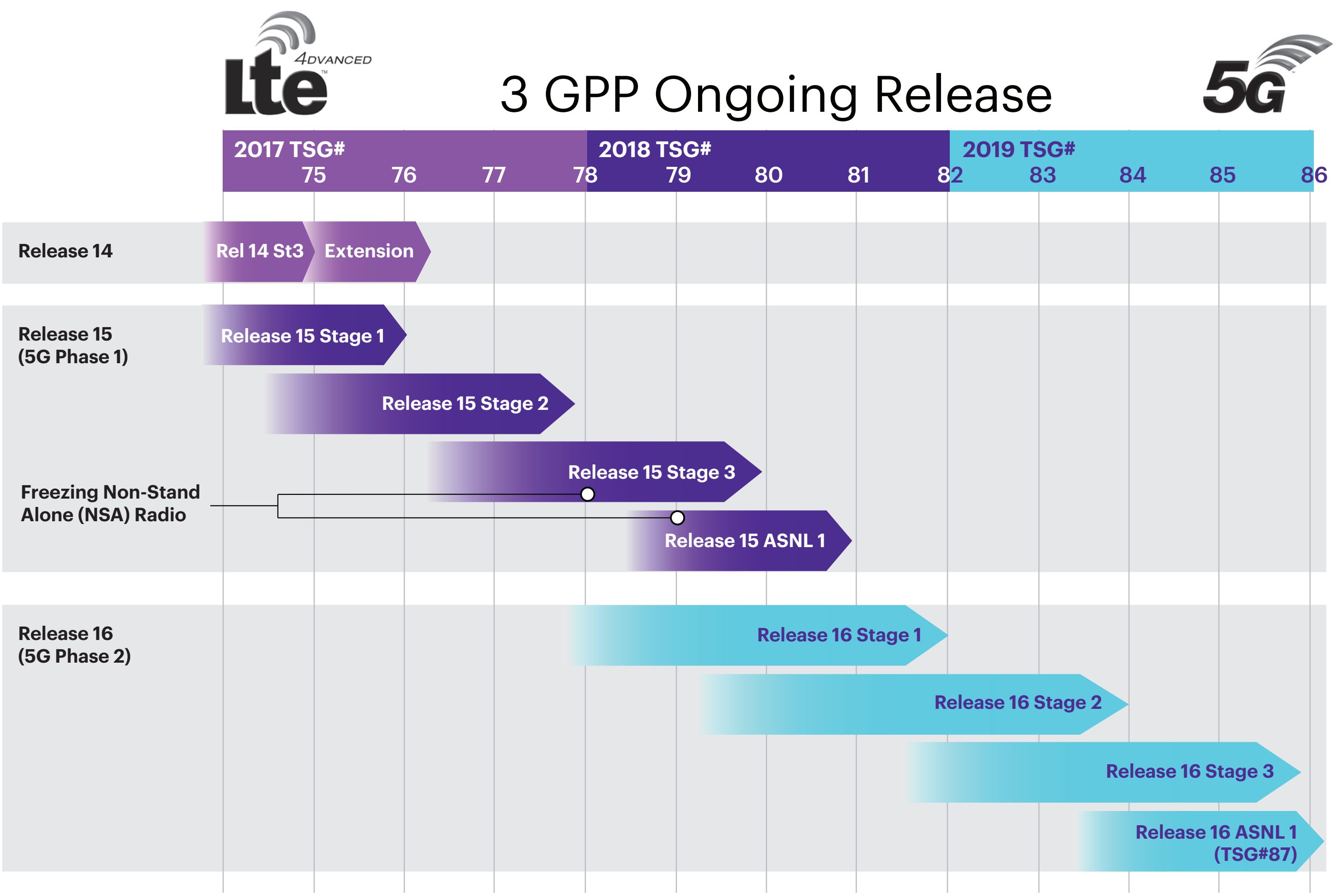


FIGURE 2:
As shown below, the release timing for 5G phase 1 and phase 2 is currently fixed. Standardization will take place in Release 15 and 16 respectively, based on the outcome of the Release 14 study [8].

2.3 ARCHITECTURE OVERVIEW

The MEC requires three enabling technologies to be implemented: Network Function Virtualization (NFV), Software Defined Networks (SDN) and Edge Cloud, whose relations are shown in Figure 3. The following sections give a brief overview of the MEC's hardware, services, and network requirements, so that their relationship may be better understood.

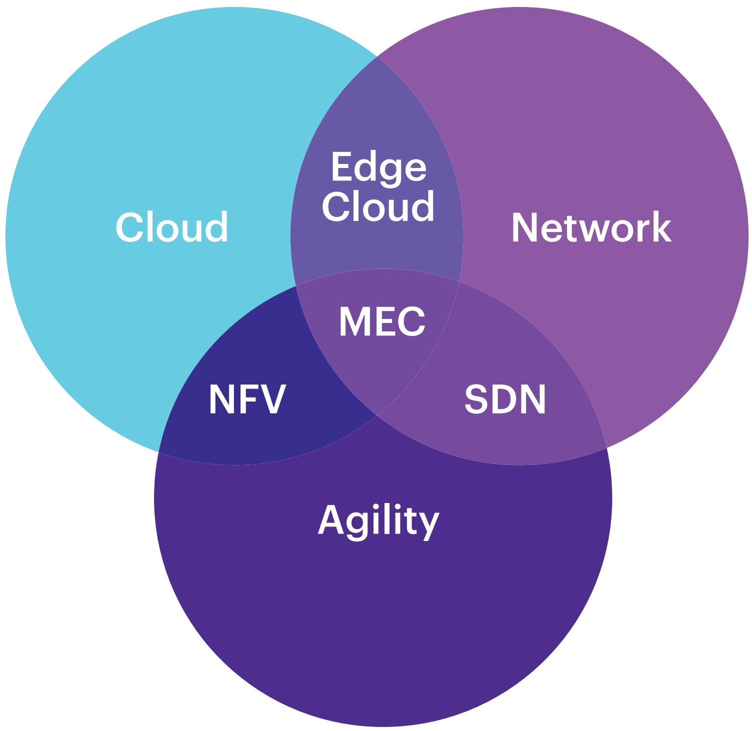
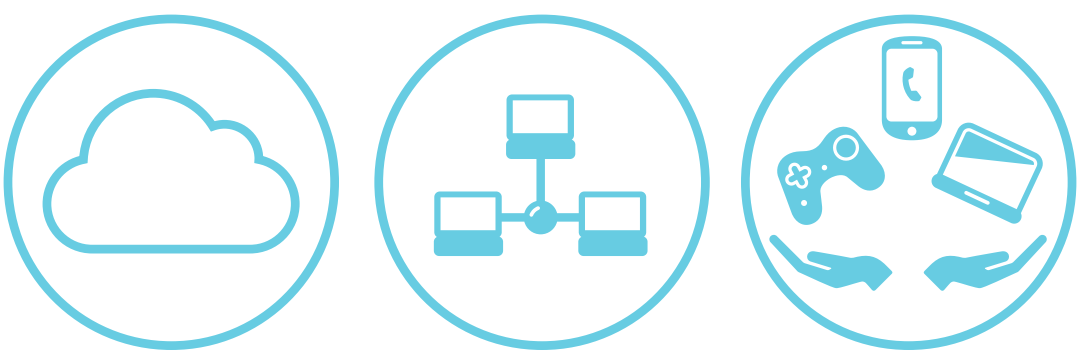


FIGURE 3:

The intersection of NFV and SDN forms the MEC.



2.3.1 HARDWARE

It is quite obvious that there will not be a big cloud center at every base station. Nevertheless, to help ensure wide availability, every edge location should be able to run a limited number of services with the fidelity of a data center. The computing performance between different nodes or even operators should be comparable. Even more important is that different services need to be run without influencing each other in terms of hardware resources; and of course, the base station functionality should not be disturbed. Since the services running on an edge cloud are very agile, the edge location needs to have a performant network connection for fast migrations and handovers to other locations.

Furthermore, the concept of an MEC is not binary. This means that if the requirements of a specific service allow it to be deployed farther away from the user, the cloud can also be located deeper in the network, to obtain economies of scale. The handover between different edge clouds, and the concept of edge, fog and cloud computing, is shown in Figure 4.

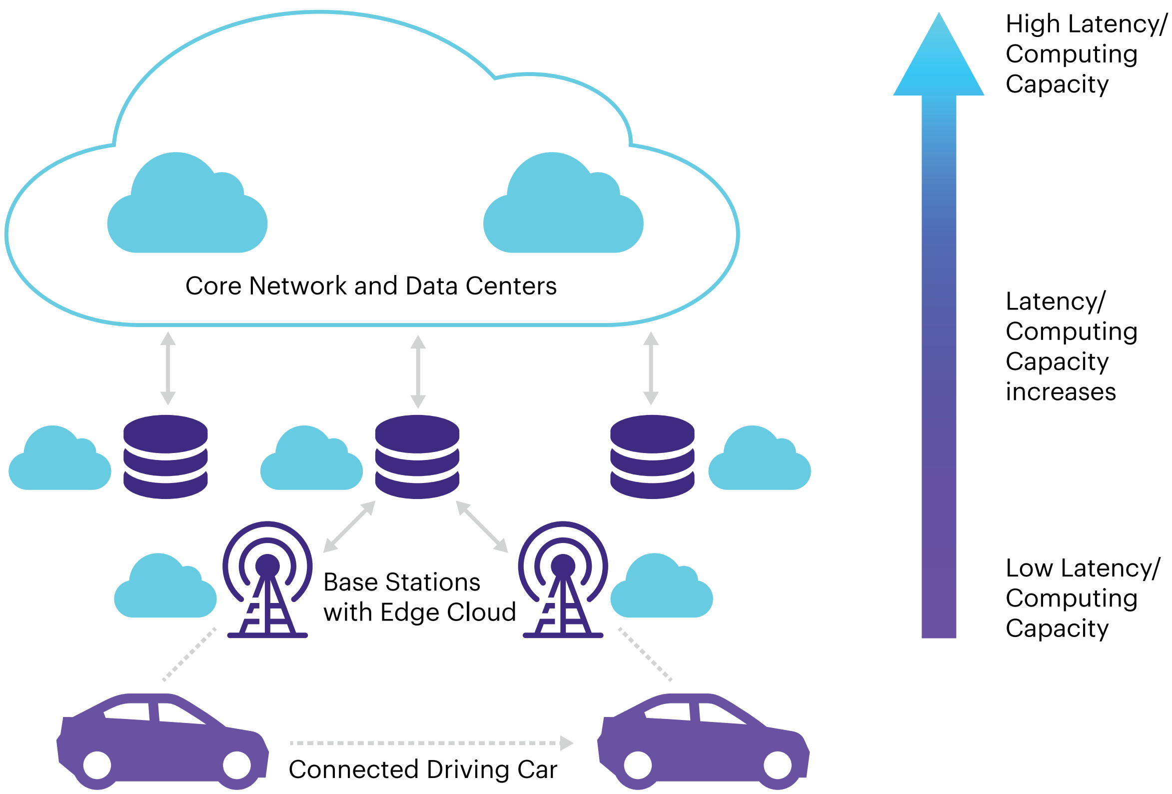


FIGURE 4:
The concept illustrates how the MEC moves, if a connected car is driving. Depending on the service level, the latency and computing capacity increases.

2.3.2 SERVICES

The hardware limitation at the edge comes with some challenges for applications and services. For instance, there is the need for an orchestrator that manages the deployment and migration of services within the network. The orchestrator needs to have an overview of all the services which are currently running in the network, as well as the network's remaining hardware capacities. If a service wants to run in the network, it needs to perform a service agreement handshake with the orchestrator. In the first step, the service would send to the orchestrator a request containing its requirements for latency, computational power, storage, throughput and other factors.

To fulfil the latency requirement, it is also important for the orchestrator to know the physical position of the service-requesting entity. Depending on the current capacity of the network hardware of the edge cloud, the orchestrator can either satisfy the requirements, even if this means migration to some other services, or cannot meet them. In the first case, the orchestrator will answer the service regarding where it can be deployed, and the service can then agree to that. In the second case, the orchestrator can send a counterproposal or deny the request. When a counteroffer is sent, the service can either accept the offer and run with less functionality, or deny it. Depending on the application, it is a huge advantage to separate the state of a service from the actual engine. This helps in lowering the actual handover time, and therefore in decreasing possible downtimes. One sample use case for this is connected autonomous driving.

Furthermore, services, especially computationally intensive ones, can increase the utilization of the hardware or network and the throughput when they are designed as several independent functions, so that they can run in a service function chain.

2.3.3 NETWORK

The client must be agnostic of network knowledge. The orchestrator will manage the routing with the help of SDN. If the user changes its location, then the location of the corresponding MEC also needs to be adapted. From the user or service point of view, nothing will change when this happens. As the location changes, the SDN controller will change the routing between the user and the MEC.

Compared to the use of data centers with fixed network conditions, this handover is more complex, since it needs to happen on both sides of the connection, instead of only on the client side. Also, network functions such as firewalls or deep-packet inspection should either be instantiated together with the MEC, or else deployed in a manner that is as dynamic as the service itself. So, the MEC as a whole can also be considered as a virtualized network function – i.e., one established through NFV – and the usage of network slicing can guarantee specific network parameters for the service provider.

3.0 CHALLENGES AND BENEFITS FOR COMMUNICATION SERVICE PROVIDERS (CSPS)

MEC introduction will be based on a trade-off between two requirements: an operator will target the minimum number of MECs that will enable the offering of dimensioned density of required 1ms services, with sufficient coverage. Depending on the use cases driving the business case behind the MEC, we expect that MEC infrastructure will be deployed in a phased manner, first addressing quick wins with limited coverage density, and potentially evolving in terms of latency (close or co-located with eNB) and coverage.

The operator will need to be able to intelligently predict both the density of MEC-needed services and their capacity. With smart planning models, an operator needs to be able to bring intelligence into the planning, predicting both the usage of 1ms services and the number of users for which an MEC is needed. Since the investment in MEC will be significant, and an intelligent approach to its dimensioning is beneficial, it is crucial to come up with smart models to perform this task.

The benefits of the MEC will depend on the modularity of the architecture and the accessibility of the MEC capabilities by various services. Since the end service provider, the MEC provider and service functionality vendors are likely different entities, the previously mentioned Standard Development Organizations (SDOs), alliances and partnerships will play a crucial role in assuring the openness of the interfaces and the modularity of the architectures, to ensure that the overall ecosystem can rapidly grow and that new services can be easily instantiated.

Operation of the future network, including the MEC, may bring additional complexity to the networks. Networks will gradually evolve towards 5G with MEC, which means that current components from various Operations Support System (OSS) domains (Plan/Build, Assure, Configure, etc.) will not completely disappear. Having this in mind, an operator will need to operate a complex OSS landscape involving the existing OSS and the new, real-time OSS. MEC and its orchestration need to be fully integrated into OSS, with an overall orchestration enabling full visibility from the control point of view. In addition, an MEC will collect a huge amount of valuable data that should be properly used by analytics engines, with the aim of enabling new service offerings, improved network optimization and better planning of further MECs. Because its information will have to be used in near-real-time network optimization, the MEC needs to be integrated into future self-optimizing networks and orchestration solutions.

There will be cases where an optimization of the RAN resources is needed within milliseconds so that the MEC can play an important role in improving the overall performance of future Ultra-high Density (UHD) networks. One example could be bandwidth on request for location-based services, where a user utilizing MEC service needs more bandwidth to meet the required Quality of Service (QoS), and the OSS needs to be able to perform the quick RAN resource reassignment. In addition, MEC will have to be able to operate multi-operator applications, such as connected-cars apps, that cannot be provisioned by only one operator.

Even though many challenges are evident when introducing MEC, the benefits it brings make these challenges worth addressing. These benefits include reductions in latency; huge QoS improvements; the ability to offer new, faster, location-based services; security improvements; and the ability to keep traffic local and closer to the user. For an operator, processing the information closer to the user opens up an entirely new range of possibilities. Using MEC, the operator can better control the infrastructure and its revenue by designing and offering new services specifically shaped for MEC usage. With this approach, an operator can also limit the disruption caused by Over the Top (OTT) players, and tailor the services specifically for MEC in order to have a faster return of investment. Additionally, processing the information closer to the user will help enable more capacity in the backbone/core that can be used for a better utilization of the operator´s core processing capabilities and for network optimization. Since superfluous information in the core could be an unwanted drawback for a high-performing network of the future, it makes sense to offload the core using this approach. Benefits are not only visible to CSPs in improved QoS, but also to other stakeholders, such as venue owners, who can provide faster, more responsive services, or companies that offer low-latency Internet of Things (IoT) services.

We believe that mobile operators will continue investing in MECs and driving the funding of MEC deployments. MEC is becoming an integral part of future networks, and it makes sense for operators to continue driving the integration of MEC. Also, because MEC is not exposing the services to the Internet and to third parties as today´s network is doing, it gives operators an improved opportunity to control revenue share by restricting the disruption caused by OTTs, who will have to work more closely with operators if they want to participate in this revenue stream. In this scenario, an operator would own the MEC and would rent its access to the third party, such as a vehicle manufacturer who could improve current location-based service offerings and offer new ones.

Another option could be for several operators to jointly invest in MECs and share the resources, using an approach similar to today’s antenna tower-sharing scenarios. This is probably a better option from a CAPEX perspective, because the application users will come from different CSPs; however, it has a drawback in terms of capacity utilization, security and controlling the interoperability between different networks.

The option of a shared model between an operator and industry partner, such as a vehicle manufacturer, would be also an interesting option for an operator, but a high network exposure towards the third party makes this option more challenging in terms of security.

AS DESCRIBED IN THE PRECEDING SECTIONS, MEC IS A RECENT TECHNOLOGY THAT IN THE NEXT GENERATION OF MOBILE NETWORKS WILL ALLOW CSPS TO HOST USE-CASE-SPECIFIC CONTENT AND APPLICATIONS, AS WELL AS SERVICES.

This is specified through the involvement of several standardization groups, including the Industry Specification Group by ETSI and so forth, which recommend and specify a multi-vendor platform and service orchestration. As pointed out, the MEC will be able to create new use cases and support particular scenarios where high data rates and low latency are expected. So, the ability to provide content and services at the closest point to the user unveils new opportunities to launch services requiring performance which previously would not have been satisfied by using an operator's network and centralized architectures.

In summary, these new services have the potential to positively disrupt the traditional business models and to open up new opportunities for the CSPs to be at the center of the ecosystem of various vertical services within the IoT, Intelligent Transport Systems, Multimedia and Broadcasting.

REFERENCES & SOURCES

1 https://www.mckinsey.de/files/mck_connected_car_report.pdf

2 http://www.ieeevtc.org/conf-admin/vtc2015fall/14.pdf

3 https://www.digitaltrends.com/cool-tech/connected-cow-internet-of-cows-managed-mating/

4 https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-eHealth-Vertical-Sector.pdf

5 http://www-file.huawei.com/-/media/CORPORATE/PDF/mbb/cloud-ran-the-next-generation-mobile-network-architecture.pdf?la=en

6 http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing

7 https://datatracker.ietf.org/meeting/98/materials/slides-98-nfvrg-sessb-12-multi-access-edge-computing-mec-applications

8 http://www.3gpp.org/images/articleimages/ongoing_releases_900px.JPG

AUTHORS

Gianluca Noya
gianluca.noya@accenture.com

Lead Accenture Authors:

Marc Schellenberg
marc.schellenberg@accenture.com

Martin Macuha
martin.macuha@accenture.com

Gregor Tomic
gregor.tomic@accenture.com

Lead 5G Lab Germany authors:

Robert-Steve Schmoll
robert-steve.schmoll@tu-dresden.de

Alexander Kropp
Alexander.Kropp@tu-dresden.de

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions – underpinned by the world's largest delivery network – Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With approximately 401,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at **www.accenture.com**.

ABOUT THE 5G LAB GERMANY

The 5G Lab Germany at Technische Universität Dresden is an interdisciplinary team with more than 600 researchers from 22 different research areas of the university that aims to deliver key technologies for enabling 5G. The lab carries out research in four tracks: hardware and silicon; wireless; network and cloud; and tactile internet applications. This holistic 5G view is the core theme of the lab, which will help to define the capability of 5G mobile networks in meeting the massive connectivity demands of the future across Industry 4.0; transportation systems; smart grids; health care; and others ranging from agriculture to construction. Visit us at **www.5glab.de**.