



Edge Computing – An Overview

Local and Decentralized Computing

Author: Edgar Germer, Information Technology Risk Control
Published: April 2022

Introduction

“Alexa, what’s the weather?” “Hey Siri, preheat the oven to 350 degrees.” A simple spoken command is all it takes for many people to obtain information or interact with their appliances today. Similarly, in the not-too-distant future, instead of asking for directions to a location we might simply tell our vehicles to take us there.

The commercialization and utilization of smart appliances, smart homes, the Internet of Things ([IoT](#)) and Industrial Internet of Things ([IIoT](#)), Augmented Reality (AR)/Virtual Reality (VR), and other smart technologies is exploding. The unprecedented scale and complexity of data that is being created by internet connected devices has outpaced network and infrastructure capabilities. [Gartner](#) estimates that by 2025, 75% of data will be processed outside the traditional, centralized cloud data center.¹ This staggering prediction begs several important questions:

- Where is this information coming from?
- Where is the data being processed?
- What impact will the increase in data demand and processing time have on the latency, speed, and bandwidth of the current information super-highway infrastructure?

This whitepaper provides an overview of the challenges resulting from the proliferation of a digitally interconnected society and the technology being implemented to address these challenges – specifically, the shift from a cloud-based computing to edge-based computing.

In the Beginning – clouds, latency, & bandwidth

When voice assistants came to market, they relied on back-end cloud service providers to respond to a user’s request such as “Alexa, what’s the weather?” With only a few major cloud service providers (Amazon, Google, Apple), the request would travel hundreds if not thousands of miles for processing resulting in a relatively long response time. This delay or latency increased over time with the growth of voice assistants and other devices that required cloud-based, backend processing. This increased the demand for bandwidth and congestion on the network. To address these issues of latency and bandwidth, providers are now turning to “edge computing”.²

What is Edge Computing?

In its basic form, Edge Computing is the placement of computational and storage servers closer to the devices generating the data and request for services; instead of having it performed at a central cloud processing center that is distant.³ Think of a traditional back-end cloud service as a hub and spoke model – like a bicycle wheel with the center being the mega, central cloud processing center and requests coming in

from devices at the spokes. In edge computing, the processing power is distributed out to the spokes so that it is closer to the devices.

Edge computing was developed to address issues of latency and bandwidth created by the exponential growth of smart devices which require connections to the internet - for receiving or processing information in the cloud, delivering data to the cloud, or for computationally sensitive use cases which require low latency and prompt data processing.

Edge Computing servers can be placed almost anywhere as long as it is close to where the computation or data processing is needed. It can be in cabinets along the road and on roof tops, at the base of cell towers, and even in commercial establishments. Companies such as [Vertiv](#) are building prefabricated micro datacenters complete with chassis and infrastructure (power, cooling, monitoring, racks). These look like a small shipping container which allows them to be readily transported and placed onto concrete slabs.⁴ They are designed to function in a wide variety of environmental conditions.

Another form of edge computing is the processing of data on the device. Some companies are providing chips in their devices to perform on-board computing that in the past was done at a central data center. For example: Apple's face recognition was originally performed at a datacenter but is now performed on the device. Google and Amazon are providing their smart speakers with some on-board processing capabilities as well.

With computational capacity closer to the device (or on the device), the request for "what's the weather" does not have to travel hundreds or thousands of miles to be processed, shortening response time. This frees up computing resources and bandwidth at the cloud for projects that require more complex data processing.

The major benefits of edge computing are the ability to process data quickly which reduces latency, and to store data closer to the device. Data processing for applications such as virtual and augmented reality, self-driving cars, and smart cities⁵ can be performed closer to the source/device with faster response times.

Per Stephan Blum, CTO and co-founder at [PubNub](#), Google, Apple, and Amazon have spent millions investing in their edge so that their AI (artificial intelligence) technology can answer your requests quickly.⁶ The more processing their smart speakers can perform locally (closer to the device or on device), the less they need to rely on the cloud, and the quicker the response.

Content Delivery Networks (CDN)

The concept of processing and storing data closer to the device (one that is requesting or generating it) for the purpose of providing better performance or faster response time is not new. In the late 1990's, [Akamai Technologies](#) solved the problem of web traffic congestions by introducing Content Delivery Networks (CDN).

A CDN has servers linked together to provide faster webpage response (how quickly a webpage loads on your device) and faster content delivery (how quickly content is delivered to the device). Improved response time and connectivity can be achieved by placing servers at exchange points between different networks. Internet exchange points are locations where various internet service providers connect to exchange traffic originating at different networks.⁷ CDN reduces latency by decreasing the distance content travels. CDNs do not host content but they improve performance by caching/storing content closer (at the edge) to the recipient device. Today most web/internet traffic routes through CDNs, including traffic from major streaming companies such as Netflix, Amazon, and Hulu. Imagine the buffering time if this was not the case.

While Edge Computing is like CDN, the difference is that edge deals with moving the hosting of computing services as close as possible to the people and devices who will use them, minimizing the distance the data must travel.⁸

The primary benefits of CDN are:

- **Faster website load times/decreased latency** - distributing content closer to website users results in shorter travel distance and faster service.
- **Reducing bandwidth** - caching/storing high demand content closer to the user reduces the request load on the primary content servers.
- **Increasing content availability and redundancy** – caching/storage content at various distributed CDNs ensures availability in case of a downtime or data corruption at one of the CDN locations.

Why Does 5G Matter?

By design, [5G](#) has greater bandwidth than 4G/LTE. Additionally, telecommunications providers incorporate edge computing strategies directly at the tower to reduce latency and increase the speed of communication processing. Compared to 4G/LTE, 5G is touted as being ten to twenty times faster. This delivers faster, near real-time processing of communications, and this is going to be essential for backend data processing for IoT/IIoT devices and autonomous vehicles in the future.⁹

5G edge architecture moves application hosting away from centralized public clouds and into an array of distributed servers at the telecom's network edge, closer to the customer's device thereby decreasing latency and increasing bandwidth.¹⁰

Autonomous Vehicles (AV)

One of the primary applications of Edge Computing and 5G cellular technology will be in the growth of autonomous vehicle systems.

Although we are years away from true level 5 AV, it will have a hard time succeeding without Edge Computing and 5G. To replace human drivers, an autonomous vehicle's sensors, software, and hardware stack must react to road and vehicle conditions in real-time.

In current autonomous test fleets, all the intensive computational work is handled by onboard computing hardware. When computing software needs to be updated, the OEM can do this over-the-air; a method used already by Tesla. However, if the hardware needs to be updated, it will have to be done individually for each vehicle. This works fine for a small fleet of test or operational autonomous vehicles. Now imagine a dozen years in the future when there are thousands of autonomous vehicles on the road and they all need a hardware update. The logistics of physically updating the hardware would be insurmountable.

What if a solution to this future problem is to eliminate the on-board computational work by off-loading it to the cloud; thereby eliminating the large banks of computing hardware in the vehicle and the routine need of upgrading it.

Additionally, we expect to see V2V (vehicle to vehicle) and V2I (vehicle to infrastructure) communications in the future. V2V/V2I is where sensory data from an autonomous vehicle several cars in front of you can alert your car to an accident or road conditions and slow your car down, or data from infrastructure sensors (traffic lights or road embedded) could alert your vehicle to take a different route due to congestion. Sensor packages on autonomous vehicles and V2V/V2I will generate a significant amount of data. Multiply this by

hundreds of thousands of autonomous vehicles and infrastructure sensors, this data flow would easily congest current cellular networks. This requires a more bandwidth-robust cellular platform, like 5G.

5G's bandwidth can help offload the processing of this large flow of data to the cloud. However, on average, it may take 100 milliseconds for data transmission between vehicle sensors and backend cloud datacenters. For comparison, a typical human eye blink ranges from 100 to 400 milliseconds. AV data processing would require speeds greater than a "blink of an eye." Slower response times may have significant impact on the reaction of self-driving vehicles.¹¹ This is where Edge Computing can help reduce AV response time by performing the data computation locally on the edge, thereby eliminating the latency associated with backend cloud datacenter. Both (edge computing and 5G) are needed to achieve the desired, low latency response times for autonomous vehicles - somewhere below 10 milliseconds.¹²

Too Good to be True?

Edge Computing appears to offer significant promises, or is it too good to be true? There are a few drawbacks of this technology to consider:

- **Increase in hacking targets** - The significant growth of consumer/industrial smart devices and edge servers greatly increases the overall attack surface which amplifies target opportunities for threat actors.
- **Need for additional hardware and software** - Many smart device manufacturers are off-loading numerous functions from the cloud onto the devices themselves. For example, Apple off-loaded facial recognition capability from the cloud directly onto the iPhone. This requires additional processors and/or software which increases the cost of the device.
- **Limited power and backhaul resources in rural America** - Don't expect to see Edge Computing in rural America soon due to a scarcity of three-phase power in such areas. Edge computing systems require three-phase power for the support (HVAC) and server equipment. Single phase power does not have the continuous waveform found in three-phase and wouldn't be sufficient for such systems. Additionally, to deliver digital data, 5G requires a wired fiber optic backhaul which may not exist in parts of rural America.¹³

Conclusion

Edge Computing is transforming the way data is handled, processed, and delivered for millions of IoT/IIoT devices around the world. The growth of internet-enabled smart devices along with new, computationally intensive applications (e.g., gaming, AV, AR, VR) requires near, real-time data processing. This will continue to drive edge computing systems¹⁴ in an effort to decrease processing latency, and shift from 4G/LTE to 5G for greater bandwidth to wirelessly handle the larger data flows.

In addition to the several use-cases discussed already, there could be other applications shifting to the edge computing model using 5G data transport as it may be more efficient and responsive. Some of these applications may include:

- Fleet maintenance and monitoring
- Smart building technologies
- Predictive maintenance
- Smart cities and infrastructure

Considering the rising demand for IoT and industrial IoT (IIoT) applications, and ongoing research and development in artificial intelligence and 5G connectivity technologies, there is significant growth potential with Edge Computing; it may likely reach maturity far more quickly than expected.

Contact Us

To learn more about how Intact Insurance Technology can help you manage online and other technology risks, please contact Dan Bauman, VP of Risk Control at dbauman@intactinsurance.com or 262-951-1455.

References

- ¹ Meulen, Rob van der. (October 3, 2018). "What edge computing means for infrastructure and operations leaders." Gartner. Accessed January 2022. <https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders>; <https://www.ibm.com/cloud/what-is-edge-computing>
- ² Miller, Paul. (May 7, 2018). "What is edge computing?" Accessed January 2022. <https://www.theverge.com/circuitbreaker/2018/5/7/17327584/edge-computing-cloud-google-microsoft-apple-amazon>
- ³ Gold, Jon; Shaw, Keith. (June 29, 2021). "What is edge computing and why does it matter?" Network World. Accessed January 2022. <https://www.networkworld.com/article/3224893/what-is-edge-computing-and-how-it-s-changing-the-network.html>
- ⁴ Fulton, Scott. (February 1, 2021). "5G and edge computing: How it will affect the enterprise in the next five years." ZDNet. Accessed January 2022. <https://www.zdnet.com/article/5g-and-edge-computing-how-it-will-affect-the-enterprise-in-the-next-five-years/>
- ⁵ Ibid 3
- ⁶ Overby, Steven. (November 30, 2020). "How to explain edge computing in plain English." The Enterprisers Project. Accessed January 2022. <https://enterpriseproject.com/article/2019/7/edge-computing-explained-plain-english>
- ⁷ "What is a CDN? How do CDNs work?" Cloudflare. Accessed January 2022. <https://www.cloudflare.com/learning/cdn/what-is-a-cdn/>
- ⁸ Ibid 5
- ⁹ Ibid 3
- ¹⁰ Ibid 4
- ¹¹ Raza, Muhammad. (July 18, 2018). "Edge Computing" An introduction with examples." BMC Blogs. Accessed January 2022. <https://www.bmc.com/blogs/edge-computing/>
- ¹² "What is edge computing?" STL Partners. Accessed January 2022. <https://stlpartners.com/articles/edge-computing/what-is-edge-computing/>
- ¹³ Fulton, Scott. (June 1, 2020). "What is edge computing? Here's why the edge matters and where it's headed." ZDNet. Accessed January 2022. <https://www.zdnet.com/article/where-the-edge-is-in-edge-computing-why-it-matters-and-how-we-use-it/>
- ¹⁴ Ibid 3