Schneider Electric

EDGE Computing

General Trends and implications for Industrial Automation

Gilbert BRAULT
Q1 2020

GREEN PAPER

Context

Edge Computing is a new trend in the information technology landscape triggered by the Internet of Things (IoT), which should literally drown cloud providers resources, due to the exponential growth of data production on the ground beside the fact it is not delivering today the response time required by some applications.

The IT community created this new concept to face this cloud architecture centralization downside. In a nutshell, Edge computing is a decentralized architecture which will support IoT scale-up, keeping, in the same time, the cloud promise.

The edge is like a distributed cloud with proximity close to the end user that delivers ultra-low latency, reliability, and scalability.

Edge Computing and 5G is not just incrementing performance: the network latency, the download speed and the geographical coverage. This has already been achieved by previous technology generation to a great extent. it's a new business play, potentially reshuffling the Telco and IT companies business landscape.

This green paper aims to provide some knowledge about this new market to envision a potential future for Industrial Automation in this market segment.

Table of Content

Executive Summary

What is Edge Computing?

EDGE Computing concept takeoff in the year 2016-2017 or earlier[1] to cope with the cloud centralization road-block identified by the IT community as a definitive barrier to IoT scale-up or fast response time applications.

The first large scale deployment of Edge Computing is likely to be the Multi-Access Edge Computing (MEC). Even if loosely coupled with 5G, the concepts and standards have been developed by ETSI, with 5G perspective in mind, since 2016.

MEC adds computing and storage to the traditional communication network and provides an end-user application framework which end-user lease as they used to, leasing communication lines.

MEC has the capability to be deployed at every level of the communication infrastructure and can be available[2]

- On base station (Mobile antenna or fiber junctions aggregation nodes) (250m à 1km)
- On access network (5 to 10km)
- On aggregation network (5 to 20km)
- On core network (10 to 100km)

MEC is likely to be deployed on end-users' premises. This mean MEC round trip time may be as low as 1ms (MEC distant to less than 10km from customer premise) but in the same time can also access internet.

MEC and 5G stake not only cope with greater performance and unified communication and networking architecture but also address a new generation of Internet, Cloud computing, bringing a new information services framework significant to End Users business core.

The technical side of MEC is greatly surpassed, – by far –, by non-technical issues. This might reshape the market now owned by

- Communication Service Providers, issued from the traditional Telco which have been already widely transformed in the last 2 decades
- Cloud Computing Companies, providing new services and which have significantly grown in the last decade to a point where they have now a real role in the Corporate IT landscape
- Information Service companies which have ever played an application integration role delivering end-to-end services to End Users
- Software companies, which business model is going to be heavily transformed by this new setting where the application is delivered to the end user as a service and not as a software right of use.
- Many other players from this business landscape transformation, emerging or having business at stake are impacted
    - The "Tower Companies" which are the last-miles EDGE border and likely to "host" MEC nodes in their infrastructure premises
    - Legacy or emerging Industries which application field is largely impacted by this new technology capabilities

[1] ETSI released Mobile-Edge Computing – Introductory Technical White Paper, September 2014;link
[2] See MEC in 5G networks from ETSI link

- Vehicle to Vehicle communication to bring a new car transportation and driving experience
- Industrial Automation, seen as one of the most attractive segments and to enable the Industrie 4.0 bargains
- Many B2C domains
  - Healthcare being one of the first (medical patient file, remote patient diagnostic…)
  - Gaming industry
- Many IoT players leveraging the new EDGE computing framework enabling new market creation

To summarize, edge computing is a distributed cloud with proximity close to the end user that delivers ultra-low[3] latency, scalability[4], reliability and security, providing Information Services addressing core business function for the Corporate World (business application level).

This new offer will be delivered by a set of players which will regroup the legacy resources of the Information and communication industries but organized in a new setting addressing the associated technology and business challenges.

For Industrial Automation (IA), the opportunity of this new technology is the ability to implement Cyber Physical Systems (CPS) it might also be seen by traditional IA player as a threat, being a new technology, which is able to address Industrial Automation applications implementation constraints.

Prerequisite Requirements for Cyber Physical Systems?

For Industrial Automation vendors, the opportunity is to use this new technology to implement Cyber Physical Systems (CPS).

To build Cyber Physical Systems this new technology needs to address the following issues, constraining Edge Computing implementation, including but not restricted to

- Operate multi-tenant systems, specific to the industrial field, by third parties
- Develop Application
- Deploy and Operate Application
- Allocate compute and storage resources as if “private”
- Allocate network resources as if “private”
- Provide deterministic “low” round-trip time

Operate multi-tenant systems, specific to the industrial field, by third parties

The owners of the new Edge and communication assets are faced by multiple business issues and must focus on their core competence: infrastructure scale-up and management.

Industry specific deployments, application Devops and commercial development needs to be done by 3rd party which will manage the required domain specific knowledge.

This ecosystem setting is a critical success factor for implementing this new technology and is a chicken and eggs problem.

[3] On a “restricted” geography: latency is proportional to the actual distance between client and server and the number of hops

[4] If data is processed locally or distributed across local consumer: not being centralized at wide scale, it’s more the information (signal processing) than data that is the customer focus

Edge and communication assets providers must then offer tools to 3rd parties who will deploy and manage industry specific frameworks to address end-user specific requirements from a business perspective.

This 3rd party lease Edge and Communication resources using the multitenant tools allowing them to

- Evaluate and simulate new deployment architecture corresponding to specific user requirements
- Build and deploy end-user specific system architecture tailored to their requirements: supporting application development, deployment and operation as well as networking
- Measure and invoice, for each customer, used resources per period
- Automatically setup the resources needed and require them on-demand to the "Edge and Communication" provider
- Manage resources allocation request and their follow-up
- Monitor, from the 3rd Party perspective, the global system health, alert management…
- Implement cyber security and privacy measures
- Implement availability operating modes (data, compute, network backup etc. …)
- Manage service level agreements (SLA) contracting system performance and resources for all parties
- Comply with regulatory requirements (depending upon geography, for example: GDPR in Europe, etc. …)
- Support Customer request (Customer service) and manage customer account (resources and services definition, accounting management)
- Communicate about the provided offer: 3rd party commercial website
- Invoice resources and services used by the 3rd party

This multitenant tool can be fully branded by the 3rd party.

Develop Application

Application is a partition of resources united and dedicated to solving a specific business or technical problem. Application intermingle network, compute and storage resources with different pieces of code which might be pre-existing (libraries) or developed on demand to address a specific customer business process.

Today, modern application model, is well represented by Docker Swarm. A Docker swarm is a set of compute engines, including storage, distributed over multiple 'physical' networks and seen from the application perspective as clients-servers operating over a local area network (LAN). Each node, which is a container, benefits from DNS services enabling to access network resources by name. Depending upon configuration, the swarm can also access the internet if one of the nodes has internet access. Another flavor of Docker Swarm is Kubernetes.

Application development tools provide the following features (DevOps)

- Manage code life-cycle (versioning, Documentation, multi-level testing)
- Manage container image creation and life-cycle
- Manage access to container images that will be used to deploy applications
- Manage test life-cycle (Continuous Integration)
- Manage libraries
- IDE to create code and making some tests

Deploy and Operate Application

Once container images are built, they need to be integrated in a swarm, each container providing on-going or on-demand services. One needs now

- To deploy a set of containers: may be according to workload, working in an on-going manner or on-demand
- To monitor the health of the system and respond to downgraded operations
- To change incrementally, as new business features, get implemented

Specific languages enable this “orchestration” which enable definition and automation of those deployment and monitoring tasks.

Docker offer one language, Kubernetes another but both use Docker containers as run time embodiment which is now deeply embedded in Windows or Linux technology.

Allocate compute and storage resources as if “private”

Compute engines can be allocated on-demand and must support basic virtualization feature like spatial and temporal partitioning. Those requirements are today partially fulfilled by “cloud computing” frameworks like Google Compute Engine, Amazon AWS EC2 or Microsoft Azure Virtual Machines and by many actors providing VPS (virtual private servers) services

Spatial partitioning

Spatial partitioning consists in ensuring that it is not possible that an application writes into the memory or data (storage) of an application running on a different partition.

Temporal Partitioning

Temporal partitioning consists in ensuring that the activities in one partition do not affect the timing of the activities in other partition.

Beside the above key virtualization features, the following is also needed:

Storage location and privacy

On the base ground service, data storage can be shared but the location of data must be set by the end-user. The storage can also, on-demand, be private.

Privacy of this storage is supported by features like ciphering or user owned storage (with the right to get the physical device in and out)

On demand resource allocation

Compute and storage resources can grow on-demand. The capabilities of computing nodes, like Artificial Intelligence dedicated processors or computer graphics processors, the memory size etc.… can be defined according to user needs. Network services must be also linked to the compute and storage allocation mechanism.

Computing resources can be deployed according to a given geography for the network to comply with round trip time requirements, if geographically feasible.

Manage High-Availability

Compute and storage resources can be reallocated to new resources, real-time, in case of system crash. This feature can be set on-demand, not all resources need this feature.

Capability to deploy containerized application

Support execution and monitoring of compute loads defined as container on-demand or continuously.

Allocate network resources as if "private"

This paradigm is part of the privacy and security requirements but also address performance issues as some Edge resources might be pre-allocated for higher throughput or lower latency. Let's see it as communication partitioning: consists in ensuring that the activities in one communication channel do not cross over the activities in other communication channels.

In the 5G world this is denominated "Network Slicing"[5]:

"5G network slicing is a network architecture that enables the multiplexing of virtualized and independent logical networks on the same physical network infrastructure. Each network slice is an isolated end-to-end network tailored to fulfil diverse requirements requested by an application.

For this reason, this technology assumes a central role to support 5G mobile networks that are designed to efficiently embrace a plethora of services with very different service level requirements (SLA). The realization of this service-oriented view of the network leverages on the concepts of software-defined networking (SDN) and network function virtualization (NFV) that allow the implementation of flexible and scalable network slices on top of a common network infrastructure.

From a business model perspective, each network slice is administrated by a mobile virtual network operator (MVNO). The infrastructure provider (the owner of the telecommunication infrastructure) leases its physical resources to the MVNOs that share the underlying physical network. According to the availability of the assigned resources, a MVNO can autonomously deploy multiple network slices that are customized to the various applications provided to its own users."

Provide deterministic "low" round-trip time[6]

Implementation of Cyber Physical Systems requires stringent Round-Trip time[7]. 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.

[5] Network Slicing Wikipedia link
[6] Accenture MULTI-ACCESS EDGE COMPUTING FOR PERVASIVE NETWORKS link
[7] See 3GPP Service requirements for the 5G system (Release 17) section in this document

Scenario for Industrial Automation

To Be Continued

Benefits for End Users (Industrial Automation scope)

To Be Continued

Conclusion

Edge Computing is really to provide a next level user experience, at a fair cost, which might impact favorably its core business. As such Edge Computing breaks a lot of existing business practice and requires many changes which are going to disrupt habits if not business of many today players.

The essence of this new technology benefit best when end-users need to share information at a high pace rate with low latency, which is the basic of signal processing of course.

A basic question is to understand the business rationale that will make possible this technology funding: what are the “killer apps” and who will pay for it?

Even if V2x “Vehicle to Everything” is quite an appealing case, who wants to pay for it? The increase in car cost is likely to be several thousand dollars not considering the Edge computing design, infrastructure and operation… If a small portion of cars are populated with this technology what is the benefit for the society and for the driver! How to migrate the installed base with this new technology?

5G inception is one opportunity but Edge Computing is much more than 5G… In another hand 5G without Edge computing might not be very attractive…

Of course, this technology might be very well adapted to Industry 4.0, as the core of Industry 4.0 is the interaction of a web of suppliers, manufacturers and customers. But it’s not likely it will benefit from the latency and network performance as they may be spread all over the world…

The integrated services might well balance this lack of attractivity in another hand.

Of course, it will be very attractive in geographically concentrated cluster with a high level of digital exchanges… Maybe a new paradigm for economical country development!

Concerning Cyber Physical System implementation, Edge computing is very well fitted and could replace micro-data centers that would play this role on-premise, the benefits being the one provided by the cloud model (the life-cycle cost might not be very attractive so), but a much better performance and user experience compared to today internet and cloud implementation. The level of service can also improve by a significant factor, as not every corporation will be likely to implement sophisticated IA and analytics due to the lack of expertise and struggle to create and maintain it.

Make a simple analogy with power generation: Utilities are specialized company to generate and distribute electricity: not everyone is generating its own electricity! It’s likely to be the same with information management as it becomes more and more techy…

We believe Edge Computing will happen if they are other major applications which will justify this reshaping of communication and information technology landscape – IA will then benefit from the creation of this new infrastructure and workforce –

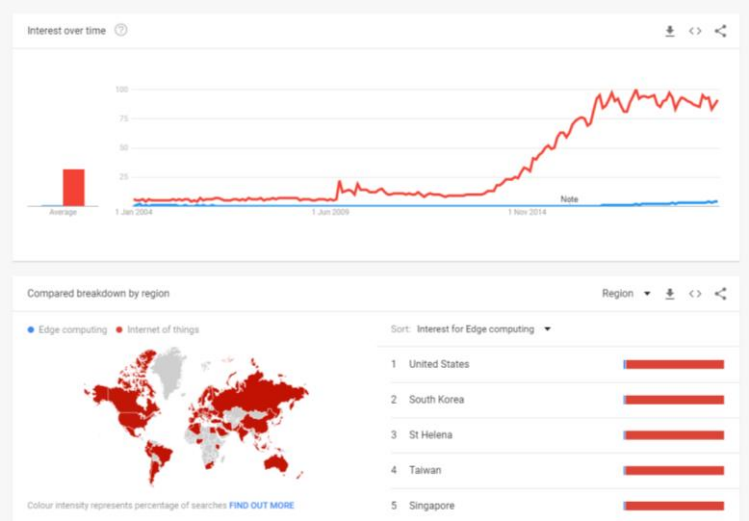
Industrial Automation vendors must then very carefully scan the landscape changes and adapt their offers, as the landscape changes, being very opportunistic and agile! Edge computing is likely to happen but might take some time.

Edge Computing landscape

Edge Computing a topic for specialist?

Using Google Trends tool (https://trends.google.com/) , which provides some statistics about Google topic searched over time, Comparing the two topics during the 2004-2020 period:

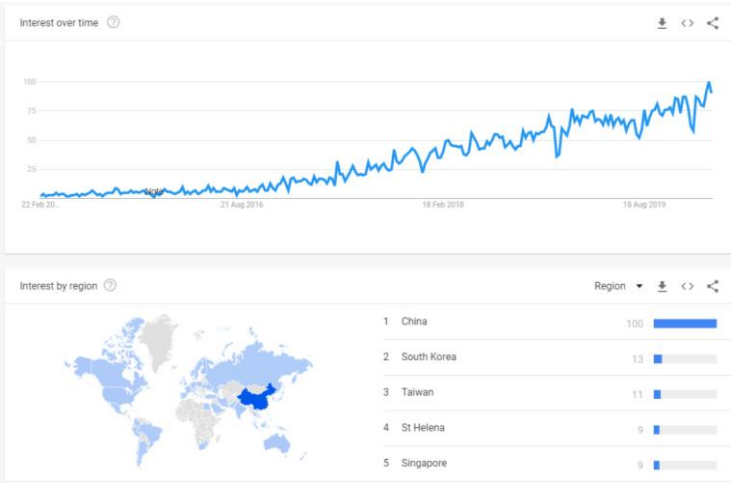
- Edge computing
- Internet of things



On can derive, from the above figure, the following thoughts:

1. IoT is much more popular: IoT index is 100 when Edge Computing is 5
2. IoT is a more 'mature' topic: IoT takeoff is in 2014 Edge Computing 2017
3. Since 2017 IoT entered a steady-state mode

"Zooming" over Edge computing we have the following information



It seems then:

1. China and Asia are making the most request about Edge Computing
2. USA and India are 3 on a scale of 100 (where is China)

Who coined the "Edge Computing" term?

According to Wikipedia (https://en.wikipedia.org/wiki/Edge_computing), **Karim Arabi**, head of Corp. R&D ASIC department at Qualcomm at that time, in an IEEE DAC 2014 Keynote[8] and subsequently in an invited talk at MIT's MTL Seminar in 2015[9] defined edge computing broadly as all computing outside the cloud happening at the edge of the network, and more specifically in applications where real-time processing of data is required. In his definition, cloud computing operates on big data while edge computing operates on "instant data" that is real-time data generated by sensors or users.

For BayTech Consulting[10], **Alex Reznik**, Chairman of the European Telecommunications Standards Institute (ETSI) Multi-access Edge Computing (MEC)[11] industry specification group (ISG) defines edge computing more broadly. He considers edge computing to be any computing performed outside a traditional data center, since such a location would be the edge of a network for someone. In their website[12], in 2016, ETSI defined the Mobile Edge Computing (MEC) as:

> "MEC offers IT service and cloud-computing capabilities at the edge of the mobile network in an environment that is characterized by proximity, ultra-low latency and high bandwidth. Furthermore, it provides exposure to real-time radio network and context information."
>
> ….
>
> "MEC can be considered as the "Cloud" for real-time and personalized mobile services, providing an unparalleled experience with reduced latency, highly efficient network operation and service delivery."

[8] Mobile Computing Opportunities, Challenges and Technology Drivers link
[9] MTL Seminar in 2015 link
[10] What is Edge Computing? link
[11] MEC was created in 2016 link
[12] Transforming the mobile-broadband experience is no longer a pipe dream link

Evolution of wireless communication networks to 5G[13]

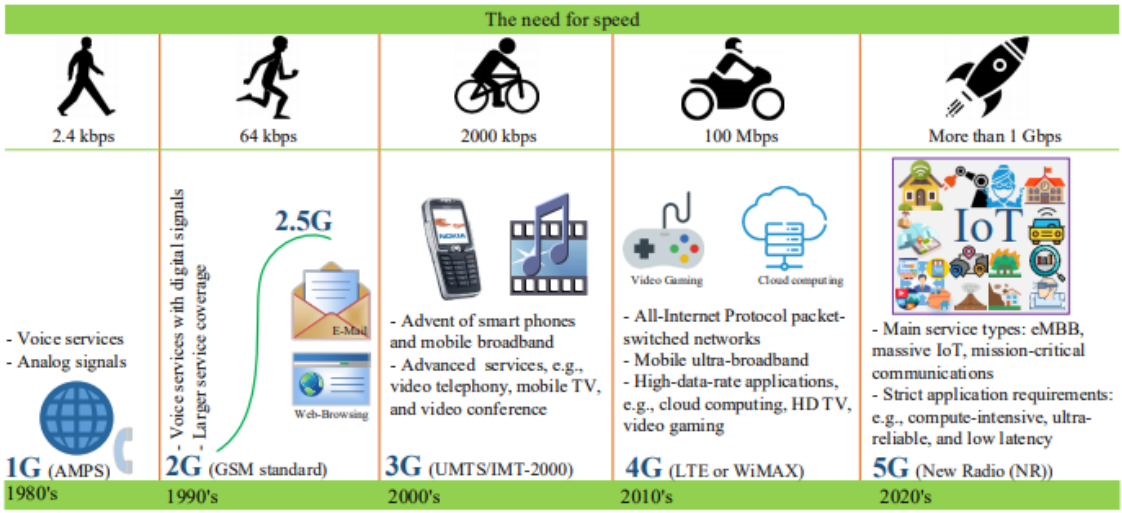


Figure 1: Evolution of wireless communication.

During the last four decades, the evolution of wireless communication networks has changed every aspect of our lives, society, culture, politics, and economics. Since the commercialization of the first generation (1G) of cellular networks in early 1980's, generations have been launched with enormous differences in terms of the network architectures, key technologies, coverage, mobility, security and privacy, data, spectral efficiency, cost optimality, and so on. The summary of wireless communication evolution is shown in Fig. 1. Now, both academic and industry communities are making tremendous efforts to finalize the 5G standardization and commercialization in 2019. 5G communications can be categorized into three categories:

- enhanced mobile broadband (eMBB),
- ultra-reliable low-latency communication (URLLC),
- and massive Internet of Things (IoT).

Compared with previous generations, 5G will support not only communication, but also computation, control, and content delivery (4C) functions.

Moreover, many new applications and use cases are expected with the advent of 5G, for example, virtual/augmented reality (VR/AR), autonomous vehicle, Tactile Internet, and IoT scenarios. These applications are poised to induce a significant surge in demand for not only communication resources but also computation resources. To meet such ever-growing demands, various technological concepts have been developed for 5G in terms of radio access, network resource management, applications, network architectures and scenarios, power supply, and performance improvement [2]. For example, nonorthogonal multiple access (NOMA), dense heterogeneous networks (HetNets), cloud radio access network (C-RAN), unmanned aerial vehicle (UAV), IoT, wireless power transfer (WPT) and energy harvesting (EH), and machine learning (ML), have been considered as key enabling technologies. The Cisco white paper [3] showed that global data traffic will grow at a compound annual growth rate (CAGR) of 26 percent between 2017 and 2022 (i.e., increase more than threefold) and reach 122 exabytes (EB) per month by 2022.

[13] A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art link

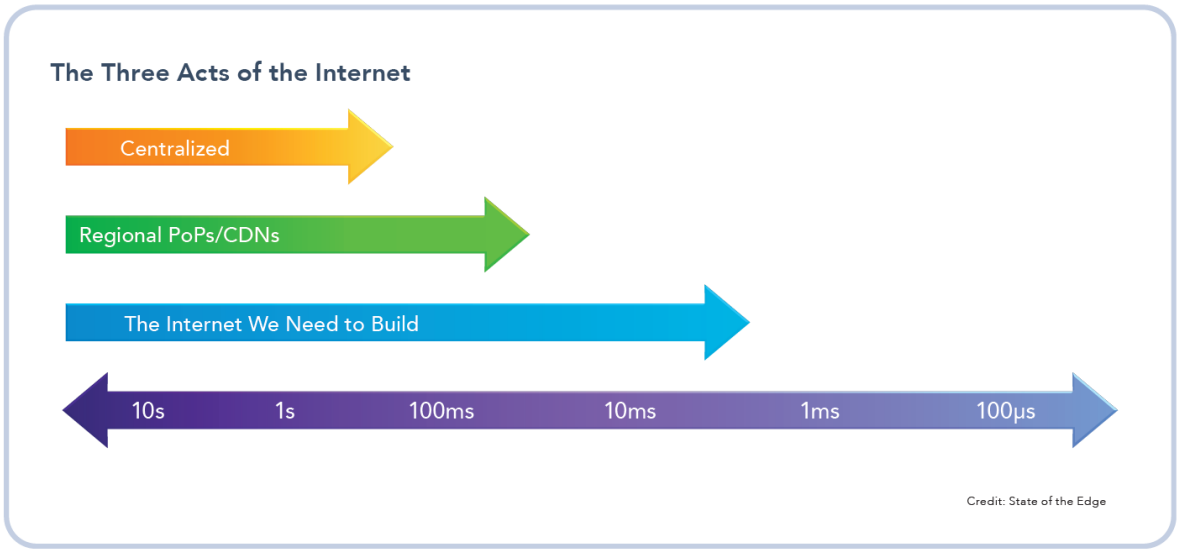
Mobile and wireless networks carried 11.51 EB per month in 2017, 28.56 EB per month in 2019, and 77.49 EB per month at the end of 2022. Moreover, traffic generated by new applications and services will increase at a much higher CAGR, for example, 12-fold for AR and VR, nine-fold for Internet gaming, and sevenfold for Internet video surveillance.

It is also anticipated that the number of connected things (e.g., sensors and wearable devices) will reach 28.5 billion by 2022, up from 21.5 billion in 2019. However, most connected devices have limited communication and storage resources and finite processing capabilities, which show the mismatch between the stringent requirements for emerging applications and the actual device capabilities. Despite recent advancements in the hardware capability, mobile computing still cannot cope with the demand of many applications that need to generate, process, and store a massive amount of data and require large computing resources. One potential solution to these challenges is to transfer computations to centralized clouds, which can be, however, burdened by many issues, such as network congestion and privacy policies. This has driven the development of mobile edge computing (MEC).

The Edge, the third generation of Internet?[14]

According to the State of The EDGE organization:

- Over $700 billion in cumulative CAPEX will be spent within the next decade on edge IT infrastructure and data center facilities. This includes edge equipment at access network sites (e.g. radio base station tower sites), pre-aggregation sites (e.g. street-side cabinets), and aggregation and central office sites.
- **The edge is the "Third Act of the Internet"**, building upon previous phases of regionalization and origination. While there remain significant unknown challenges at the edge, substantial business opportunities have emerged.
- The Edge today is a solution-specific story. Equipment and architectures are purpose-built for specific use cases, such as 5G and network function virtualization, next-generation CDNs and cloud, and streaming games.
- Edge computing is a global phenomenon. Asia Pacific currently has the largest edge computing equipment footprint of all the global regions, estimated to be 187MW today. Buoyed by markets like Australia, Japan, Korea and China
- The deployed global power footprint of the edge IT and data center facilities is forecast to reach 102 thousand MW by 2028[15] (1.5% of WW Power generation capacity, today Cloud Computing is estimated 1% of the WW Power Generation capacity). Datacenters, servers, local breakout and middle mile architecture are areas of intense investment.
- A common definition of the edge is gaining momentum. Community efforts including the LF Edge foundation, the Glossary of Edge Computing, and the Edge Computing Landscape provide cohesion.



[14] STATE OF THE EDGE 2020 link

[15] According to: WORLD ENERGY OUTLOOK 2020, WW Installed Power Generation capacity = 6500GW link, it would be 102/6500= 1.5% of WW capacity

IA Edge Controller "Emerging market Analysis" ARC Advisory Group

Stefan Miksch and Florian Güldner issued this report in February 2020. ARC recognized the inception of Edge in the Industrial Automation, based on the following timeline

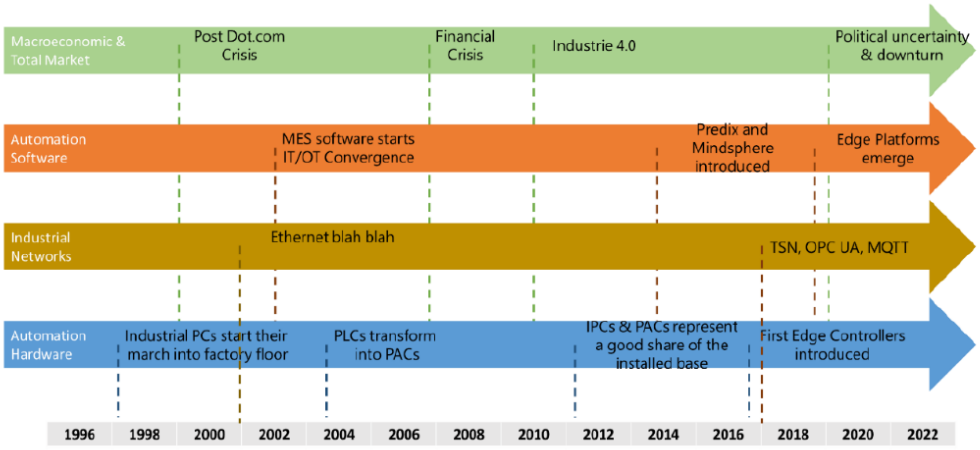


Figure 2: Historical Time Line for Edge Controllers

They define the "Edge Controller" as:

"The term "edge controller" refers to a device used for combined control and edge-to-cloud applications at the IoT edge, especially in machine control. Edge controllers merge the functionality of an industrial controller and an IoT edge device (edge computer, gateway) into a single device. This creates a hybrid device that can control a machine (logic, motion, numerical control) while simultaneously communicating bi-directionally with cloud-based apps (public, private, on-premise, hybrid). For cloud-based apps, an edge controller may run an edge software platform and applications in parallel with the controller function."

For them DCS does not fall in this category:

"The DCS does not fall under the controller category as its uses centralized intelligence and decentralized control to guarantee system integrity, safety, and availability."

In a nutshell, the scenario in this report is as follow:

- EDGE controllers are built reengineering the traditional controllers used in Discrete industry
  - PLC
  - CNC
  - GMC
  - IPC
- They are extended with 'cloud connectivity' and replace IoT boxes that have emerged in the recent years

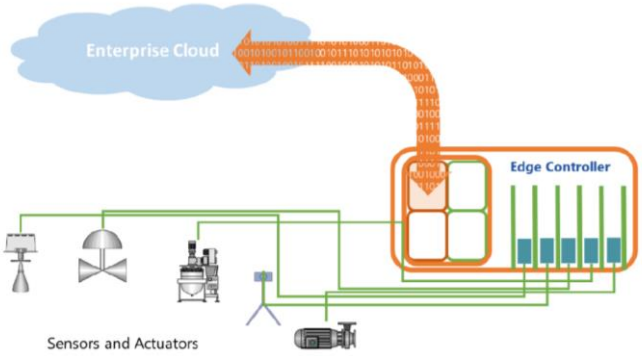


Figure 3: EDGE Controller According to ARC

Given this scenario framework, they quantify the market as follow

	Revenue (Millions of US Dollars)						CAGR
Controller Type	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2019-2024**
PLC	152.2	193.3	254.0	345.4	456.9	575.2	30.5%
IPC	70.0	93.6	121.7	165.5	217.3	273.6	31.3%
GMC	6.8	9.3	13.3	18.6	25.2	34.1	38.2%
CNC	22.9	28.1	39.4	54.6	70.5	91.3	31.9%
Other	2.0	2.5	3.9	5.3	7.7	10.8	40.4%
Total	253.8	326.8	432.2	589.4	777.6	985.0	31.2%

Figure 4: Edge Controller revenue Forecast by Controller Type

They further provide a definition of the EDGE:

- Represents a top-down perspective from the enterprise or datacenter level that results in production equipment appearing at the outer edge of the architecture
- Is distinctive from distributed automation & control systems due to emphasis on service to cloud or datacenter-based enterprise applications and not execution of traditional distributed automation & control.
- Introduced concurrently with the Industrial IoT initiative
- Is the point where IT/enterprise architecture, data, applications, and services meet OT/production

As a conclusion, the “EDGE Controller” business is significant but not very big as well (~B$1 in 2024), growth is impressive (>30% CAGR). If Industrial Automation players want a greater share of this new pie, they must then, not only adapt their core offer to interoperate in the new eco-system, but as well develop new Edge services to increase significantly their revenues in this market segment.

3GPP Service requirements for the 5G system (Release 17)

the 3rd Generation Partnership Project (3GPP) unites [Seven] telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC), known as “Organizational Partners” and provides their members with a stable environment to produce the Reports and Specifications that define 3GPP technologies.

They have released a “Service requirements for the 5G system (Release 17)” which scope is:

“

The need to support different kinds of User Equipment’s (UEs) (e.g. for the Internet of Things (IoT)), services, and technologies is driving the technology revolution to a high-performance and highly efficient 3GPP system. The drivers include IoT, Virtual Reality (VR), industrial control, ubiquitous on-demand coverage, as well as the opportunity to meet customized market needs. These drivers require enhancements to the devices, services, and technologies well established by 3GPP. The key objective with the 5G system is to be able to support new deployment scenarios to address diverse market segments.

This document compiles requirements that define a 5G system.

The 5G system is characterized, for example, by:

- Support for multiple Access technologies
- Scalable and customizable network
- Advanced Key Performance Indicators (KPIs) (e.g. availability, latency, reliability, user experienced data rates, area traffic capacity)
- Flexibility and programmability (e.g. network slicing, diverse mobility management, Network Function Virtualization)
- Resource efficiency (both user plane and control plane)
- Seamless mobility in densely populated and heterogeneous environment
- Support for real time and non-real time multimedia services and applications with advanced Quality of Experience (QoE)

”

This document captures the requirements for Industrial Automation

In the document body: 6.28 Cyber-physical control applications in vertical domains

“6.28.1 Description

The 5G system is expected to meet the service requirements for cyber-physical control applications in vertical domains.

A vertical domain is an industry or group of enterprises in which similar products or services are developed, produced, and provided. Automation refers to the control of processes, devices, or systems in vertical domains by automatic means. The main control functions of automated control systems include taking measurements, comparing results, computing any detected or anticipated errors, and correcting the process to avoid future errors. These functions are performed by sensors, transmitters, controllers, and actuators.

Cyber-physical systems are to be understood as systems that include engineered, interacting networks of physical and computational components. Cyber-physical control applications are to be understood as applications that control physical processes. Cyber-physical control applications in automation follow certain activity patterns, which are open-loop control, closed-loop control, sequence control, and batch control.

Communication services supporting cyber-physical control applications need to be ultra-reliable, dependable with a high communication service availability, and often require low or (in some cases) very low end-to-end latency.

Communication in automation in vertical domains follows certain communication patterns. The most well-known is periodic deterministic communication, others are a-periodic deterministic communication and non-deterministic communication.

Communication for cyber-physical control applications supports operation in various vertical domains, for instance industrial automation and energy automation.

6.28.2 Requirements

The 5G system supports the communication services for cyber-physical control applications in the vertical domains of factories of the future (smart manufacturing), electric power distribution, central power generation, and rail-bound mass transit. The associated requirements are described in 3GPP TS 22.104 [21].”

in annex D1 to D3:

- Motion Control
- Discrete Automation
- Process Automation

D.1 Discrete automation – motion control

Industrial factory automation requires communications for closed-loop control applications. Examples for such applications are motion control of robots, machine tools, as well as packaging and printing machines. All other discrete-automation applications are addressed in Annex D.2.

The corresponding industrial communication solutions are referred to as fieldbuses. The pertinent standard suite is IEC 61158. Note that clock synchronization is an integral part of fieldbuses used for motion control.

In motion control applications, a controller interacts with many sensors and actuators (e.g. up to 100), which are integrated in a manufacturing unit. The resulting sensor/actuator density is often very high (up to 1 m-3). Many such manufacturing units may have to be supported within proximity within a factory (e.g. up to 100 in automobile assembly line production).

In a closed-loop control application, the controller periodically submits instructions to a set of sensor/actuator devices, which return a response within a cycle time. The messages, referred to as telegrams, are typically small (≤ 56 bytes). The cycle time can be as low as 2 ms, setting stringent end-to-end latency constraints on telegram forwarding (1 ms). Additional constraints on isochronous telegram delivery add tight constraints on jitter (1 ?s), and the communication service has also to be highly available (99,9999%).

Multi-robot cooperation is a case in closed-loop control where a group of robots collaborate to conduct an action, for example, symmetrical welding of a car body to minimize deformation. This

requires isochronous operation between all robots. For multi-robot cooperation, the jitter (1µs) is among the command messages of a control event to the group robots.

To meet the stringent requirements of closed-loop factory automation, the following considerations may have to be taken:

- Limitation to short-range communications.
- Use of direct device connection between the controller and actuators.
- Allocation of licensed spectrum for closed-loop control operations. Licensed spectrum may further be used as a complement to unlicensed spectrum, e.g. to enhance reliability.
- Reservation of dedicated air-interface resources for each link.
- Combination of multiple diversity techniques to approach the high reliability target within stringent end-to-end latency constraints such as frequency, antenna, and various forms of spatial diversity, e.g. via relaying
- Utilizing OTA time synchronization to satisfy jitter constraints for isochronous operation.

A typical industrial closed-loop motion control application is based on individual control events. Each closed-loop control event consists of a downlink transaction followed by a synchronous uplink transaction, both of which are executed within a cycle time. Control events within a manufacturing unit may have to occur isochronously. Factory automation considers application-layer transaction cycles between controller devices and sensor/actuator devices. Each transaction cycle consists of (1) a command sent by the controller to the sensor/actuator (downlink), (2) application-layer processing on the sensor/actuator device, and (3) a subsequent response by the sensor/actuator to the controller (uplink). Cycle time includes the entire transaction from the transmission of a command by the controller to the reception of a response by the controller. It includes all lower layer processes and latencies on the air interface as well the application-layer processing time on the sensor/actuator.

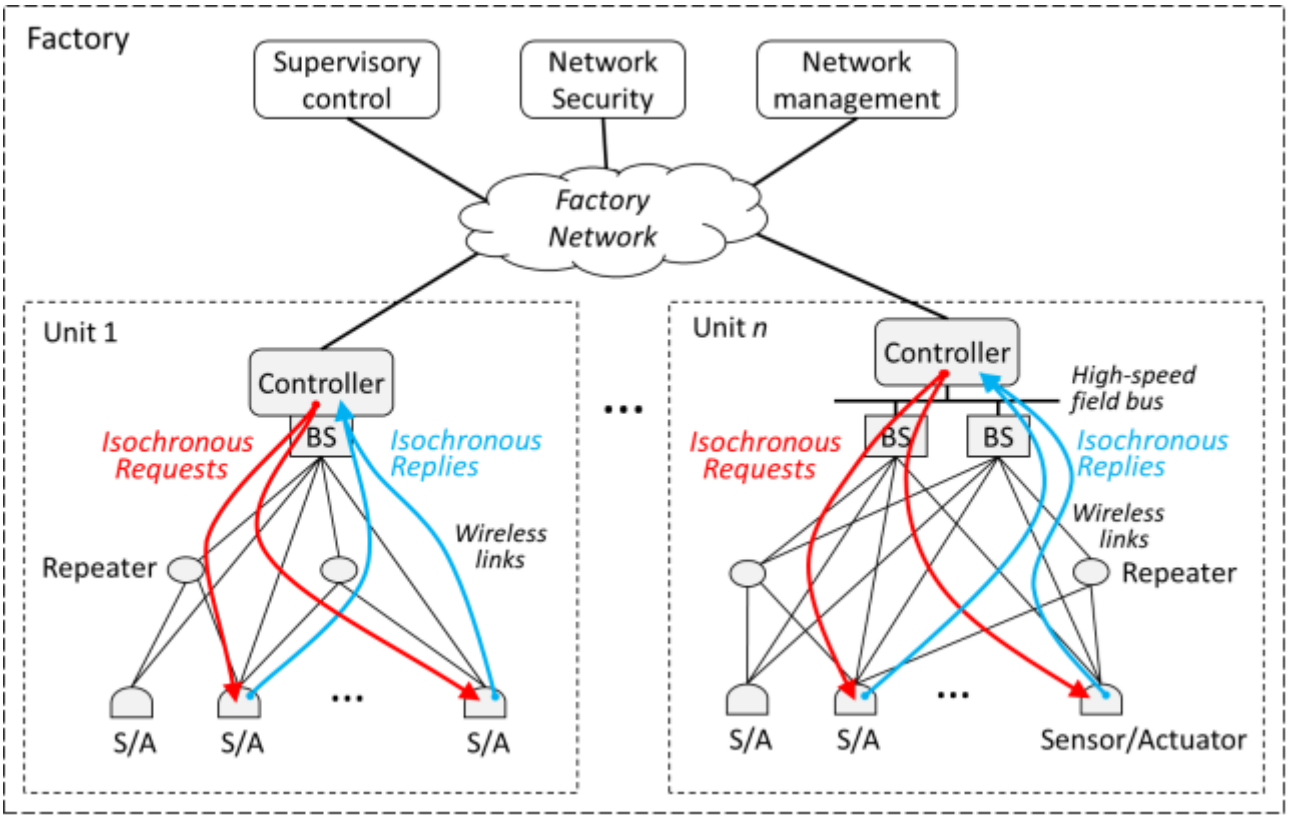


Figure D.1-1: Communication path for isochronous control cycles within factory units. Step 1 (red): controller requests sensor data (or an actuator to conduct actuation) from the sensor/actuator (S/A). Step 2 (blue): sensor sends measurement information (or acknowledges actuation) to controller.

Figure D.1-1 depicts how communication may occur in factory automation. In this use case, communication is confined to local controller-to-sensor/actuator interaction within each manufacturing unit. Repeaters may provide spatial diversity to enhance reliability.

D.1.1 Service area and connection density

The maximum service volume in motion control is currently set by hoisting solutions, i.e. cranes, and by the manipulation of large machine components, e.g. propeller blades of wind-energy generators. Cranes can be rather wide and quite high above the shop floor, even within a factory hall. In addition, they typically travel along an entire factory hall.

An approximate dimension of the service area is 100 x 100 x 30 m.

Note that production cells are commonly much smaller (< 10 x 10 x 3 m). There are typically about 10 motion-control connections in a production cell, which results in a connection density of up to 105 km-2.

D.1.2 Security

Network access and authorization in an industrial factory deployment is typically provided and managed by the factory owner with its ID management, authentication, confidentiality and integrity.

Note that motion control telegrams usually are not encrypted due to stringent cycle time requirements.

A comprehensive security framework for factories has been described in IEC 62443.

D.2 Discrete automation

Discrete automation encompasses all types of production that result in discrete products: cars, chocolate bars, etc. Automation that addresses the control of flows and chemical reactions is referred to as process automation (see clause D.3). Discrete automation requires communications for supervisory and open-loop control applications, as well as process monitoring and tracking operations inside an industrial plant. In these applications, many sensors distributed over the plant forward measurement data to process controllers on a periodic or event-driven base. Traditionally, wireline field bus technologies have been used to interconnect sensors and control equipment. Due to the sizable extension of a plant (up to10 km2), the large number of sensors, rotary joints, and the high deployment complexity of wired infrastructure, wireless solutions have made inroads into industrial process automation.

This use case requires support of many sensor devices per plant as well as high communication service availability (99,99%). Furthermore, power consumption is relevant since some sensor devices are battery-powered with a targeted battery lifetime of several years while providing measurement updates every few seconds. Range also becomes a critical factor due to the low transmit power levels of the sensors, the large size of the plant and the high reliability requirements on transport. End-to-end latency requirements typically range between 10 ms and 1 s. User experienced data rate can be rather low since each transaction typically comprises less than 256 bytes. However, there has been a shift from field busses featuring somewhat modest data rates (~ 2 Mbit/s) to those with higher data rates (~ 10 Mbit/s) due to the increasing number of distributed applications and "data-hungry"

applications. An example for the latter is the visual control of production processes. For this application, the user experienced data rate is typically around 10 Mbit/s and the transmitted packets are much larger.

The existing wireless technologies rely on unlicensed bands. Communication is therefore vulnerable to interference caused by other technologies (e.g. WLAN). With the stringent requirements on transport reliability, such interference is detrimental to proper operation.

The use of licensed spectrum could overcome the vulnerability to same-band interference and therefore enable higher reliability. Utilization of licensed spectrum can be confined to those events where high interference bursts in unlicensed bands jeopardizes reliability and end-to-end latency constraints. This allows sharing the licensed spectrum between process automation and conventional mobile services.

Multi-hop topologies can provide range extension and mesh topologies can increase reliability through path redundancy. Clock synchronization will be highly beneficial since it enables more power-efficient sensor operation and mesh forwarding.

The corresponding industrial communication solutions are referred to as fieldbuses. The related standard suite is IEC 61158.

A typical discrete automation application supports downstream and upstream data flows between process controllers and sensors/actuators. The communication consists of individual transactions. The process controller resides in the plant network. This network interconnects via base stations to the wireless (mesh-) network which hosts the sensor/actuator devices. Typically, each transaction uses less than 256 bytes. An example of a controller-initiated transaction service flow is:

1. The process controller requests sensor data (or an actuator to conduct actuation). The request is forwarded via the plant network and the wireless network to the sensors/actuators.
2. The sensors/actuators process the request and send a replay in upstream direction to the controller. This reply may contain an acknowledgement or a measurement reading.

An example of a sensor/actuator device-initiated transaction service flow:

1. The sensor sends a measurement reading to the process controller. The request is forwarded via the wireless (mesh) network and the plant network.
2. The process controller may send an acknowledgement in opposite direction.

For both controller- and sensor/actuator-initiated service flows, upstream and downstream transactions may occur asynchronously.

Figure D.2-1 depicts how communication may occur in discrete automation. In this use case, communication runs between process controller and sensor/actuator device via the plant network and the wireless (mesh) network. The wireless (mesh) network may also support access for handheld devices for supervisory control or process monitoring purposes.

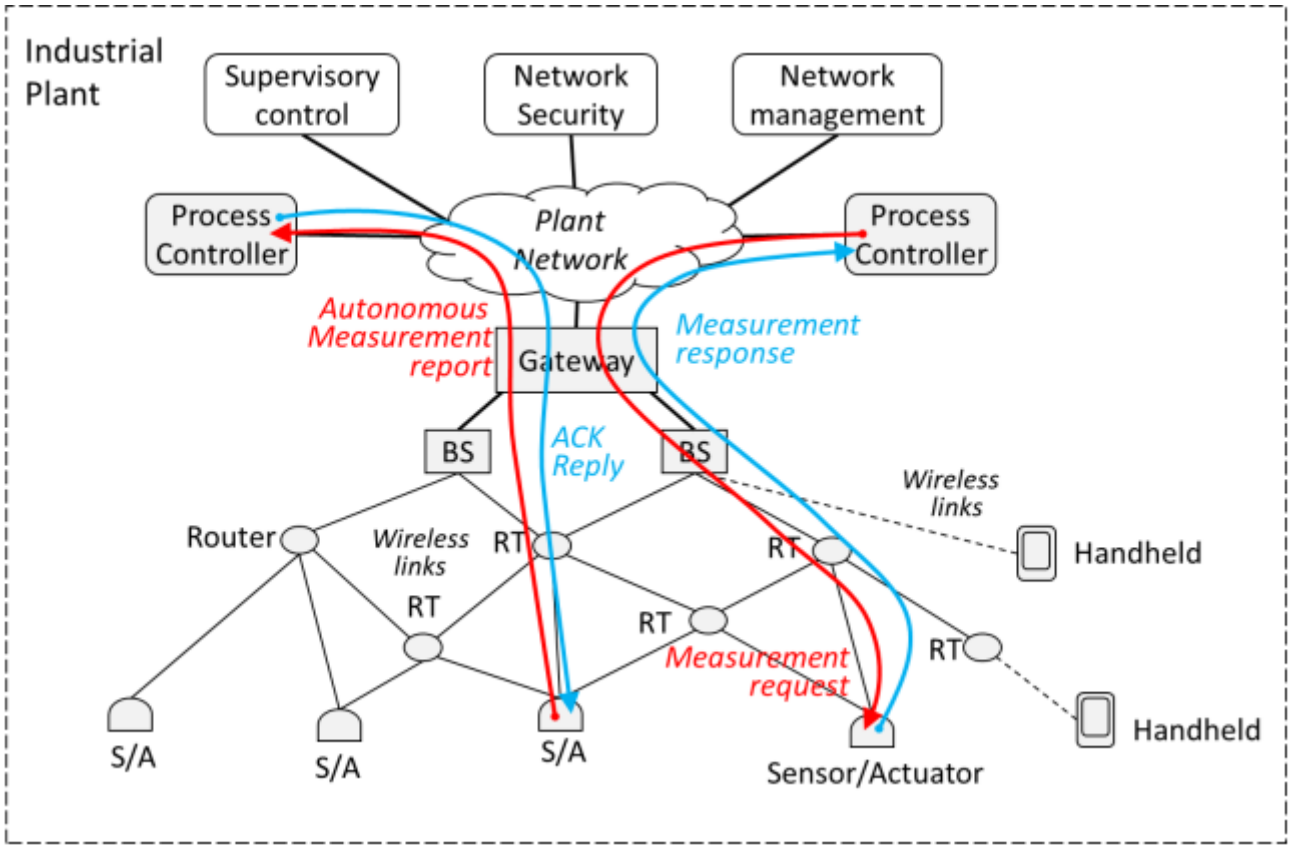


Figure D.2-1: Communication path for service flows between process controllers and sensor/actuator devices. Left-hand side: Step 1 (red) – the sensor/actuator (S/A) sends measurement report autonomously, Step 2 (blue) controller acknowledges. Right-hand side: Step 1 (red) - controller requests sensor data (or an actuator to conduct actuation), Step 2 (blue): S/A sends measurement information (or acknowledges actuation) to controller.

D.2.1 Service area and connection density

Factory halls can be rather large and even quite high. We set the upper limit at 1000 x 1000 x 30 m. Note that the connection density might vary quite a bit throughout factory halls. It is, for instance much higher along an assembly line than in an overflow buffer. Also, the density usually increases towards the factory floor. Typically, there is at least one connection per 10 m2, which results in a connection density of up to 105 km-2.

D.2.2 Security

Network access and authorization in an industrial factory deployment is typically provided and managed by the factory owner with its ID management, authentication, confidentiality and integrity.

A comprehensive security framework for factories has been described in IEC 62443.

D.3 Process automation

Process automation has much in common with discrete automation (see Annex D.2). Instead of discrete products (cars, chocolate bars, etc.), process automation addresses the production of bulk products such as petrol and reactive gases. In contrast to discrete automation, motion control is of limited or no importance. Typical end-to-end latencies are 50 ms. User experienced data rates, communication service availability, and connection density vary noticeably between applications.

Below we describe one emerging use case (remote control via mobile computational units, see Annex D.3.1) and a contemporary use case (monitoring, see Annex D.3.2).

Note that discrete automation fieldbuses (see Annex D.2) are also used in process automation.

D.3.1 Remote control

Some of the interactions within a plant are conducted by automated control applications like those described in Annex D.2. Here too, sensor output is requested in a cyclic fashion, and actuator commands are sent via the communication network between a controller and the actuator. Furthermore, there is an emerging need for the control of the plant from personnel on location. Typically, monitoring and managing of distributed control systems takes place in a dedicated control room.

Staff deployment to the plant itself occurs, for instance, during construction and commissioning of a plant and in the start-up phase of the processes. In this scenario, the locally deployed staff taps into the same real-time data as provided to the control room. These remote applications require high data rates (~ 100 Mps) since the staff on location needs to view inaccessible locations with high definition (e.g. emergency valves) and since their colleagues in the control room benefit from high-definition footage from body cameras (HD or even 4K).

For both kinds of applications, a very high communication service availability is needed (99,9999%). Typically, only a few control loops are fully automated and only handful of control personnel is deployed on location, so that the connection density is rather modest (~ 1000 km-2).

D.3.2 Monitoring

The monitoring of states, e.g. the level of the liquid of process reactors, is a paramount task. Due to the ever-changing states, measurement data is either pulled or pushed from the sensors in a cyclic manner. Some sensors are more conveniently accessed via wireless links, and monitoring via handheld terminals of these sensors during, e.g. maintenance is also on the rise. This kind of application entails rather modest user experienced data rates (~ 1 Mps), and since this kind of data is "only" indicator for, e.g. what process should be stopped in order to avoid an overflow, and not for automated control loops, the requirement on communication service availability is comparably low (99,9%). Note that emergency valves and such typically are operated locally and do not rely on communication networks. However, many sensors are deployed in chemical plants etc., so that connection density can readily reach 10 000 km-2.

D.3.3 Service area

While, for instance, chemical plants and refineries readily can span over several square kilometers, the dedicated control rooms are typically only responsible for a subset of that area. Such subsets are often referred to as plant, and their typical size is 300 x 300 x 50 m.

5G, the edge, and the disruption of the cloud: Why now is the time for change[16]

If one were to put together a linguistic analysis of all the conversations held at MWC 2020, later this month in Barcelona, there is a fair chance that the most spoken word would be 5G.

Not surprisingly, the term will be everywhere this year – much of course as it was last year and the year before. Yet whether it is smartphone vendors looking to showcase their latest, speediest wares, or thought leaders looking to where the enterprise needs to focus, things have turned up a notch over the past 12 months.

Take, for instance, what Bejoy Pankajakshan, executive vice president of Mavenir, had to say for sister publication Telecoms last month. The need for discussion is vital for future strategy, Pankajakshan affirms, as the options are legion. "A 5G network is envisaged as the most open, powerful, flexible, and advanced network the telecoms world has ever seen," he wrote. "At its heart, [it] is a software network and its development and deployment require a new approach and a new way of thinking.

"If a 5G network isn't built the right way, users may not come to the telco, and the OTTs could win again."

Getting everything moving, across various stakeholders, will be no easy task. Charting a blueprint for 5G in concert with other technologies requires detailed planning. Take a session from Accenture set to take place on February 25 around unlocking the power of the cloud. The rise of edge computing, setting the stage for network transformation, will bring huge long-term benefits, but immediate challenges.

"Network cloudification and the disaggregation of hardware and software become necessary – with CSPs now embarking on the critical journey to move their networks to be fully in the cloud, built around cloud at the edge and mobile edge compute, with the ability to cater for all the new applications and use cases unlocked by 5G," the session materials read.

For some, **5G will inevitably disrupt cloud computing as we know it today**. Writing for this publication in August, Marty Puranik, founder, president and CEO of Atlantic.Net, noted how 5G would effectively kill latency – and in one fell swoop, **potentially eradicate the need for cloud solutions**.

"One of the main reasons the cloud is so beneficial is for numerous devices – either in an organization for a private cloud or any user with an Internet connection for a public cloud – to connect to and transmit data with a central machine or hard drive located on the cloud," Puranik wrote. "**For an employee to share a large video file with a colleague who's working from home that day, the cloud made it simple. But why go through all that if your device can connect with your colleague's device with only a millisecond of latency and a minimum connection speed of 20 Gbps down and 10 Gbps up?**"

Edge computing, essentially the older, more streetwise sister of cloud computing, is expected to receive a lot of attention in Barcelona. Microsoft for instance, from whom the taglines of intelligent cloud and intelligent edge are never far away, are expected to be unveiling edge computing services at MWC.

[16] Cloudcomputing-news.net link

A report from Omdia, which looks to preview 5G developments at MWC20, noted that **5G and AI technologies could utilize edge computing, to the detriment of the cloud**. "By 2025, two of three smartphones will include built-in AI capabilities, and global revenue for AI smartphones will increase to $378 billion," the report notes. "**To alleviate consumers' privacy concerns, smartphone and smart speaker manufacturers will introduce 5G products which perform visual AI processing tasks on edge servers and appliances, bypassing the privacy risks involved in sending data to the cloud.**"

The current technological landscape feels like the calm before the storm. Organizations need to fully research the terrain and find out the best use cases for edge, 5G and AI to ensure smooth sailing ahead.

MEC, the Edge Computing "Killer Application"?

The killer application is Mobile Edge Computing MEC triggered by the progress of radio technologies (5G). ETSI soon rebranded it Multi-Access Edge Computing in 2017 to encompass a shift in both networking (convergence) and computing business model and not just standardizing a next generation mobile technology.

EDGE computing first applications are likely to be V2X[17]: Vehicle to everything communication, to embrace a new era of safer vehicle driving: it is likely to be one of the most useful Edge computing application (may be not MEC compliant day 1). We introduce a GSMA white paper about this topic.

[17] Not sure this V2x feature will be implement with ETSI MEC however…

Mobile Edge Computing vs. Multi-Access Edge Computing[18]

The acronym MEC is used interchangeably to stand for mobile edge computing or multi-access edge computing. What is the definition of mobile edge computing versus multi-access edge computing? The definition is the same for both terms with one small distinction that formed during the evolution of MEC research.

MEC computing refers to computing at the edge of a network. The edge is like a distributed cloud with proximity close to the end user that delivers ultra-low latency, reliability, and scalability. When first conceptualized, the edge of a network meant the edge of a mobile network, hence the name mobile edge computing. As MEC research progressed, technology experts realized that the term leaves out several access points that may also construct the edge of a network. Thus, prompted the change from mobile edge computing to multi-access computing to reflect that **the edge is not solely based on mobile networks**.

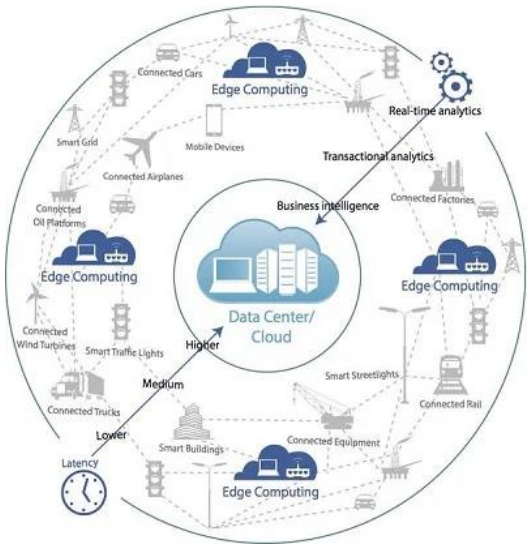
In September 2017, the European Telecommunications Standards Institute (ETSI) Industry Specification Group (ISG) officially changed its name from Mobile Edge Computing ISG to Multi-Access Edge Computing ISG to "to embrace the challenges in the second phase of work and better reflect non-cellular operators' requirements."

Tapping into the edge of a network elevates computing to handle the onslaught of connected devices, and it helps enterprises with their business-critical missions. Edge computing reduces latency to milliseconds and allows for constant connectivity. Plus, when the edge network experiences high traffic, the edge may offload data to the cloud to maintain a quick and reliable connection.

As we've noted the difference in terminology Let's review where the edge is it located with the types of access points currently in use, and the ones envisioned for the near-future computing that Create the Edge in MEC:

The access point is typically one hop away from the user. The access point can be either of these items to establish the network edge:

- Base Stations, including mobile base stations, cell towers, central office base stations
- RAN for LTE/5G
- Radio network controller for WiFi
- Cable modem termination systems (CMTS) for cable
- PON OLT for fiber or the access points for other networks
  such as Zigbee, CBRS, LoRA, DSL, MuLTEfire, private LTE.
- Hot spots
- Small cells
- Data centers (and micro-data centers)
- Routers
- Switches
- WiFi access points



[18] 2018 sdxcentral article link

What's the Difference Between Edge Computing and MEC?[19]

The terms edge computing and multi-access edge computing are commonly used interchangeably. However, the two have important distinctions. Edge Computing is a concept, and MEC is a standard architecture.

Not Interchangeable

Within the broad topic of edge computing, MEC is the widely accepted standard that must be met for a technology to be considered edge computing. While not an industry mandate that products meet MEC standards to be billed as edge solutions, many vendors are building around the standard. The general term of edge computing covers the practice of offloading computing processes (and in some cases the handling of storage and networking resources) from the user's computer or device to a local network node or other computer.

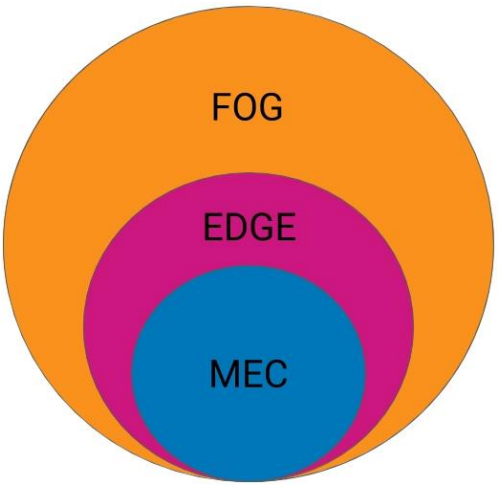


Figure 5:Edge computing is a more general concept than MEC and less general than fog computing.

Edge Computing

Edge Computing is an overarching term for the practice of placing compute and storage resources local to the end-user (at the "edge" of the network), to augment a distant central cloud. The edge can exist in places such as on the customer's premises or the bas e of cell towers. In addition to traditional server racks, edge computing can take place on smaller pieces of hardware like routers or WiFi hotspots where practical.

Not sending data to a distant central data center for computing power means faster network speeds, lower latencies, and more reliable network connections.

Standards by ETSI

ETSI created a catalog of over a dozen standards and papers for various aspects of edge computing, or to use the term they created and use, MEC. ETSI's MEC standards are guided by the following principles:

- Edge technology should have a virtualization platform to be considered MEC (ETSI uses their NFV architecture in the standard).
- MEC can be deployed at radio nodes, aggregation points, and the edge of the core network.
- APIs in a MEC environment should be simple, controllable, and if possible, reusable for other tasks.
- Since the compute, storage, and network resources that a MEC application requires may not match what are available at a node, a MEC network needs a system-wide lifecycle management of applications to handle these variables correctly.
- MEC systems must be able to relocate a mobile edge application running in an external cloud to a MEC host and back while fulfilling all the application's requirements (ETSI admits this principle needs further study).

[19] Sdxcentral April,2019 Article link

For more detail on these principles and MEC's generic technical requirements, check out the paper here.

Note that ETSI previously referred to MEC as Edge Computing. Over time, the field expanded the focus of edge computing beyond mobile implementations.

Cellular Vehicle-to-Everything C-V2X[20]

GSMA, Representing the worldwide mobile communications industry, has released a white paper in August 2019: "Connecting Vehicles Today and in the 5G Era with C-V2X (Cellular Vehicle-to-everything). Here are the few items to notice:

C-V2X?

The industry standard for vehicle communication, Cellular Vehicle-to-Everything (C-V2X) technology is commercially available globally today. Standardized by 3GPP, this 4G and 5G-based technology is designed to connect vehicles to each other, to roadside infrastructure, to pedestrians and cyclists, and to cloud-based services.

Many automakers are adding C-V2X connectivity to their vehicles because it has several key advantages over DSRC (a variant of Wi-Fi) and other technologies designed to enable vehicle-to-vehicle communications. C-V2X can:

- Provide levels of security, range, latency and reliability that have been proven to greatly exceed the capabilities of DSRC.
- Leverage the comprehensive coverage of secure and well-established LTE networks, reducing the amount of roadside infrastructure that needs to be installed and maintained by municipalities and highway agencies in both urban and rural areas.
- Enable highly reliable, real-time communication at high speeds and in high-density traffic.
- Support both short-range and long-range transmissions between vehicles and roadside infrastructure using one cost effective module.
- Harness 5G networks to support fully autonomous driving, as well as being backwards compatible with 4G.
- Leverage the robust security built into cellular networks.

C-V2X builds upon the services already offered in the connected vehicle today, such as pay-as-you drive insurance, vehicle diagnostics, eCall and connected infotainment by adding critical safety features that allow the vehicle to share data in real time. C-V2X helps vehicles avoid accidents, work together on the road and even detect road hazards beyond the driver's field of vision.

C-V2X employs two complementary transmission modes:

1. Short-range direct communications between vehicles (V2V), between vehicles and infrastructure (V2I), and vehicles and other road users (V2P), such as cyclists and pedestrians. In this mode, C-V2X works independently of the cellular networks in dedicated ITS 5.9GHz spectrum.
2. Long-range network communications (V2N), in which C-V2X employs the conventional mobile network to enable a vehicle to receive information about road conditions and traffic in the area, beyond the driver's line of sight.

How C-V2X Is Changing Driving?

C-V2X can be used in many ways to improve road safety, while making more efficient use of transport networks and infrastructure. For example, it can support:

- **Collision avoidance**: Each vehicle on the road could use C-V2X to broadcast its identity, position, speed and direction. An on-board computer could combine that data with that from

[20] Connecting Vehicles gsma 2019 link

other vehicles to build its own real-time map of the immediate surroundings and alert the driver to any potential collisions.

- **Platooning**: The formation of a convoy in which the vehicles are much closer together than can be safely achieved with human drivers, making better use of road space, saving fuel and making the transport of goods more efficient.
- **Cooperative driving**: By sharing sensor data, vehicles can use C-V2X to work together to minimize the disruption caused by lane changes and sudden braking.
- **Queue warning**: Roadside infrastructure can use C-V2X to warn vehicles of queues or road works ahead of them, so they can slow down smoothly and avoid hard braking.
- **Supporting the emergency services**: C-V2X can be used to warn road-users about emergency vehicles on route to an incident.
- **Hazards ahead warning**: C-V2X can be used to extend a vehicle's electronic horizon, so it can detect hazards around a blind corner, obscured by fog or other obstructions, such as high vehicles or undulations in the landscape.
- **Increasingly autonomous driving**: Along with other sensors and communications systems, C-V2X will play an important role in enabling vehicles to become increasingly autonomous.
- **Collecting road tolls**: designed to reduce congestion and the impact of motor transport on the environment through reduced emissions.
- **Avoiding vulnerable road-users**: by detecting pedestrians and cyclists' smartphones, C-V2X will help vehicles to avoid other road users.

The Roadmap for Deployment

C-V2X has the support of mobile operators, leading mobile equipment makers and many automakers, including Audi, BMW, Daimler, Ford, Lexus, Nissan, PSA, SAIC and Tesla. Ford, for example, has committed to deploying C-V2X in all new U.S. vehicle models beginning in 2022, "pending a technology neutral regulatory environment." It also plans to begin deploying C-V2X technology in Ford vehicles in China in 2021.

In Europe, Bosch, Huawei, and Vodafone Germany have successfully tested C-V2X on the A9 freeway in Bavaria, Germany, while Deutsche Telekom has announced C-V2X tests together with Skoda Auto in the Czech Republic as part of the European C-Roads project. Vodafone is also working with Continental to create a "digital safety-shield" for cyclists and pedestrians, using C-V2X direct communication and **edge computing** in the first 5G deployments. The 5G-ready tests are taking place under real-life conditions at Vodafone's 5G Mobility Lab in Aldenhoven, Germany

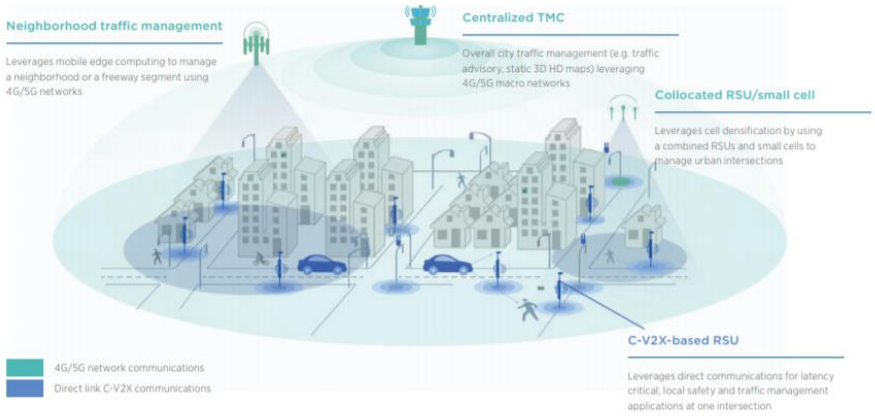
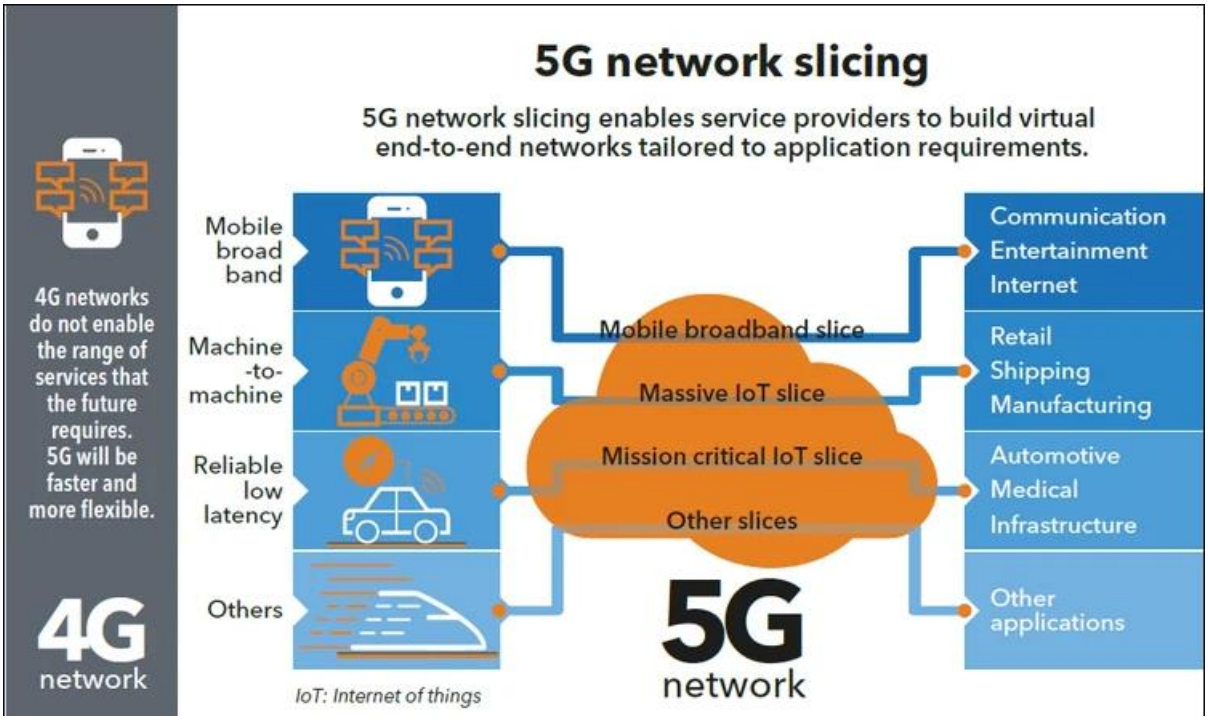


Figure 6: V2X architecture

What is 5G network slicing?



ZTE, a leading supplier of communication networking from China says[21]: "In the 5G "Slicing Store", consumers or enterprise customers can select the predefined slice template and set the SLA parameters according to the characteristics of the user or individual industry requirements. After the user logs into the online sliced based service store, the user can choose services with different SLA, such as guaranteed bandwidth, maximum latency, etc."

In sdxcentral[22]: "One of the most innovative aspects of the 5G architecture will be its reliance on 5G network slicing, which will let operators provide portions of their networks for specific customer uses cases — whether that use case is the smart home, the Internet of Things (IoT) factory, the connected car, or the smart energy grid.

Each use case receives a unique set of optimized resources and network topology — covering certain SLA-specified factors such as connectivity, speed, and capacity — that suit the needs of that application.

The Basics of 5G Network Slicing

Network slicing is a type of virtual networking architecture in the same family as software-defined networking (SDN) and network functions virtualization (NFV) — two closely related network virtualization technologies that are moving modern networks toward software-based automation. SDN and NFV allow far better network flexibility through the partitioning of network architectures into virtual elements. In essence, network slicing allows the creation of multiple virtual networks atop a shared physical infrastructure.

In this virtualized network scenario, physical components are secondary and logical (software-based) partitions are paramount, devoting capacity to certain purposes dynamically, according to need. As needs change, so can the devoted resources. Using common resources such as storage and

[21] See ZTE demonstrates the first 5G slicing store in Europe link
[22] Source: What Is 5G Network Slicing? link

processors, network slicing permits the creation of slices devoted to logical, self-contained, and partitioned network functions.

5G Network Slicing

According to 5G Americas, a clear benefit of 5G network slicing for network operators will be the ability to deploy only the functions necessary to support particular customers and particular market segments. “This results directly in savings compared to being required to deploy full functionality to support devices that will use only a part of that functionality. And a derivative benefit is the ability to deploy 5G systems more quickly because fewer functions need to be deployed, enabling faster time-to-market.”

Some vendors — such as Ericsson — believe that 5G network slicing will be the key ingredient necessary for 5G to meet its technical requirements. The new era of 5G connectivity will be characterized by its wide diversity of use cases and their varied requirements in terms of power, bandwidth, and speed. According to Ericsson, “The greater elasticity brought about by network slicing will help to address the cost, efficiency, and flexibility requirements imposed by future.”

Network Slicing Is Essential to 5G

GSMA Intelligence estimates that there will be 1.2 billion 5G connections by 2025, accounting for 40 percent of the global population, or approximately 2.7 billion people. It hypothesizes that the coming 5G network architecture is “a real opportunity to create an agile network that adapts to the different needs of specific industries and the economy.” And a key enabler of that 5G reality will be network slicing.

MEC in 5G Network Architecture

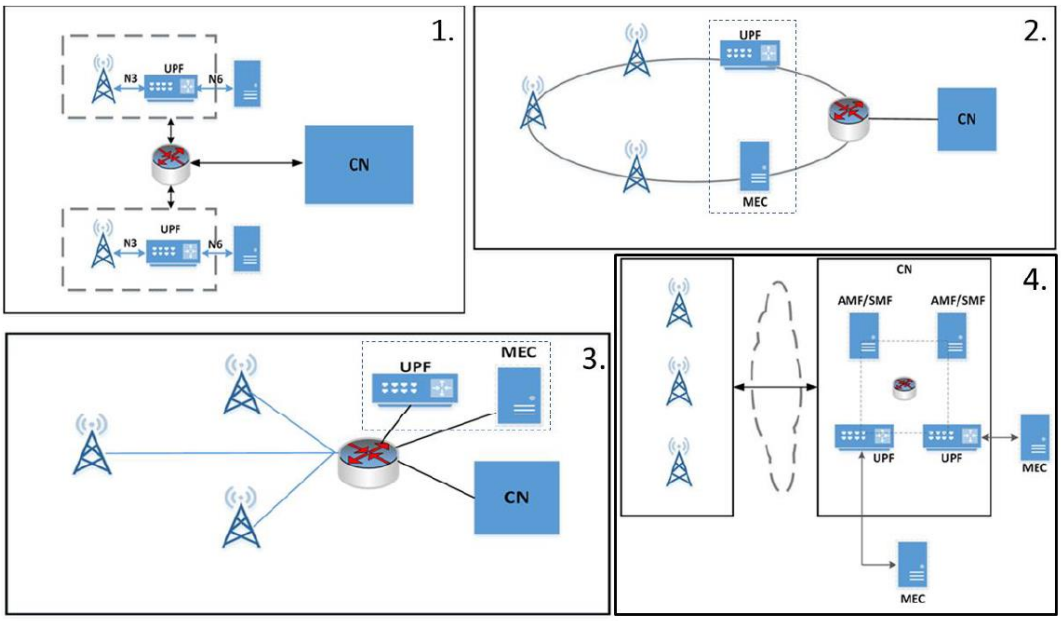


Figure 3. Examples of the physical deployment of MEC.

1. MEC and the local UPF collocated with the Base Station.
2. MEC collocated with a transmission node, possibly with a local UPF
3. MEC and the local UPF collocated with a network aggregation point
4. MEC collocated with the Core Network functions (i.e. in the same data center)

Provide more from ETSI MEC in 5G networks

Who are the Edge Computing players?

Communications service provider (CSP)

MEC is a game changer and CSP are historical provider of communication but lack of experience in deployment of End-to-End services which is a core capability of MEC business architecture.

According to Ovum[23],

> "to achieve the service-related goals of 5G and deliver on the business case, computing and connectivity need to converge, and MEC provides the best available technical solution if network operators are to move beyond offering an enhanced mobile broadband service and begin to offer new services and penetrate new markets such as the enterprise.
>
> As CSPs move toward a software-based cloud-network architecture in preparation for 5G, the case for the development of an edge-computing model will become stronger. The combination of network functions virtualization (NFV) and telco cloud is providing CSPs with the foundation of a platform for MEC, given additional support by the move from today's virtual machine (VM) model to a cloud-native, container-based reference architecture."

According to Gartner[24], communications service providers (CSP) offers telecommunications services or some combination of information and media services, content, entertainment and application services over networks, leveraging the network infrastructure as a rich, functional platform. CSPs include the following categories: telecommunications carrier, content and application service provider (CASP), cable service provider, satellite broadcasting operator, and cloud communications service provider.

CGI[25] says: "What a difference a decade has made for communication service providers (CSPs). The original flagships of the industry are gone—AT&T, MCI, BellSouth and Qwest. What governments didn't manage to do through legislation, deregulation and the destruction of monopolies, IP technology has done by turning what were once high margin services into commodity applications—essentially democratizing the industry while at the same time eroding margins."

This can be seen for Europe (Sp,It,Fr,Ger,UK) CSP revenues on the following graphics

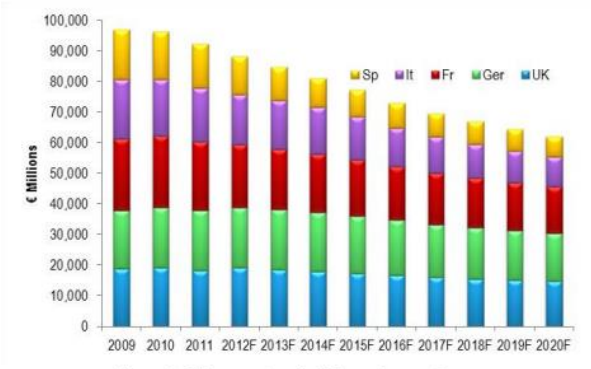


Figure 1: EU revenue trends in five major countries
(Source: European regulators, Mobile operators, Barclays Capital, STL Partners assumptions and analysis)

The competition landscape has changed from one of minimal competition, where traditional CSPs had a monopoly in designated functional and geographical areas, to one of heavy competition where traditional CSPs are competing against one another and against a multitude of new, agile CSPs.

[23] Driving New Business Opportunities with Multi-Access Edge Computing and 5G link
[24] Gartner Glossary link
[25] Communication service providers in the next decade link

But even if this industry is under technology disruption and revenue pressure (-20% in 10 years), they have strong organizations, like ETSI who is leading the standardization of the MEC (Mobile Edge Computing) technologies and subsequent business models.

"Tower companies"[26]

We can qualify "Tower Companies" as a new entrant in the MEC business. But who are they?

The increasing data usage by consumers has led to a competitive telecom landscape across regions. With **tower sharing becoming increasingly popular** among the Mobile Network operators (MNOs, a category of CSP), the tower operators have been able to reach operational efficiencies. **Independent tower companies** owned around 70% of the total 4.10 million towers around the world (as of 2017). China has the highest number of telecom towers in the world, owned by the state-run China Tower Corporation. It has around 1,968,000 towers and it was estimated that it is leasing over 550,000 towers.

The telecom towers market was valued at 4.82 million units in 2019 and is expected to reach 6.29 million by 2025, at a CAGR of 4.56% over the forecast period 2020 - 2025. The leasing concept has enabled the MNOs to invest heavily in developing their infrastructure and reach across rural regions, thus bringing new revenues to the tower operators through tower installations. Tower-sharing is one of the major growth drivers for the telecom industry, as it provides benefits like cost reduction and faster data rollout. The telecom tower industry has gained high prominence as an independent industry, mainly in India and the United States.

Even in Europe this market is moving fast. For example, FPS [27] is the raising « tower company » in France. The company, which specializes in hosting telecom equipment for operators, was created at the end of 2012, after the acquisition of the pylons of Bouygues Telecom. It operates 2,500 pylons and 20,000 flat roofs, which it makes available to operators. The company employs 90 people and shows growth of around 40% per year. Its turnover should amount to 70 million euros in 2016, for an Ebitda (gross operating surplus) of 45 million.

In US[28], Packet is already running one edge computing location for SBA in Boston and two in Chicago for Crown Castle and the tower company's edge computing partner, Vapor IO. Packet is also gearing up to launch an edge computing location for American Tower (one of the biggest "tower companies" in US).

Packet is having some success in the area: Packet is already supplying edge computing services in 190 locations for Sprint Corp. (NYSE: S)'s newly established IoT service, called Curiosity. Further, Packet's products are powering some initial edge computing deployments from all the nation's major tower companies: Crown Castle International Corp. (NYSE: CCI), SBA Communications and American Tower Corp. (NYSE: AMT).

Altogether Packet hopes to have 15 edge computing locations up and running this year, on the way to the company's eventual goal of scaling up to 50 edge computing sites

[26] TELECOM TOWERS MARKET - GROWTH, TRENDS AND FORECAST (2020 - 2025) link
[27] Les ECHOS 2016 : Le marché français des « tower companies » en pleine mutation link
[28] All Major Tower Companies Are Sniffing Around Edge Computing link

Rank	Company	Tower Count
1	American Tower	40586
2	Crown Castle	40039
3	SBA Communications	14873
4	United States Cellular Co.	4207
5	Vertical Bridge	3198
6	InSite Wireless Group	1196
7	Peppertree Capital (AT&T Towers to close 2019/2020)	1023
8	BNSF Railroad	941
9	Time Warner	653
10	Uniti	651
11	Others	16309
		123 676

Figure 7: Some Tower Companies in US[29]

The Cloud Contenders

Other Contenders

[29] List of Tower Companies in US link

Edge Computing for Industrial Automation

Scenario for an EDGE offer landscape

Role of Industrial Automation players