Infosys®
Navigate your next

WHITE PAPER

EDGE COMPUTING AND 5G

5G promises ultralow latency in transmitting large volumes of data. But bottlenecks may occur if cloud computing cannot respond fast enough. Processing power at the edge of a network will be an essential part in 5G delivering on its potential, and may provide a new revenue stream for telecom providers.

Infosys® | Knowledge Institute

By 2025, there will be more than 42 billion active IoT devices[1] creating 90 zettabytes[2] (1 zettabyte = 1 trillion gigabytes) of data. An average person will have more than 5,000 digital interactions per day.[2] Increasingly, these device and human interactions will be in real time, through streaming data, rather than intermittent interactions, as is typically the case today.

In theory, this growth will all be enabled seamlessly by the new 5G standard, which should reduce latency by 50 times (from 50 milliseconds to 1 millisecond) and increase bandwidth between four times, on average, to about 100 Mbps to 10,000 Mbps at its peak.[3]

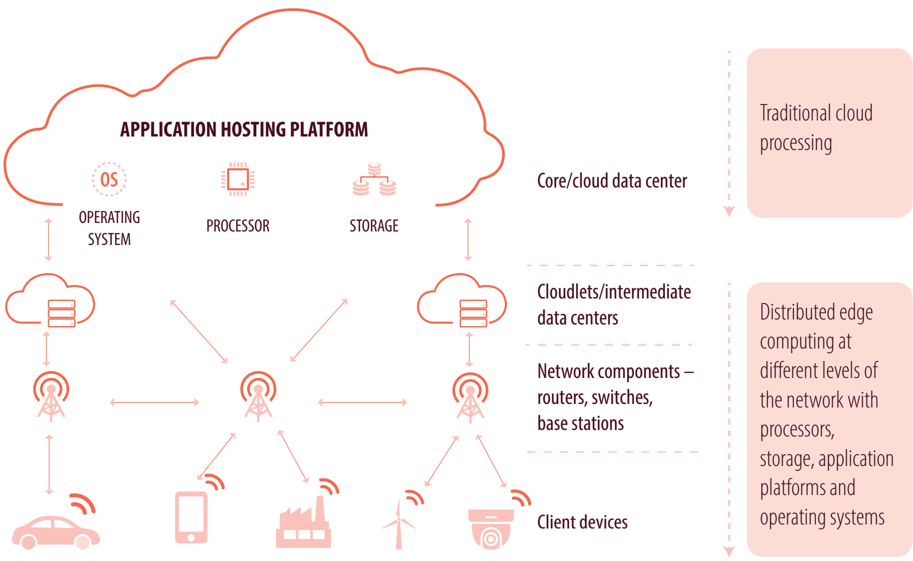
But as more rich media and time-critical applications are deployed to benefit from this wireless network, the concern is that processing bottlenecks will start emerging back at the data centers or in network nodes.

Consider facial recognition. For example, if the New York Police Department wants to use facial recognition for threat detection in Times Square, it would need an application that continuously captures images/videos of thousands of moving people and objects, analyzes the images and compares them with existing data to identify suspicious elements, and alerts the police. To be effective, this complex process would have to be done in a fraction of a second. This might not be possible if data has to travel through a choked physical network or compete for resources at the data center.

The issue becomes more of a challenge if, for example, during a potential terrorist attack or another emergency, the police want to quickly shift their focus to a different part of the city. This would suddenly move the pressure to another segment of the network, which may not be ready for it.

It's not just emergencies either. As autonomous cars become more prevalent, more passengers will expect to use their travel time productively while on the move. This most likely will involve two-way rich media connectivity, either for video conferencing or online collaboration. And as IoT devices proliferate and begin supporting "smart city" management, the reliance on large volumes of data to manage daily civil tasks will become much more critical.

Figure 1: Distributing processing throughout the network


APPLICATION HOSTING PLATFORM
OS
OPERATING SYSTEM
PROCESSOR
STORAGE
Core/cloud data center
Cloudlets/intermediate data centers
Network components – routers, switches, base stations
Client devices
Traditional cloud processing
Distributed edge computing at different levels of the network with processors, storage, application platforms and operating systems

Source: Infosys

Edge computing

The solution, at least in part, lies in pushing more processing power away from the core, and decentralizing throughout the network, closer to the edge devices (Figure 1). This new evolution of "edge computing" will be crucial in helping 5G deliver its promise of ultralow latency.

Historically, computing power has swung back and forth between the edge and the core, depending on the technology architecture of the time. Applications that benefit from a lower-latency and higher-bandwidth 5G wireless network will start the pendulum swinging toward a more powerful edge.

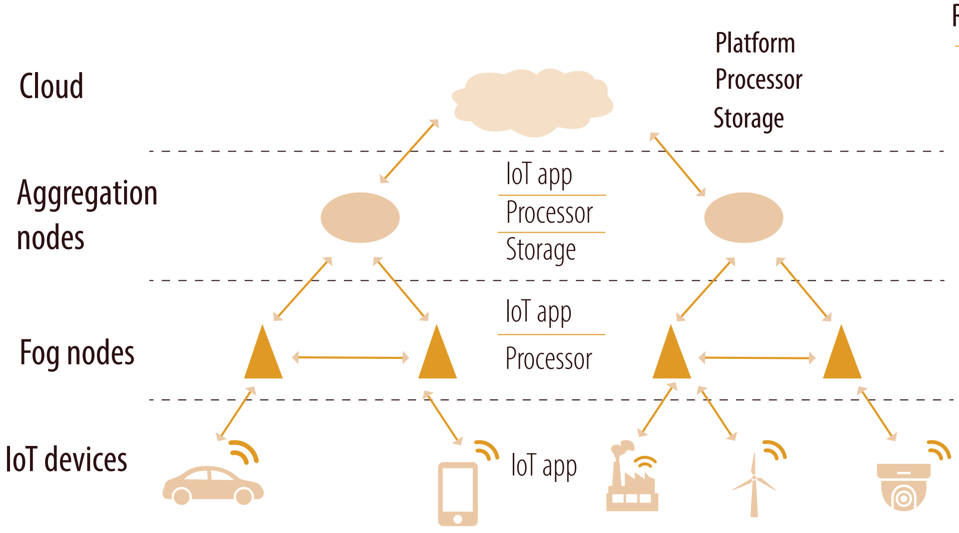
Many standards and flavors of edge computing are evolving, based on who is developing the technology and what application it is being used for. Some of those popular architectures being adopted today are multi-access edge computing, fog computing and cloudlets.

Multi-access edge computing

Multi-access edge computing, or MEC, is an edge computing architecture developed by European Telecommunications Standards Institute, ETSI, which brings a cloud-enabled service environment closer to the edge of the network. The idea is to bring computing close to the user, but to not have it residing on the user device.

This model creates three regions: the client or 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.[4] The MEC platform is a secure environment where services can be created and delivered. Low-latency

Figure 2: The different layers within a fog computing architecture



Response time	Storage time	Area covered
Minutes, hours, days	Months, years	Global
Seconds to minutes	Hours, days, weeks	Cities or wider area
Milliseconds	Transient storage	Local, blocks

Sources: Infosys, Cisco

applications are offered or serviced by the edge server or MEC host while complex requests are routed to the remote server. It is interesting to note that, in MEC, developers need to clearly specify which application elements will run on the edge cloud or MEC host and which ones should be routed to the cloud.

Fog computing

Fog computing is an edge computing standard defined by Cisco. Fog computing too, like MEC, aims to bring the cloud close to devices, with a focus on IoT devices. Fog nodes are devices with storage, network and computing capabilities, and can be deployed anywhere in the network as shown in figure 2.[5] Fog nodes nearest to the IoT device collect data from the source. Then, based on the time sensitivity and complexity of data, the fog IoT application directs the data to the right processing station. The most time-sensitive data is processed at the nearest free fog node. The decisions, which can take a couple of seconds, are directed to the aggregation nodes placed in the network, and the decisions that can wait are sent to data centers for storage and processing. The fog nodes also act as intermediate storage locations. The data collected is stored there for a few hours and then transmitted to the cloud for long-term storage.[5]

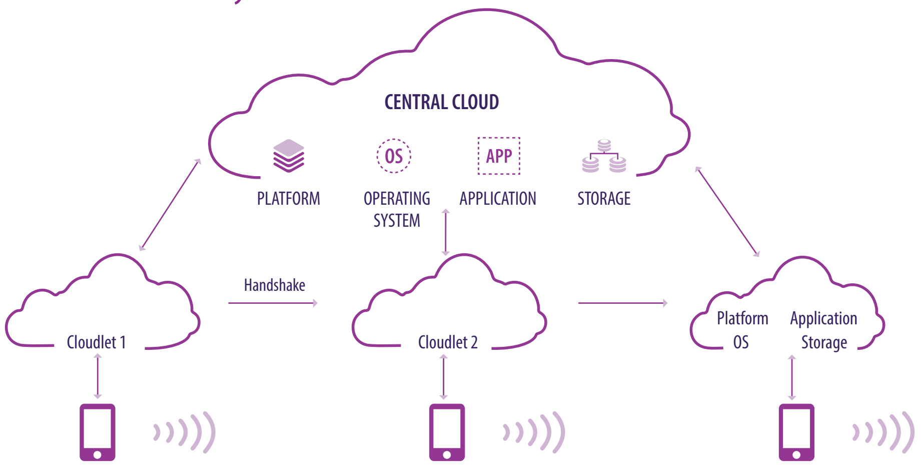
Cloudlets

Cloudlet is a term coined by Mahadev Satyanarayanan 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 known as a "data center 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 a connected device has to transmit data, it offloads the data onto the nearest cloudlet (Figure 3). The cloudlet stores and processes this data. It might decide to pass on the data to the main data center if it sees doing so as being necessary. When the connected device moves, the cloudlet detects the next cloudlet closest to the device and transfers all the data using defined protocols, essentially orchestrating a cellular handshake.[6] This ensures seamless low-latency processing as the device travels.

All of these types of edge computing will likely interact and work together.

Figure 3: Cloudlets create mini clouds to manage mobile devices effectively



Source: Infosys

The industry is at a stage where standards are still being set, and it won't be until 5G networks are much more established that we see the types of applications that are in demand. This will determine the specific architectures and approaches that are taken.

As edge computing evolves, however, it's worth bearing in mind the many challenges and complexities that need to be overcome to scale this concept in the 5G era.

Device capabilities

The source/end devices, substations and routers all need to be equipped with the hardware and software capabilities required to support edge computing. The hardware providers need to decide what kind of computing and storage capacities these devices will have. If the device has a lot of processing power, it would be very fast. But it would be heavier, expensive and require more battery power. Therefore, the hardware providers would need to find a balance between processing power and the size of the device.

Switching nodes

For a seamless user experience, the software needs to be designed to intelligently route the data to different processing nodes. For this, it will need to consider many parameters, such as the volume of data, current traffic in the network, nearest edge node that is free, the best path to route the data, which part of the data is to be processed by an edge node and which one is to be routed further, how critical the decision is, how time-sensitive the decision is, and many more. The way software is architected and how it interacts with the operating system of each device will be incredibly complex, and will require a virtualized control layer to support developers (see below).

Moving devices

The end points and sources of data, meaning the IoT devices, will not always be stationary. As these devices move, the system should automatically be able to connect them to the nearest edge node based on the location. If this is not done instantly, the data would travel to the previously connected edge node, which might now be far away. Then, the advantage of edge computing will be lost. This also has to be done perfectly to ensure a seamless user experience.

Security risks

With increasing points of storage and processing, security vulnerabilities are bound to increase. The network and service providers will have to enhance security measures to address the possibilities of malware attacks. As the network becomes more connected, malware or security breaches can be contagious.

Intelligent edge

Edge nodes need to be equipped with intelligent decision-making capabilities to ensure quick response time. For example, in case of facial recognition application, the edge nodes should be able to analyze the images to identify any suspicious activity. The edge node in an autonomous car should be able to apply machine learning algorithms to predict the behavior of other vehicles on the road dynamically and move accordingly.

Distributed control

There are two management challenges that this type of edge network will create: managing the interconnections across the multiple infrastructure nodes, and enabling a virtualized layer for software developers to easily create applications that can benefit from a distrusted edge network. On the infrastructure side, different edge computing platforms — MEC hosts, fog nodes and cloudlets — will need to communicate flawlessly with each other and also with the core data centers. The data has to be routed to the right nodes in the shortest time, as low latency is the goal here.

Also, as the user moves, the edge node should do a handshake with the next-nearest node/cloudlet. All of this requires a sophisticated controlling layer or control plane that can handle this routing. Software-defined network architecture enables cloud providers to centrally control the network switching to find the most effective route; this will have to be extended to accommodate edge computing. The control plane should create a virtual system to classify data, balance the load and distribute the traffic between the edge computing nodes. It should also be able to identify the best possible node to which all the data can be transferred as the device moves.

Costs and commercial model

Perhaps the biggest challenge is overcoming the fact that to build an edge network in 5G would incur significant costs. Certainly, the ever-declining price of chips and storage will help. But the fact remains that, in the early days, the business case for upgrading to an edge network may be difficult to prove, until it's clear what applications can be monetized on the network. It does leave 5G edge computing in somewhat of a "chicken and egg" situation, whereby the value can't be proved without the network, and the network can't be built without proven value. That said, progress is occurring, and we expect that it will only take the emergence of one "killer app" to quickly incentivize the industry to invest.

Once we reach that stage, there are many ways in which telecom companies can monetize edge computing. They can develop their own edge networks to facilitate zero-latency use cases for their subscribers — both consumer and corporate. They can also develop a virtual edge platform where software developers can build and sell new low-latency, highly critical applications to network subscribers.

Telecom companies can define a management plane to handle the data flow and manage service level agreements. This could orchestrate the entire flow of data between the devices, edge nodes and core, and generate commercial value by being an intelligent layer between content providers and consumers.

Edge — a lifeline for telcos?

Cloud providers such as Amazon, Google and Microsoft have been the power players in the cloud computing era, both in terms of owning the infrastructure as well as playing a part in the content and value that is generated over it. However, with the arrival of 5G edge computing, the power balance could soon tilt away from them.

Today telecom carriers are the closest ones to the users and IoT devices, and will also control the 5G infrastructure. Edge computing and 5G provide a unique opportunity for telecom providers to move up the value chain, after having lately been characterized as "dumb pipe" service providers. This point has not been missed by the industry. Telecom companies have already started exploring opportunities edge computing on 4G network. Telstra is looking at converting its base stations and towers into computing nodes.

The cloud leaders see this threat and have started investing in the ecosystem. They are looking for partnerships with telecom companies. The hyperscale cloud providers like AWS and Microsoft are already looking to develop edge computing solutions for 5G. They intend to partner with telecom companies for last mile delivery. Microsoft recently announced partnership with Telefonica to explore intelligent applications and services in the network.[7]

It's clear that 5G and the potential creation of a powerful edge network will start to reorder the industry. Many new opportunities will emerge for all involved, from infrastructure manufacturers to content providers and software developers — and telecom providers need to embrace this opportunity too. To ensure that they do not miss the opportunity, telecom carriers need to up their game technically and commercially to truly position themselves as the smart "middleman" in the era of connected computing and 5G.

References

1. "IoT Signals report," Microsoft, July 2019, https://blogs.microsoft.com/blog/2019/07/30/iot-signals-report-iots-promise-will-be-unlocked-by-addressing-skills-shortage-complexity-and-security/
2. "IDC: Expect 175 zetabytes of data worldwide by 2025," Network World, December 2018, https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html
3. "5G technology needs edge computing architecture"; Cisco; https://www.cisco.com/c/en/us/solutions/enterprise-networks/edge-computing-architecture-5g.html
4. "MEC software development"; ETSI; February 2019; https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp20ed2_MEC_Software-Development.pdf
5. "Fog Computing and the Internet of Things"; Cisco; 2015; https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf
6. "Cloudlets: at the Leading Edge of Mobile-Cloud Convergence"; CMU; 2014; https://www.cs.cmu.edu/~satya/docdir/satya-mobicase2014.pdf
7. "Telefónica and Microsoft establish strategic partnership"; Microsoft; February 2019; https://news.microsoft.com/2019/02/25/telefonica-and-microsoft-establish-strategic-partnership-to-design-the-telco-of-the-future/

Authors

Gnanapriya C

Associate Vice President

gnanapriyac@infosys.com

Samad Masood

Infosys Knowledge Institute

samad.masood@infosys.com

Rachana Hasyagar

Infosys Knowledge Institute

rachana.hasyagar@infosys.com

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI

For more information, contact askus@infosys.com

Infosys®
Navigate your next