Mobile Edge Computing in 5G Era

Manjit Singh
B-54, Sector 14, Noida, 201301 UP India.
ms.54@dataone.in

Abstract -- **Mobile Edge computing, now called Multi-access edge computing (MEC) after ETSI changed the name in 2017 is an emerging ecosystem, which aims at converging telecommunication and IT services, providing a cloud computing platform at the edge of the radio access network. MEC provides a highly distributed computing environment that can be used to deploy applications and services as well as to store and process content in close proximity to mobile users. Edge application services reduce the volumes of data that must be moved, the consequent traffic, and the distance that data must travel. That provides lower latency and reduces transmission costs.**

With more data being created each day, the edge will become increasingly important to activate products, services and business models in a faster and more operationally efficient manner. Driven by the emergence of new compute-intensive applications and the vision of the Internet of Things (IoT), it is foreseen that the emerging 5G network will face an unprecedented increase in traffic volume and computation demands. MEC offers storage and computational resources at the edge, reducing latency for mobile end users and utilizing more efficiently the mobile backhaul and core networks. Though technically highly-feasible, the commercial and logistical challenges of mobile edge computing are such as to make this a rich seam for a technology analyst.

This paper introduces a survey on MEC and focuses on the fundamental key enabling technologies. It elaborates MEC deployment scenarios considering both individual services and a network of MEC platforms supporting mobility, bringing light into the different deployment options. In addition, this paper analyses the MEC reference architecture and main deployment scenarios, which offer multitenancy support for application developers, content providers, and third parties. Finally, this paper overviews the current standardization activities and elaborates further on open research challenges.

Keywords: Edge computing, Mobile edge computing, Ultra low latency, 5G, Cloud gaming, Autonomous driving, Data sovereignty, Smart cities

I. INTRODUCTION

EDGE COMPUTING is a distributed computing paradigm which brings computation and data storage closer to the location where it is needed, to improve response times and save bandwidth. The origins of edge computing lie in content delivery networks that were created in the late 1990s to serve web and video content from edge servers that were deployed close to users. In the early 2000s, these networks evolved to host applications and application components at the edge servers, resulting in the first commercial edge computing services that hosted applications such as dealer locators, shopping carts, real-time data aggregators, and ad insertion engines. Modern edge computing significantly extends this approach through virtualization technology that makes it easier to deploy and run a wider range of applications on the edge servers.

One definition of edge computing is any type of computer program that delivers low latency nearer to the requests and more specifically in applications where real-time processing of data is required. In this definition, cloud computing operates on big data while edge computing operates on 'instant data' that is real-time data generated by sensors or users.

The full capabilities of 5G can only be realized by using edge computing in some form. 5G networks will need edge computing to support a range of ultra-low-latency use cases, and so 5G will also be the major driver of edge deployments.

Enterprises will be a key early target for edge computing services, along with services for specific industry verticals and mission critical industrial processes. Services such as AR/VR and cloud gaming are also generating a great deal of interest, with autonomous driving being a potentially major market further into the future.

The industry has generally recognised that edge computing can deliver multiple benefits. Pushing computing to the edge will allow mobile operators to support new services for customers, especially applications reliant on cloud computing and demanding low latency. It will also help them to cut spending on backhaul and avoid congestion in the core.There are also non-technical reasons why edge computing is desirable. Data sovereignty requirements in many countries mean data storage and processing must happen within national borders. Meanwhile, some companies are concerned about data leaks and demanding that information be stored on their own premises.

II. NEED FOR EDGE COMPUTING

The Cisco white paper [1] showed that global data traffic will grow at a compound annual growth rate (CAGR) of 26 percent

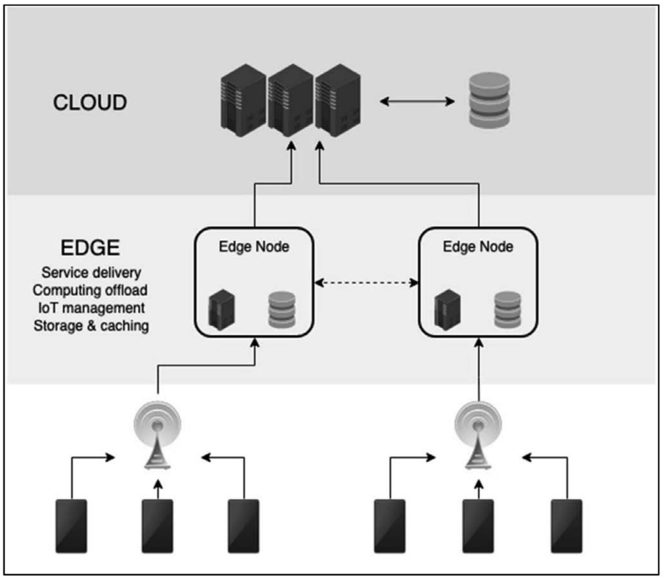


Figure 1. Concept of edge computing.

between 2017 and 2022 (*i.e*., increase over threefold) and reach 122 exabytes (EB) per month by 2022 (1 EB = 1 billion GB = 10^{18} bytes). Mobile and wireless networks carried 11.51 EB per month in 2017, 28.56 EB per month in 2018, and 77.49 EB per month by end of 2019. 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.

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 increase of IoT devices at the edge of the network is producing a massive amount of data to be computed to data centers, pushing network bandwidth requirements to the limit. Despite the improvements of network technology, data centers cannot guarantee acceptable transfer rates and response times, which could be a critical requirement for many applications. Furthermore, devices at the edge constantly consume data coming from the cloud, forcing companies to build content delivery networks to decentralize data and service provisioning, leveraging physical proximity to the end user. In a similar way, the aim of Edge Computing is to move the computation away from data centers towards the edge of the network, exploiting smart objects, mobile phones or network gateways to perform tasks and provide services on behalf of the cloud. By moving services to the edge, it is possible to provide content caching, service delivery, storage and IoT management resulting in better response times and transfer rates. At the same time, distributing the logic in different network nodes introduces new issues and challenges.

Relevance of cloud computing to mobile networks is on an upward spiral. Social network services like Facebook and Twitter, the content from YouTube and Netflix, and navigation tools from Google Maps are all on clouds. Besides, users' increasing reliance on mobile devices to carry out compute and storage intensive operations, whether personal or business related, require offloading to the clouds for achieving better performance extending battery life. These objectives would be difficult and expensive to realize without bringing the cloud closer to the edge of the network and to the users. In response to this requirement, the mobile operators are working on MEC in which the computing, storage and networking resources are integrated with the base station. Compute intensive and latency sensitive applications like augmented reality and image processing can be hosted at the edge of the network. Figure 2 shows this concept.

MEC provides a highly distributed computing environment that can be used to deploy applications and services as well as to store and process content in close proximity to mobile users. This would enable applications to be split into small tasks with some of the tasks performed at the local or regional clouds as long as the latency and accuracy are preserved. A number of challenging issues arise in distributing sub-tasks of an application among edge and other clouds. When splitting of an application does happen, the mobile edge cloud takes care of the low latency, high bandwidth and locally relevant jobs.

The Internet, was designed to be open and easy to attach to and use. It wasn't optimised for any known or anticipated usage. Instead the Internet has adapted to support unimagined uses that far exceeded any initial expectations. There is a lot in common with today's mobile edge computing.

Bringing compute, storage and processing functionality closer to the end user potentially opens up a wealth of new service opportunities, with support for new applications and use cases. Though technically highly-feasible, the commercial and logistical challenges of MEC are such as to make this a rich seam for a technology analyst.

By using MEC technology, a cellular operator can efficiently deploy new services for specific customers or classes of customers. The technology also reduces the signal load of the core network, and can host applications and services in a less costly way. It also collects data about storage, network

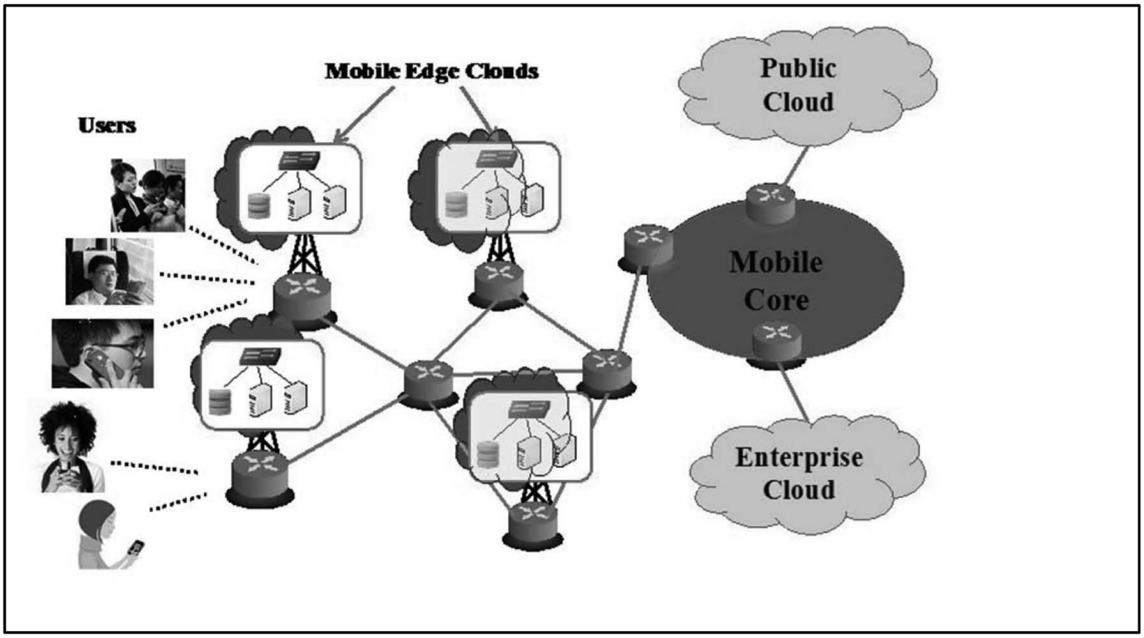


Figure 2. Concept of mobile edge clouds and computing.

bandwidth, CPU utilization, etc. for each application or service deployed by a third party. Application developers and content providers can take advantage of close proximity to cellular subscribers and real-time RAN (Radio Access Network) information.

MEC has been created using open standards and application programming interfaces (APIs), using common programming models, relevant tool chains and software development kits to encourage and expedite the development of new applications for the new MEC environment.

Edge application services reduce the volumes of data that must be moved, the consequent traffic, and the distance that data must travel. That provides lower latency and reduces transmission costs. Computation offloading for real-time applications, such as facial recognition algorithms, showed considerable improvements in response times as demonstrated in early research. Further research showed that using resource-rich machines called cloudlets near mobile users, offering services typically found in the cloud, provided improvements in execution time when some of the tasks are offloaded to the edge node. On the other hand, offloading every task may result in a slowdown due to transfer times between device and nodes, so depending on the workload an optimal configuration can be defined.

Another use of the architecture is cloud gaming, where some aspects of a game could run in the cloud, while the rendered video is transferred to lightweight clients such as mobile, VR glasses, etc. Such type of streaming is also known as *pixel streaming*. Conventional cloud games may suffer from high latency and insufficient bandwidth, since the amount of data transferred is huge due to the high resolutions required by some services. Edge nodes used for game streaming are known as gamelets, which are usually one or two hops away from the client. The edge node is mostly one or two hops away from the mobile client to meet the response time constraints for real-time games in the cloud gaming context.

Other notable applications include connected, autonomous cars, smart cities, Industry 4.0 (smart industry) and home automation systems.

III. RELEVANCE OF MEC TO 5G

5G is a collective name for technologies and methods that would go into the future networks to meet the extreme capacity and performance demands. The phrase ‘no latency, gigabit experience’ summarizes the user expectations that the industry is aspiring to meet. Both of the major standardization bodies, International Telecommunications Union (ITU) and European Telecommunications Standards Institute (ETSI) have initiated activities relating to 5G with commercial deployments expected in 2020.

5G will support not only **C**ommunication, but also **C**omputation, **C**ontrol, and **C**ontent 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.

5G gives ultra latency in transmitting large high volume of data. Bottlenecks may occur if cloud computing is not able to respond at high speed for which 5G technology is being promoted as a key point. So processing of data at the edge of the network will be the essential part of the 5G deployment.

Some of the key performance parameters targeted to be achieved in 5G networks are: per device data rates up to 20 Gbps, less than 1ms latency contribution of the radio part, mobility at 500 km/hour and terminal localization within 1 meter. It will aim for service continuity in trains, sparse and dense areas, support for connecting 20 million user devices and more than a trillion Internet of Things (IoT)/Machine to Machine (M2M) devices with high reliability.

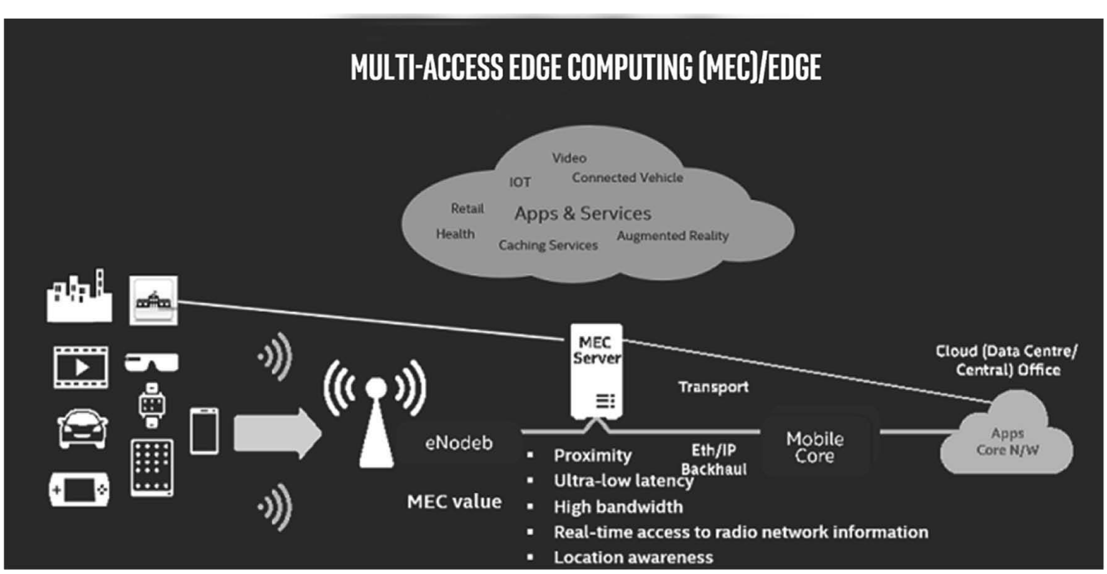
The greatest benefits of 5G to the telcos would appear to be diversification as opposed to increased squeeze on the wallets of consumers. With more data being created each day, the edge will become increasingly important to activate products, services and business models in a faster and more operationally efficient manner. Enterprise organizations will largely be unaware of how to reap the greatest benefits, a pleasant niche the telcos could certainly profit from.

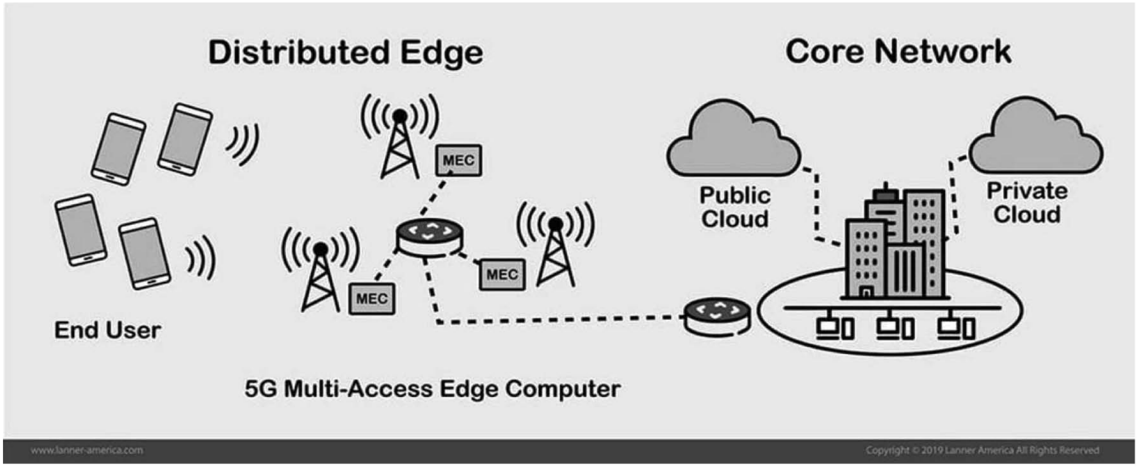
Edge computing refers to distributed data centers that reduce the physical distance between the cloud and the edge of the network – *i.e.* the RAN. The main point of this is to reduce the lag resulting from interacting with the cloud in real time and to allow the kind of low-latency communication services that promise to be the most novel new feature of 5G. Edge computing is also expected to help with things like bandwidth flexibility for IoT, cloud security and data localisation.

5G technology will make the happenings of Robots, AR/VR , Image/Video recognition, Gaming everywhere feasible. Telcos will reap rich harvests from these facilities to their enterprises. Expectations are high with deployment of 5G technology, therefore MEC will be the front runner for the network. Edge computing, which is tied to network transformation towards NFV and #CUPS (Control & User Plane Separation), is already happening in 4G, and the evolution from Edge Computing to Edge Cloud is a cloud play, not a telco play . It is also argued that 5G probably needs the edge more than the edge needs 5G.

This presents a unique opportunity for operators, for whom distributed infrastructure is a core competence. They also own, or at least have access to, a lot of remote locations, so they have a head start over cloud specialists and IT companies. Edge computing was said to be the perfect example of the convergence of networks and IT, which raises the question of which of those worlds will define and own it. Driven by the emergence of new compute-intensive applications and the vision of the Internet of Things (IoT), it is foreseen that the emerging 5G network will face an unprecedented increase in traffic volume and computation demands.

However, end users mostly have limited storage capacities and finite processing capabilities, thus how to run compute-intensive applications on resource-constrained users has recently become a natural concern. MEC, a key technology in the emerging 5G network, can optimize mobile resources by hosting compute-intensive applications, process large data before sending to the cloud, provide the cloud computing capabilities within the radio access network in close proximity to mobile users, and offer context-aware services with the help of RAN information. Therefore, MEC enables a wide





variety of applications, where the real-time response is strictly required, *e.g.,* driverless vehicles, augmented reality, robotics, and immerse media. Indeed, the paradigm shift from 4G to 5G could become a reality with the advent of new technological concepts. The successful realization of MEC in the 5G network is still in its infancy and demands for constant efforts from both academic and industry communities.

After several years of preparation, Japan and South Korea are pioneers in development of 5G Technology. They have developed networks with AI, IOT, NVF and MEC.

IV. KEY TECHNOLOGIES AND STATUS

MEC, formerly mobile edge computing, is an ETSI-defined network architecture concept that enables cloud computing capabilities and an IT service environment at the edge of the cellular network and, more in general at the edge of any network. MEC technology is designed to be implemented at the cellular base stations or other edge nodes, and enables flexible and rapid deployment of new applications and services for customers. Combining elements of information technology and telecommunications networking, MEC also allows cellular operators to open their radio access network to authorized third parties, such as application developers and content providers.

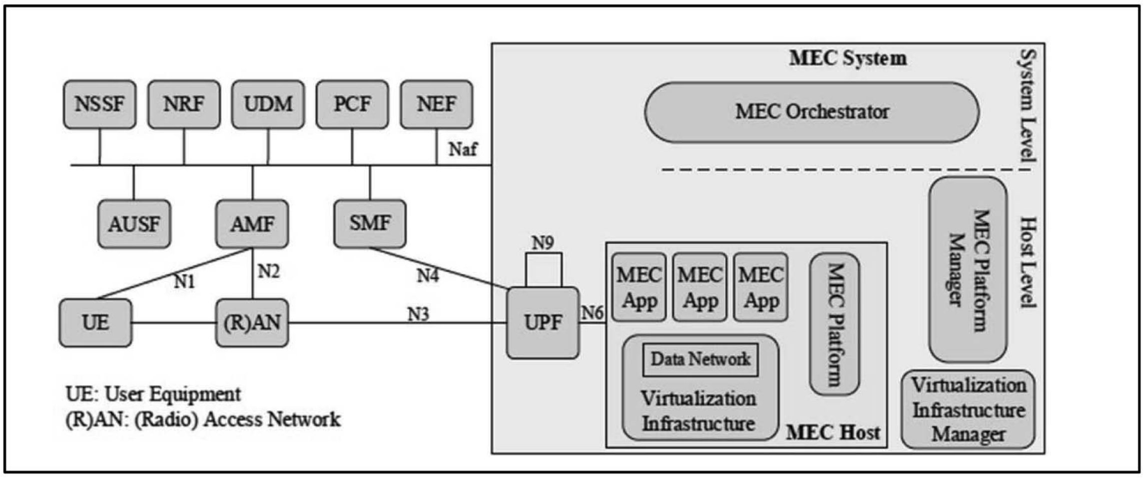
In late 2014, the European Telecommunications Standards Institute (ETSI) Mobile Edge Computing Industry Specification Group (MEC ISG) initiated the MEC concept. As a complement of the C-RAN architecture, MEC aims to unite the telecommunication and IT cloud services to provide the cloud-computing capabilities within radio access networks in the close vicinity of mobile users. Therefore, MEC enables a wide variety of applications. In order to reap additional benefits of MEC with heterogeneous access technologies, *e.g.*, 4G, 5G, WiFi, and fixed connection, ETSIISG officially changed the name of mobile edge computing to mean multi-access edge computing in 2017 .

After this scope expansion, MEC servers can be deployed by the network operators at various locations within RAN and/or collocated with different elements of the network edge, such as BSs (*aka* eNB in 4G and gNB in 5G), optical network units, radio network controller sites, and WiFi access points. This transformation pushes intelligence towards the edge so that not only communication functionalities but also computation, caching, and control services can be better facilitated.

The design of 5G networks would revolve around virtualization and programmability of networks and services. It is envisioned that transition to 5G will be facilitated by today's emerging technologies such as Software Defined Networking (SDN), Network Functions Virtualization (NFV), Mobile Edge Computing (MEC) and Fog Computing (FC). SDN and NFV provide new tools that enhance flexibility in designing networks. These complementary technologies enable programmability of control and network functions and eventual migration of these key constituents of the network to the cloud.

The network functions defined in the 5G architecture and their roles can be briefly summarized as follows.

a) AMF: Access and Mobility Management Function (AMF) establishes mobility and access procedures, *e.g.*, connection management, reachability management, mobility event notification, termination of the RAN control plane, and access authentication/authorization.
b) SMF: Session Management Function (SMF) performs functionalities related to session management, *e.g.,* session establishment, termination of interfaces towards policy control functions, and downlink data notification.
c) NSSF: Network Slice Selection Function (NSSF) executes the allocation of slicing resources and AMF set to serve users.
d) NRF: Network Repository Function (NRF) supports the discovery of network functions and their supported services.
e) UDM: Unified Data Management (UDM) handles user subscription and identification services.



f) PCF: Policy Control Function (PCF) unifies the network policies and provides policy rules to control plane functions.
g) NEF: Network Exposure Function (NEF) acts as a service aware border gateway for providing secure communication with the services supported by the network functions.
h) AUSF: Authentication Server Function (AUSF) performs authentication procedures.
i) UPF: User Plane Function (UPF) provides functionalities to facilitate user plane operations, *e.g.*, packet routing and forwarding, data buffering, and allocation of IP address.

MEC model has three components, the user, the edge server, and the remote server. The MEC host placed at the edge of the network consists of an MEC platform and resources to store and process data. The MEC platform is secured where applications are offered and serviced by the edge server.

Fog Computing: Fog computing, a term put forward by Cisco in 2012, refers to the extension of the cloud computing from the core to the network edge, thus it reduces the amount of data needed to transfer to the central cloud. Therefore, most intensive computations from and data collected by end users can be processed and analyzed by fog nodes (*i.e.*, nodes in fog computing) at the network edge, thus reducing the execution latency and network congestion . Since fog nodes are generally deployed in distributed locations, they are widespread and geographically available in large numbers. Due to its characteristics, fog computing plays an important role in many use cases and applications *e.g.*, smart cities, connected vehicle, smart grid, wireless sensor and actuator networks, smart buildings, and decentralized smart building control. However, a fog node cannot act as a self-managed cloud data center (DC) and needs the support of the cloud.

Cloudlets: Cloudlets is a term coined by Mahadev Satyanarayana and a team of people working in Carnegie Mellon University and Intel labs to describe an intermediate component between the edge and the cloud. Cloudlet also called “the data centre in a box” is a mini cloud that has all the capabilities of a cloud but in smaller portions. It has an operating system, storage area and a platform that hosts applications. Whenever the connected device has to transmit data, it offloads onto the nearest cloudlet. The cloudlet stores to process this data. It might decide to pass on the data to the main data centre if it sees doing so is necessary.

Although the commercial implementation of edge computing has yet to happen in a big way, stakeholders in the telecom industry value chain are clearly motivated to address new use cases with emerging edge technologies. This is becoming more obvious with the accelerated rollout of 5G services in different parts of the world. A few recent high-profile cases include the following:

- AT&T is developing a 5G-based MEC solution for a retail environment in partnership with Badger Technologies. It will enable in-store robots to perform tasks that boost efficiency and improve the customer experience.
- Verizon tested edge computing on its 5G network, shortly after the 5G network went live. The carrier claimed that it was able to cut latency in half during tests. Shortly afterwards, it unveiled a MEC platform from which it expects to generate B2B revenues starting in 2021.
- Deutsche Telekom’s approach is more cautious. After a two-year internal analysis and assessment, the leading German operator established a separate entity, Mobiledge X, to develop edge computing on its behalf. The independent company has subsequently worked with other mobile operators to develop a software platform, Edge Navigator, which developers can use to write applications that take advantage of edge computing.
- Vodafone believes the market is quickly evolving from a centralised cloud to distributed cloud, expecting 75% of enterprise generated data will be processed outside of a centralised data centre by 2020. According to Vodafone’s

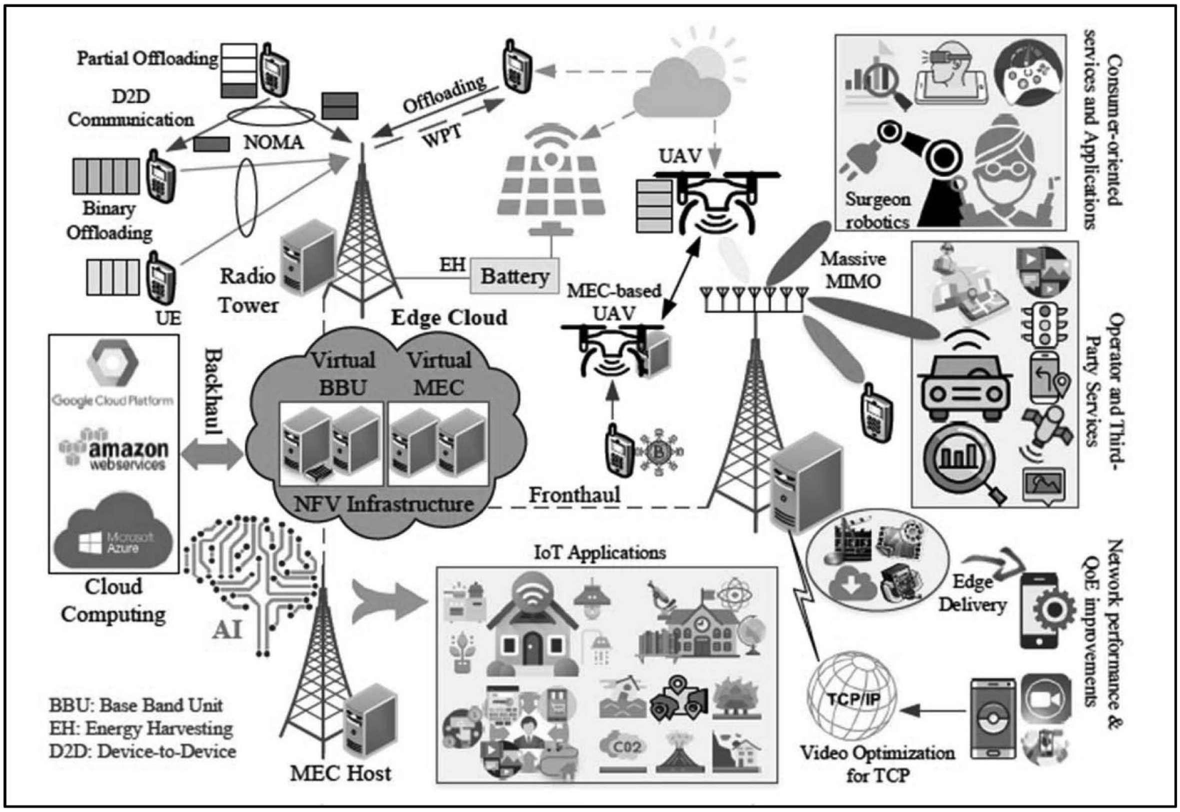


Figure 3. Integration of MEC with 5G technologies (Source IEEE paper Jan 2020).

data, 27% of businesses are already implementing edge, and a further 18% plan to do so the next year. The operator also predicted that, as a result of the trend towards edge cloud, 90% of customer deployments will be critically dependant on latency and bandwidth, the key technology properties 5G will offer. To serve the fast-moving market, Vodafone is pursuing a multi-cloud strategy and is offering enterprise customers with two different solutions: dedicated and distributed. The company is also working on a few use cases with its partners, including supporting connected factory with dedicated edge, next generation retail with augmented reality on the edge, and worker insights through augmented operation.

Telecom operators are not the only companies that have been active in edge computing. Many technology vendors, including the three biggest telecom equipment makers, have also been involved in trials and deployments, often in collaboration with operator partners. The following are a few examples:

- Huawei launched its MEC@Cloud Edge platform in 2016, in preparation for upcoming 5G launches.
- The vendor has already worked with China Telecom to test smart factory solutions and has also collaborated with the developer of a smart museum application on processing data at the network edge to improve the visitor experience.
- Ericsson has built 22 core locations across the world as local nodes for its unified delivery network, which is supporting the edge computing needs of operators and OTT (over-the-top) services.
- Nokia returned from last year's MWC to demonstrate virtual reality (VR) gaming based on its edge cloud, this time in collaboration with China Mobile. This happened six years after it introduced Radio Applications Cloud Server (RACS) and is based on the vendor's converged edge cloud solution, built on the AirFrame open edge servers that Nokia launched in the first half of 2018.
- Apart from the CSP (Communications Service Providers) and equipment suppliers, there are some web companies who are equally putting in their time, effort and money for development of edge computing as below:

■ Microsoft was the first to make it possible for customers to replicate the cloud environment on a smaller scale in their own data centres with its Azure Stack, which was launched in 2017.

■ Amazon has made some of the boldest moves. In 2018, AWS, the company's cloud service, launched Outposts, a scaled-down cloud solution that uses AWS functions but provides an on-premises managed service for enterprise customers.

- The latest arrival in the market is Google Cloud's Anthos, which can run applications on-premises and leverages the power of Google's cloud computing. What sets Google apart from Microsoft and Amazon is that Anthos can run on top of VMware vSphere, which means VMware customers using hardware from the mainstream original equipment manufacturers, like Cisco, Dell/EMC, HPE, and Lenovo, do not need to buy new equipment, making implementation of Anthos more cost-efficient.

V. USE CASES

While there is no standard way of classifying the different use cases for edge computing, 5G's technology properties can neatly map to edge requirements for low latency, high bandwidth and reliability. These capabilities of 5G technology can be best described in the following three parameters; namely

- enhanced mobile broadband (eMBB),
- ultra-reliable low-latency communication (uRLCC) and
- massive machine-typ ecommunications (mMTC).

Edge Computing can require different setups depending on the scenario. For example, if cutting latency is the priority, (uRLCC case) the base station could be the best edge location. In other cases, such as the provision of managed services, the edge could be located at the customer premises.

5G, by design a flattened, distributed network that is mainly software-based, will need to push data computing, storage and heterogeneous access as close to the network edge as possible. *Edge computing is therefore a fundamental enabler of 5G's capabilities:* enhanced mobile broadband access (eMBB); massive machine type communications (mMTC); and ultra-reliable low latency communications (uRLLC). In that sense, 5G operators, especially those aspiring to serve adjacent vertical industries, have little choice other than to actively invest in edge computing.

Hewlett Packard Enterprise (HPE) has analysed more than 200 cases in the last two years and put them into five categories: consumer services, enterprise AR/VR, enterprise analytics, IoT and subscriber services. HPE then categorized the most attractive cases to customers according to the 5G technology capabilities of enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (uRLCC) and massive machine-type communications (mMTC).

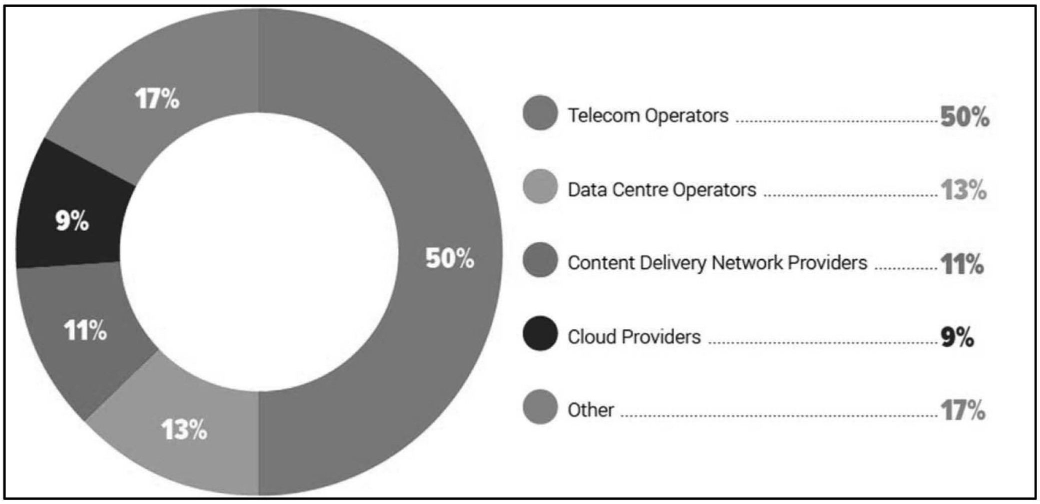
Communications service providers appear well-placed to build and run the edge cloud but they need to gain an early understanding as to how and where it can benefit them. Interest is growing among a number of other potential stakeholders. These include public cloud network providers who will want to build their own edge clouds to either sell to CSPs for their networks or to sell directly to end-users, internet players such as AWS, Google and Microsoft who will want to adapt their hyperscale cloud infrastructures to meet the demands of smaller workloads, and even owners of physical assets suitable for hosting edge cloud infrastructure such as tower companies.
Some of the Use cases of Mobile Edge Computing are enumerated as below:

IoT Gateways: Edge Computing and the Internet of Things (IoT) go hand-in-hand. With the explosion of new connected devices, everything from your car to your toaster now has an IP address. These new devices are producing a lot of data. So much data that your limited Internet uplink can't keep up. Connected devices can consume less backhaul bandwidth by processing the majority of that data at the Edge in an IoT Gateway close

Leading Edge Computing USE case with 5G

eMBB Use Cases	uRLCC Use Cases	mMTC Use Cases
High definition audio/video	Smart cities	Smart cities
AR/VR gaming event enhancement	Drones	Drones
Context aware services	Autonomous driving	Autonomous driving
Overlay identification	Real time monitoring	AI
Overlay navigation	Real time detection	
Maintenance applications	Productivity tracking	
Enterprise pinning	Experience AI	
Sponsored service		
RAT selection		

Who Should Own The Edge



to the source, rather than in the Cloud. And should the uplink go down, the IoT gateway can continue to function so you're not stuck in the dark when your connected light switch and your connected light bulb lose their connections to the Cloud and each other.

Industrial IoT (IIoT) takes these concepts to an industrial scale. Many existing industrial devices expose telemetry that is ignored or lost over time (as much as 97% in fact). By collecting and rationalizing this data at the Edge, industrial IoT gateways can improve the efficiency and efficacy of industrial automation. IoT and IIoT also serve as a catch-all for several loosely correlated Edge Computing use cases, including connected cars, transportation, and energy. These use cases share the IoT functions of telemetry analytics and localized actuation at the Edge.

Network Functions: Routers, switches, and firewalls usually conjure up images of big metal boxes. While network functions have traditionally run on purpose-built appliances, Network Function Virtualization (NFV) has taken these closed systems and turned them into software that can run in a Virtual Machine or even a Docker Container. Unlike the other use cases on this page, network functions **must** run at the Edge. Packet forwarding and security functions just can't be outsourced to the Cloud while maintaining real-time performance.

Fortunately, the combination of NFV and Edge Computing make it easier than ever to manage the lifecycle and configuration of these newly virtualized network functions. Network service providers are already considering the shift to NFV as an opportunity for rolling out general-purpose Edge Computing platforms like Multi-access / Mobile Edge Computing.

Gaming: With Edge Computing, services like Cloud Gaming become viable. Rather than investing in a new PC, PlayStation, or Xbox every year or two, gamers could subscribe to an Edge-hosted gaming service. The Edge hardware is kept up to date and users connect to it remotely. While previous attempts at Cloud Gaming (like OnLive) failed due to issues with real-time latency, the emergence of managed Edge platforms may reignite this model.

Content Delivery: The original Edge Computing use case. By caching content — whether it's a web page, a video, or a piece of music — at the Edge, end-users enjoy a better experience. Edge caching provides a massive improvement over traditional web servers, providing latency on the order of single milliseconds.

This was news 20 years ago; what does it have to do with Edge Computing today? The Content Delivery Network (CDN) market has been dominated by a handful of players, like Akamai and Limelight. They have built out sprawling global cache networks. What if the Edge looked more like the Cloud? Rather than rely on a CDN provider, any content provider, like HBO or Netflix, could spin up their own custom micro-cache at the Edge of the network with greater flexibility and customization than if they had used a general-purpose CDN provider.

Machine Learning for Voice and Video Recognition: Voice recognition through Alexa, Google Assistant, and Siri are now mainstream. And, in a world of increasingly connected devices, the number of voice and video capable devices is only going up. Hauling all this voice and video content back to the cloud is costly. Users expect low-latency responsiveness without taking hit to their data plans.

Edge Computing enables the execution of Machine Learning inference models, such as those used for voice and video analysis, to run closer than ever to end-users and their devices. By converting speech to text at the Edge, a megabyte (MB) voice recording can be converted to just a few bytes of text.

Virtual Desktop: While Cloud software is the future of Enterprise applications, many still rely on virtual desktop to get things done. Whether these virtual desktop environments are hosted in a corporate data center or running in the public cloud, they can be slow for remote workers and can even halt productivity under the wrong network conditions.

Edge Computing has the potential to free virtual desktop environments from the data center and enable highly controlled yet localized access for remote workers through Virtual Machine migration. Imagine working on your virtual desktop environment in the office and then having that virtual desktop instance follow you home to your closest Edge site. Like the Gaming use case, lower latency makes all the difference.

Video Conferencing: Whether you attend meetings every day or once a quarter, most of us have experienced conferencing software and its occasional shortcomings. Voice delays, bad video quality, and frozen screen-shares are not unusual. These problems can be the result of a slow link back to the cloud, where multiple voice and video streams are multiplexed together.

By placing the server side of voice and video conference software closer to the participants, these kinds of quality problems can be reduced significantly. A globally distributed Edge fabric of voice and video servers enables a much more resilient and responsive user experience for conference participants.

Cloud Storage: Everyone has data in the Cloud, but sometimes it can be hard to access. It can be slow or even unavailable. Whether you're using a service like Dropbox and Google Drive or just trying to access a network file system (NFS) over the Internet, remote storage doesn't always work the way it should.

Edge Computing can help fix this problem. By placing a storage gateway at the Edge, the gateway can act as a read/write cache. If the Cloud storage service is unavailable, it can offer up the working set of files that are cached locally. On a slow link, it can give the appearance of a fast local storage array, despite an exceptionally slow uplink.

Augmented Reality: Finally, the flashy Edge use case. In reality, augmented reality (AR) is an unlikely driver of Edge Computing adoption. While it is possible to offload computation from lightweight AR headsets, the power tradeoffs and current dependencies on specialized hardware make it difficult to justify AR as a use case for a general-purpose Edge Computing platform today. That said, AR certainly fits the mold of the prototypical Edge use case: bandwidth-hungry and latency-sensitive. As this technology advances, augmented reality will no doubt benefit from the augmented computation of the Edge.

VI. CHALLENGES

Figure 4 gives the barriers which the operators consider for adoption of Edge computing in 5G era. More than half (54%) of the industry professionals believe the lack of knowledge about enterprise edge computing will pose the highest barrier to its broad adoption. Several other options were each chosen by more than 30% of respondents, including the lack of any clear direction for edge computing development, fragmentation of technologies, security concerns and the financial costs. Looking at the question from more of a business perspective, the financial viability of edge computing for the operators will be the biggest impediment to its success. The CSPs are spending to win the race and win the hearts of customers. The challenge will be achieving a compelling ROI (return on investment) while previous and current charging scenarios have relied on price per gigabyte or bandwidth, or both.

Privacy Concerns: The idea of using facial recognition through video surveillance has started to create some privacy concerns in recent months, as there is little awareness in the general public who have not consented to being monitored.

With processing power stored on the edge of the network the data can be processed, insight captured, before being deleted. Useless information can be sifted out on the edge, with only relevant data or the insight sent back to the core. By empowering the edge, privacy concerns are negated as personal information is not actually being stored, simply the insight which would not be considered sensitive.

Impediments to the broad adoption of edge computing in 5G: The location, distribution, cost and ownership of edge computing resources are all issues that need to be resolved, but the key factor will be whether or not providers can build a coherent and sustainable business case around edge computing deployments. This will require a deeper understanding of how the technology can work in a real-world commercial setting. If this is not possible on a broad basis, it is likely that edge computing, like some earlier technologies, will be limited to specific use cases rather than as a ubiquitous resource.

If telcos want to be able to improve customer experience, data needs to be collected and analysed. This might sound like a very obvious statement to make, but the growing privacy movement across the world, and the potential of new regulatory restraints, might make this more difficult.

Some might say artificial intelligence and data analytics are solutions looking for a problem, but in this instance, there is a very real issue to address. Improving customer experience through analytics will only be successful if implemented quickly, some might suggest in real-time, therefore the models used to improve performance should be hosted on the edge.

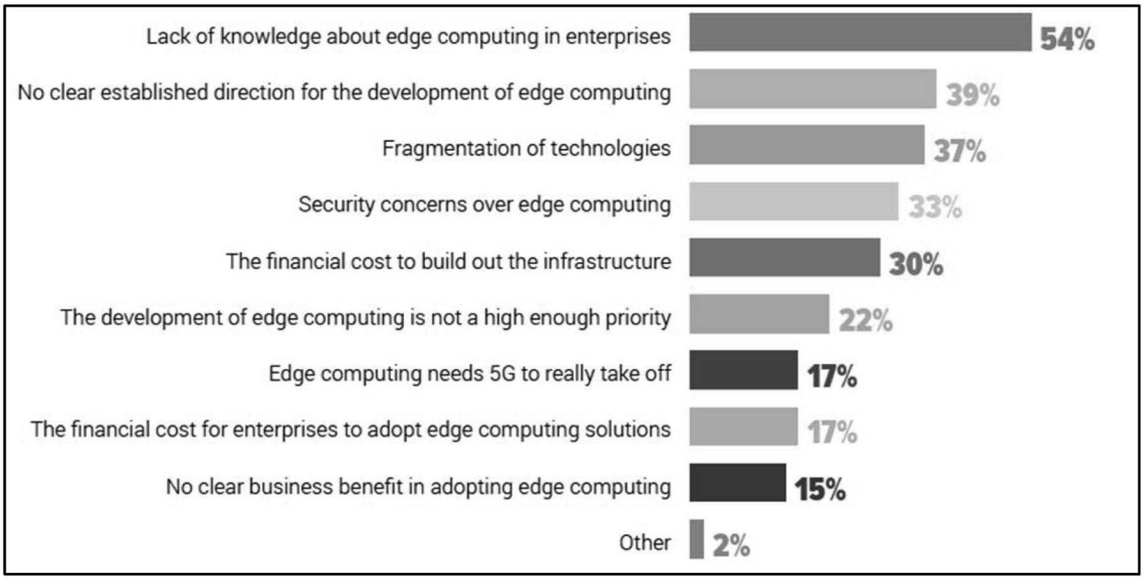


Figure 4. Barriers for adoption of Edge computing.

This is an example of where the latency business model can directly impact operations.

It also addresses another few issues, firstly, the cost of sending data back to a central data centre. It may be that telcos cannot afford to send all customer data back to be analysed , it is simply an unreasonable quantity, therefore the more insight which can be actioned on the edge, with only the genuinely important insight being sent back to train models, the more palatable customer experience management becomes.

Secondly, the privacy issue is partly addressed. The more which is actioned on the edge, as close to the customer as possible, the lesser the concerns of the privacy advocates. Yes, data is still being collected, analysed and (potentially) actioned upon, but as soon as the insight is realised, the sooner it can be deleted. There are still sceptics when it comes to the edge, the latency business case, artificial intelligence and data analytics, but slowly more cases are starting to emerge to add credibility. 75% of enterprise data expected to be processed on the edge by 2020-end.

Edge computing does not have to wait for 5G to happen. One of the most broadly adopted edge computing cases is private LTE for campus, for enterprise, etc. Yet this is an area that has become controversial to telecom operators. It is believed that private LTE, and in the future private 5G, may prove a new business opportunity for mobile operators if it is a network slice bought from the generic mobile network.

It would be a challenger if it was operated independently—for example the discussion in Germany that 5G frequencies could be awarded locally to private networks. However what worries the operator the most, is the webscale companies (AWS, Google, Microsoft) getting frequencies and offering services on the edge. This is already happening. Amazon has filed to FCC to expand its test on 3.5GHz band and, as Light Reading reports, this could be related to Amazon's plan to "offer cloud-native, private mobile networks in the CBRS band (Citizens Broadband Radio Service) to developers, telecom operators, public sector operators, enterprises and others."

A key issue for edge computing concerns interoperability. As an extension of the public cloud, it needs to be usable by all stakeholders. One way to ensure this is standardisation, something the telecoms world is very familiar with. Standardisation typically takes a long time, however, and the panel warned that operators are likely to lose their advantages in this space if they allow themselves to be bogged down by it.

There are also cultural dynamics involved. The IT world typically moves faster and is less risk-averse than the networking world. While telcos are used to significant infrastructure capex, this is typically in areas where there is proven demand and ROI. Heavy investment in edge computing will require more of a 'build it and they will come' strategic philosophy.

VII. CONCLUSION

Edge computing on mobile networks is by no means a new topic. One of the early demonstrations took place at Mobile World Congress-2013, when Nokia (then Nokia Siemens Networks, NSN) introduced its base station-based technology Radio Applications Cloud Server, which supported real-time data processing of the radio and network information from the base station.

In 2014, the standardisation body ETSI established a working group, the Industry Specification Group (ISG) on MEC, which initially referred to Mobile Edge Computing but was expanded to cover Multi-access Edge Computing in March 2017.

What may seem disappointing to the edge computing enthusiasts is that no serious commercial rollout of MEC on a big scale has yet happened, though there have been plenty of proofs of concept (PoCs). As a matter of fact, part of the ETSI MEC group's mission is to develop specifications for and provide support to PoC cases

There have been enormous efforts from academia and industry to realize MEC as the key enabler for applications and services (*e.g*., V2X, Tactile Internet, AR/VR, and big data) in the 5G and beyond network. MEC provides a great number of opportunities and potentials; however, some challenges exist and need to be further studied and tackled, *e.g*., distributed resource management, reliability and mobility, network integration and application portability, the coexistence of heterogeneous (*i.e*., H2H and MEC) traffic, data privacy, and security.

Sources:
1) Cisco white paper Feb 2019
2) Edge computing and 5G, Infosys Knowledge Institute
3) Telecom.com Intelligence- Aug 2019 , Edge computing -Will 5G make it reality
4) IEEE paper :arXiv:1906.08452v2 [cs.NI] 2 Jan 2020
5) Edge Computing Wikipedia
6) Slide share- Mobile Edge Computing.

Manjit Singh, FIETE was the President of Himachal Futuristic Communications Ltd. (HFCL), a pioneering Indian telecommunications company involved in the design, manufacturing, and deployment of wireless, optical, and PSTN equipment, as well as providing turnkey IT solutions to customers all over India. He was heading HFCL's technical operations and business development. Directed successful implementation of CDMA, GSM and WLL equipment including PDSN, billing, customer care, and SMS/VMS systems.

Prior to joining HFCL, worked for Compton Greaves Limited as Vice President Telecom Services.

Served Indian Telecom Service for over 27 years, in key positions like Director responsible for Maintenance and Development in the Department of Telecommunications, Govt. of India. Was General Manager in Jaipur, for the upkeep and development in Rajasthan area. Also served MTNL, Delhi as General Manger. He has expertise in planning, installation, commissioning, and monitoring of several large telecom networks.

Presently, he is running a small software export company for the development of software for different applications.

He is a prolific writer on Telecommunications subjects which have been published in various Technical magazines. Mr. Singh has served on the technical groups of two world telecom forums: the ITU (International Telecommunications Union) and the APT (Asia Pacific Telecom). Holds a B.Sc. (gold medalist) degree in Electrical Engineering from the University of Punjab, India.