Edge Computing and Mobility

Rafat Khan

March 13, 2022

Abstract

This essay concludes that there is an increasing demand for Edge Computing and why IT worlds adopt it easily and why it is becoming more popular over time? Also, why we need Edge Computing for modern IoT and mobile services like 5G. As opposed to improving latency, edge computing reduces unplanned downtime of IoT devices, so that they can run, process, and compute the data within the allocated time. From a business perspective, edge computing offers fast processing and fast computing at a lower maintenance cost.

1 Introduction

Understanding edge computing, cloudlets, and fog computing requires understanding what cloud computing is and why it is needed. The cloud provides on-demand computing services over the internet. Several datacenters are linked through optical networks to form a data center network (DCN) that is at the heart of cloud computing. The datacenters are responsible for storing and computing data. Due to its scalability, flexibility, and affordability, cloud computing has gained popularity among organizations. But with the internet of things (IOT) and technologies like 5G expanding, it has a new set of demands, which are:

1. Realtime processing of sensor data, the communication latency between end devices and DCN might prove to be bottleneck.
2. As most of IOT devices have agile nature, mobility support is necessary.

In edge computing, the computing and storage capacity of a large number of mobile, wearable devices or sensors is leveraged to overcome these shortfalls. It serves as a bridge between end devices and the cloud.

This reduces the latency in resolving requests because these "Edge devices" handle requests locally rather than sending them to the cloud, which allows real-time handling of requests. Furthermore, the abundance of edge devices facilitates mobility [1]. Essentially, Edge computing reduces the need for data to be sent to a remote cloud or server by bringing the processing to the data store.

It can be classified as Mobile Edge Computing (MEC), Fog Computing (FC) and Cloudlet Computing (CC). An article co-authored by M. Satyanarayana, Paramvir Bahl, Rámon Cáceres, and Nigel Davies laid the conceptual foundation for edge computing. They presented a two-level architecture, in which the first level represents present day cloud infrastructure, and the second level is dispersed elements called cloudlets where the state is cached from the first level clouds. Cloudlets can be arranged in a multilevel hierarchy with this cloudlet concept. In 2012, Flavio Bonomi and his colleagues introduced the term fog computing to refer to dispersed cloud infrastructure, where data is cached from a main data server [1]. In addition, we can say that one of the standards for implementing edge computing is called fog computing. It helps in computing, storing, and networking between devices and centralised data centers. Fog extends the cloud infrastructure close to the IoT devices that produce and process data. Fog nodes can be deployed anywhere in the network, with the data-producing devices. We can say any device that can compute, store, and have network connectivity can be called a fog node.

The Cloudlet is a small part of the cloud that is placed close to end nodes and has data cached from main cloud. Further, it can be defined as a resource-rich system placed close to mobile devices and connected to the Internet. Proximity of cloudlet matters because it helps us achieving goals like [4]:

- Low end to end latency
- Scalability: As only exact information is transmitted after analysis.
- Privacy- As privacy policy can be enforced prior to release of data to cloud.
- Cloud failure protection-In case cloud services become unavailable, a fallback service of nearby cloudlet can mask the failure.

2 Mobile Edge Computing

In telecommunications, edge computing, also called Mobile Edge Computing, Multi-Access Edge Computing, or Mobile Edge Computing, provides execution resources (compute and storage) for applications utilizing local networks close to the end users, typically within or at the edge of operator networks. In addition to enterprise premises, edge computing can also be installed inside buildings such as factories, homes, and vehicles, including trains, planes, and private cars. A communication service provider or other type of service provider can manage and host edge infrastructure. For use cases that require multiple applications to be deployed at different places, a distributed cloud is helpful, which acts as an execution environment for applications across multiple sites with connectivity managed as one solution.

Low latency, high bandwidth, device processing and data offloading, as well as trusted computing and storage are some of the main benefits edge solutions provide. Edge Computing has transformed the infrastructure of Data Centers as well as the way Cloud Computing is utilized. It offers increased elasticity, security, and reliability. By putting data closer to the users who connect to it, information can be shared quickly, securely, and without latency. It provides IT and cloud computing capabilities within the Radio Access Network (RAN) in close proximity to mobile subscribers.

Typically, Mobile-edge Computing is characterized by:

- On-Premises: As an Edge node, it can run isolated from the rest of the network, but still have access to local resources. This is especially important in Machine-to-Machine scenarios, for example when dealing with security or safety systems that need high levels of resilience.
- Proximity: As Edge Computing captures information near the source, it is especially useful for analytics and big data. Edge computing may also have direct access to devices, which can be leveraged by business-specific applications.
- Lower latency: Edge services run close to end users, reducing latency and allowing them to react faster, to improve user experience, or to minimize congestion in other parts of the network.
- Location awareness: A Local Service can determine the location of each connected device when a Network Edge is part of a wireless network, be it Wi-Fi or cellular. This opens up a vast array of business-oriented use cases, including Location Based Services, Analytics, and many more.
- Network context information: Applications and services can use real-time network data (such as radio conditions, network statistics, etc.) to deliver context-related services that differentiate the mobile broadband experience and can be monetized. These real-time network data can be used to develop new applications, which can connect mobile subscribers with local points-of-interest, businesses, and events.

A RAN edge offers application developers and content providers a service environment with ultra-low latency and high bandwidth, as well as direct access to real-time radio network information such as subscriber location, cell load, etc. Services and applications can use these to offer context-related services; these services can differentiate the mobile broadband experience.

Content, services, and applications can be accelerated by Mobile Edge Computing, improving responsiveness at the edge. Mobile subscribers' experience can be improved with efficient network and service operations, based on insights into radio and network conditions. The radio network edge can be opened up to third-party partners, allowing operators to rapidly deploy advanced applications and services to mobile subscribers, enterprises, and other vertical segments. By leveraging proximity, context, agility and speed, mobile operators, service and content providers, and Over the Top (OTT) players and Independent Software Vendors (ISVs) can play complementary and profitable roles within within their respective business models and allowing them to monetize the mobile broadband experience.

With this environment, we can create a new value chain and an energized ecosystem encompassing app developers, content providers, OTT players, network equipment vendors, and mobile operators. By focusing on innovation and business value, this value chain will benefit all players. To facilitate global market growth, it is intended to create favorable market conditions that will help all players within the value chain to succeed. To make this possible, a standardized, open environment must be created to facilitate the seamless integration of such applications across multivendor mobile edge computing platforms. ETSI is proposing to set up a new Industry Specification Group (ISG) to develop industry specifications for Mobile-edge Computing (MEC). The ISG MEC will also work towards enabling and accelerating the development of edge applications across the industry, increasing the market scale and improving market economics.

3 Edge Computing and CDN

As the internet has grown, businesses all over the world have become increasingly dependent on the speed, availability, and integrity of the network. Before Content Delivery Networks (CDN) were implemented, each request was sent to the main web server to be processed, which increased latency. Also, during times of congestion, multiple requests could be made to a particular website, causing it to become unavailable. In order to overcome this problem, Akamai introduced the CDN in the 1990s to improve web performance [4] [3]. The CDN network is a network that caches data from the main server on CDN servers that are placed close to the end users. CDNs have helped companies to achieve a better Quality of Service (QoS) and Quality of Experience (QoE) for their customers. Having CDN nodes close to users who have preferred or cached data can greatly reduce latency and save a lot of bandwidth [4]. Additionally, these nodes can help advertisers reach the right audiences with location-based advertising. CDNs have increased in size and number over the years; over a third of internet traffic is served by CDNs, and it is expected to reach over half in the near future [5]. Since edge computing uses nodes close to the end user, it has roots in CDN, as mentioned earlier when Akamai introduced CDN to speed up web performance in the late 1990's. As with edge computing, edgelets extend the same concept by leveraging the cloud infrastructure. For both, the proximity to the user is crucial since edgelets are only limited to caching web, while cloudlets run arbitrary code just like cloud computing.

4 Edge Computing and IoT

Edge Computing enables data generated by Internet of Things (IoT) devices to be processed closer to where it is created rather than being sent long distances to the cloud. Edge computing is entirely referred to in IoT use cases, where edge devices collect data sometimes in very large amounts and send it to clouds or data centers for further processing. The majority of the tasks are handled by IoT devices that transfer data to the local device for processing, storing, and connecting with other networks. It enables Internet of Things (IoT) data to be analyzed at the edge of the network before it is sent to the Data Centers and Cloud. The concept of Edge Computing is very reliable and the safest way to share data on cloud since it works near the user and offers more security and availability of IoT. Edge computing is increasing in popularity due to its ability to deliver tangible value. This helps to reduce connectivity loss that is common with cloud computing. In the world of IT, edge computing is very popular because it offers a lot of benefits to its users. By storing data in IOT devices, Edge computing also enhances security and privacy. Processors at the edge also reduce latency and improve the responsiveness and robustness of connected applications.

Data available on the cloud can be brought close to the source by Edge Computing, thereby reducing the latency and bandwidth requirements. With IoT, EDGE COMPUTING means running fewer processes in the cloud and more near the user. By doing so, it also reduces long-distance communication between the client and user. Edge computing stands out from other forms of computing due to its security features. Because the data would not travel across the network and be processed in IoT devices, it would be more powerful and less vulnerable as compared to the cloud because it stayed closer to where it was created.

By keeping data within IoT devices, edge computing can improve security and privacy. In applications that use computer vision or machine learning, eliminating device-to-cloud data round trips is critical - such as an enterprise identity verification system or a drone that tracks and films its owner or an object. In addition, on-device machine learning can enhance natural language interfaces, allowing smart speakers to respond more quickly to voice commands, turn on/off lights, or adjust the thermostat, regardless of internet connectivity. Furthermore, edge computing makes these systems future-proof since it enables over-the-air software updates as well as the ability to run local commands.

The opportunity for system architects with edge computing is to harness the benefits of distributed computing from end to end - harnessing the capabilities of field devices, gateways, and clouds.

5 How does Edge Computing affect industry?

There are many applications of edge computing in the IT industry. According to IT experts, billions of connected things will generate massive amounts of data that don't just require security, but also require reliability so that data is available at any time. By leveraging Big Data, advanced analytics, and machine learning, we can reduce unplanned downtime as well as improve performance.

In the past, a mobile network's edge was a place where only specialist processing was done. It contained specialized computing that was designed from scratch to perform a specific function in the overall architecture and could not be repurposed. It was also necessary to use specialized protocols to connect from the edge to the core of the network. The entire configuration was optimized before smartphones, when voice quality was the key factor in network design and before the days when IP was the norm for network communications.

The growth of IP from the internet to enterprise networks, as well as the widespread adoption of LTE, has enabled new applications to emerge that have transformed the way telecommunications networks are designed. The single vendor radio network solutions are evolving into modular, open solutions that can be integrated into a changing ecosystem of flexible components [6].

With Mobile Edge Computing, there are new ecosystems and value chains, and all players can collaborate and develop new business models they can each benefit from. MNOs are able to deploy new services rapidly for both consumer and enterprise segments, which can differentiate their portfolios. In addition to improving the MNO's bottom line, innovative services that are delivered closer to the end user can improve the QoE of the end user. Apps that are aware of the local context in which they operate (RAN conditions, locality, etc.) can open up entire new service categories and increase end user satisfaction. Applications installed on or near the base station not only benefit consumers and enterprises. MNOs can also increase their revenue by charging per resource used (storage, NW bandwidth, CPU, etc.) by each provider by reducing the amount of signaling offloaded to the core network. In order to serve this new ecosystem, software and application providers need to develop and bring to market innovative and ground-breaking services and applications that leverage radio network capabilities and conditions available at the base station. Anyone can apply for the application space: software and application providers, infrastructure vendors, and mobile network operators.

The use of open standards and APIs, as well as the use of familiar programming models, relevant tools and Software Development Kits are key pillars to encourage and expedite the development of new disruptive applications or the adaptation of existing services and applications to the new Mobile-edge Computing environment.

6 Edge computing: its future and scope

As the IoT ecosystem has grown rapidly in recent years, the amount of data generated from endpoint devices has increased accordingly. However, many organizations that have spent the past decade moving data from their own data centers to the cloud are realizing it does not make sense to send all the data generated by IoT devices to the cloud for processing.

One of the main concerns is how much IoT data will be created. According to IDC, a technology research and advisory firm, there will be 55.7 billion connected devices worldwide by 2025, of which 75% will be connected to IoT platforms. According to IDC, connected IoT devices will generate 73.1 ZB of data by 2025, up from 18.3 ZB in 2019. The majority of this data comes from security and video surveillance, but industrial IoT will also make up a substantial portion of this data.

Edge computing is driven by IoT and 5G, since they allow for an increased number of data sources and processing points to be interconnected, meaning an exponential increase in data volumes to be processed. Existing "sites to cloud" connections can quickly become overloaded with this type of data and require data processing to be performed closer to the source. Additionally, 5G has much lower latencies, which is a key component of some new applications, as well as making processing power more accessible closer to where it's consumed or generated.

Edge computing will absolutely be open in the future. Artificial intelligence and machine learning will be used to collide edge computing with the use of data to turn insights into actions that will benefit businesses and customers alike. Eventually, it will be viewed as any other place where applications can be placed seamlessly, consistently, and without compromise.

References

[1] K. Dolui and S. K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," 2017 Global Internet of Things Summit (GIoTS), 2017, pp. 1-6, doi: 10.1109/GIOTS.2017.8016213.

[2] https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

[3] A. Vakali and G. Pallis, "Content delivery networks: status and trends," in IEEE Internet Computing, vol. 7, no. 6, pp. 68-74, Nov.-Dec. 2003, doi: 10.1109/MIC.2003.1250586.

[4] S. Mehraghdam, M. Keller and H. Karl, "Specifying and placing chains of virtual network functions," 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet), 2014, pp. 7-13, doi: 10.1109/CloudNet.2014.6968961.

[5] B. Han, V. Gopalakrishnan, L. Ji and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," in IEEE Communications Magazine, vol. 53, no. 2, pp. 90-97, Feb. 2015, doi: 10.1109/MCOM.2015.7045396.

[6] M. Patel et al., "Mobile Edge Computing- Technical white paper," no. 1, p. 36, 2014.

[7] K. Dolui and S. K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," in 2017 Global Internet of Things Summit (GIoTS), Geneva, Switzerland, Jun. 2017, pp. 1– 6. doi: 10.1109/GIOTS.2017.8016213.

[8] Q.-V. Pham et al., "A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art," arXiv:1906.08452 [cs, math], Jun. 2019, Accessed: Sep. 01, 2019. [Online]. Available: http://arxiv.org/abs/1906.08452

[9] M. Satyanarayanan, G. Klas, M. Silva, and S. Mangiante, "The Seminal Role of Edge-Native Applications," in 2019 IEEE International Conference on Edge Computing (EDGE), Milan, Italy, Jul. 2019, pp. 33–40. doi: 10.1109/EDGE.2019.00022.