Edge-Cloud Computing and RAN Bandwidth Limitations

Justin Wasser

University of Maryland Global Campus

ITEC 625: Computer Systems Architecture

Professor Cirel De Laus

8/09/2022

Abstract

The following research paper examines the Internet of Things (IoT), edge computing, fog computing, cloud computing, and how Radio Access Network (RAN) 5G is facilitating a convergence of these systems; thereby creating edge-cloud computing. Moreover, conclusions reached from the analysis of these systems are used to inform the technology purchase recommendations I would make for a client.

Introduction

As a professional tasked with making technology purchase recommendations for a client, it was necessary to examine current state-of-the-art edge computing technology as well as periphery technologies that will impact future state-of-the-art edge computing systems. Therefore, research was conducted on the growth of the Internet of Things (IoT), current and future state-of-the-art Radio Access Network (RAN) bandwidth capabilities and limitations, as well as cloud computing and fog computing architecture. After synthesizing all of the research conducted an IT infrastructure purchase recommendation was reached that will allow the client's business to leverage innovative IoT applications as they emerge.

Edge-Cloud Computing and RAN Bandwidth Limitations

As the world becomes increasingly connected via the internet, more data is being collected, transmitted, and stored than ever before thanks largely in part to the expansion of the Internet of Things (IoT). In the near future, as exponentially more information is processed; transferring and housing raw data using traditional cloud-based computing will become impractical for large businesses. This development will be due to bandwidth limitations and associated costs due to the scope and scale of data needing to be transferred. Therefore, demand for edge-cloud computing systems will increase, and this computing architecture will become the preferred infrastructure used for state-of-the-art IoT applications. Therefore, if I were making a purchase recommendation for a client, I would recommend they invest in edge-cloud computing infrastructure now, so their business is prepared to leverage future IoT innovations.

Several technologies must be explored to understand why the demand for edge computing infrastructure will increase. Technologies to be discussed must include IoT, cloud computing architecture, current state-of-the-art bandwidth capabilities, an in-depth examination of edge and fog computing architecture, and finally speculation on future state-of-the-art edge-cloud computing architecture.

Internet of Things (IoT)

The world has become increasingly connected due to the proliferation of high-speed internet and as a result "the internet is evolving rapidly toward the future "Internet of Things" (IoT), which will potentially connect billions or even trillions of edge devices" (Pan & McElhannon, 2018, pg. 1). Some of the biggest IoT applications thus far include self-driving cars, smart homes, wearable health monitoring devices, and "industrial Internet of Things systems" (Edge Computing Task Group, 2018, pg. 2); with many more innovative applications in development including remote surgery (Biagi, 2021). These IoT applications have already led to several issues "such as high latency, low Spectral Efficiency (SE), and non-adaptive machine type of communication" (Ai et al., 2018, pg. 77). Furthermore, these issues are likely to worsen as more devices are connected via cloud computing, even if bandwidth capabilities improve. The reason being the rate of devices connected to the internet will likely outpace the development of increased bandwidth. Additionally, cloud computing architecture is not well suited to fit the rapidly expanding data requirements of Internet of Things applications (Ai et al., 2018).

Cloud Computing Architecture

Why is cloud computing a poor fit for the data needs of IoT applications? To answer this question, a definition of cloud computing architecture is needed. Cloud computing is defined as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources" (Ruparelia, 2016, pg. 4). Cloud computing has many strengths, one of which offers users the ability to scale up or down their computing resources very easily based on their current business needs (Ruparelia, 2016). Furthermore, cloud computing services offer users and devices connected to it the flexibility to access its services from anywhere there is an internet connection. However, the same attributes that make cloud computing architecture a good choice for many business needs are ultimately a detriment concerning its use for supporting the IoT. To that point, one of the defining characteristics of cloud architecture is that it relies on massive, centralized data centers (Pan & McElhannon, 2018). Furthermore, due to their size, there is a limited number of data centers, which results in devices connected to cloud services being physically located very far from the data center it needs to communicate with (Pan & McElhannon, 2018). This design is problematic for IoT deployment because it is very costly to constantly be sending large data streams to far-off data centers. Moreover, some of the more innovative IoT applications of the future will need lower latency than cloud computing will be able to provide (Pan & McElhannon, 2018).

Current & Future Bandwidth Capacity

Lower latency is directly correlated with higher bandwidth, therefore at this point, it is necessary to define what bandwidth is referring to. To that point, bandwidth is defined as “the amount of data which a network can carry over time, usually expressed in bits per second” (Bigelow, 2021, para. 21). Furthermore, while wired connections will likely always be the preferred connection mode when speed and bandwidth are the only considerations, its lack of flexibility disqualifies it from being used for many mobile-based applications. Furthermore, because network-connected mobile applications/devices are essentially another name for the Internet of Things, the benefits that state-of-the-art wired network connections offer (fiber-optic) do not need to be examined for this discussion. For example, there is likely to be a proliferation of self-driving cars in the near future, and they will need to be connected to the environment around them to function effectively; however, that will only be possible through a wireless network (Bigelow, 2021). Furthermore, self-driving cars are just one example of a great many potential IoT applications that will require a wireless network connection to function, therefore, a discussion of wireless network capabilities and limitations is required at this time.

The appropriate name for wireless network technology is Radio Access Network (RAN), and the cutting edge of RAN technology is the fifth generation version otherwise known as 5G (Ai et al., 2018). A defining aspect of RAN 5G technology is its ability to offer "ultra-reliable and low-latency communications (URLLC)" (Mukherjee, 2020, pg. 205). Furthermore, 5G is a major advancement in RAN technology as it can transmit data "at multigigabit speeds, with potential peak speeds as high as 20 gigabits per second (Gbps) by some estimates" (Gillis & Gerwig, 2022). This technological advancement is critical to the development of the Internet of Things, as wireless communication technology that is capable of groundbreaking reliability and low latency is necessary for many cutting-edge IoT applications, with self-driving cars being just one example (Mukherjee, 2020). Nevertheless, despite its robust networking capabilities, 5G is not the end of the discussion when it comes to the question of how to best support the coming expansion of the Internet of Things but rather just the beginning.

To that point, while 5G is technologically capable of supporting the networking needs of some IoT applications, there is a limit to the bandwidth that 5G can provide. Furthermore, due to the enormous amount of data IoT applications will produce the supporting computing infrastructure must be designed in such a way as to avoid overloading available bandwidth while remaining cost-effective. The logical answer to the bandwidth concern IoT raises is to move as many essential computing services as possible closer to the origination of the data IoT applications create (Bigelow, 2021). Furthermore, redesigning the IoT systems architecture in such a way decreases latency and network congestion while also having the potential to lower costs by significantly decreasing the amount of data being sent to far-off cloud datacenters for storage and analysis (Pan & McElhannon, 2018). Lastly, the architecture just described was edge computing; therefore, further analysis of edge computing is required at this time.

Edge Computing

Edge computing is defined as "a distributed information technology (IT) architecture in which client data is processed at the periphery of the network, as close to the originating source as possible" (Bigelow, 2021, para. 1). Moreover, the edge is not objectively measurable, instead it is a "logical layer" (Edge Computing Task Group, 2018) that changes depending on the technical requirements of the application in question (Edge Computing Task Group, 2018). Furthermore, the growing demand for edge computing arises from the weaknesses of the centralized data-center architecture that defines cloud computing. In simple terms, edge computing architecture takes computing responsibilities away from the giant data center that is the cloud and moves those computing processes close to where data is being generated. Lastly, thanks to advancements in RAN technology, the latest development in edge computing is to interlink the edge with the cloud. This development is known by several names including, "mobile edge computing (MEC)" (Ai et al., 2018, pg. 78), "multiaccess edge computing (MEC)" (Mukherjee, 2020, pg. 219), and finally "edge-cloud" (Li et al., 2022, pg. 805). Furthermore, this combination of edge computing and cloud computing gives rise to another layer that lies between the edge and the cloud, which is appropriately named fog computing (Bigelow, 2021).

Fog Computing

Fog computing is less straightforward than its edge and cloud counterparts because it only exists in relation to the edge and the cloud. While edge computing is defined as computing that is located where data is being generated and cloud computing as a massive data center located hundreds of miles away or more, fog computing is vaguely described as somewhere in between (Bigelow, 2021). Furthermore, not only is fog computing in between edge and cloud computing in terms of its physical location, but it is also in between edge and cloud computing in terms of its size, computing power, and computing responsibility as part of a larger system (Ai et al., 2018). Moreover, “from a functional point of view, a fog node has several functions, including networking, computing, accelerating, storing, and control” (Ai et al., 2018, pg. 82-83). Essentially, fog computing can perform all of the critical computing functions just listed while being close enough to the source of IoT device data generation to achieve the low latency needed for IoT applications; all while not using nearly as much bandwidth as the cloud would use to perform the same functions. This additional layer is needed in certain situations where IoT devices are too spread out to have a clearly defined edge, or for situations where it is not possible to place computing resources where data is being collected (Bigelow, 2021). Lastly, it must be noted that “fog computing and edge computing share an almost identical definition and architecture, and the terms are sometimes used interchangeably even among technology experts” (Bigelow, 2021, para. 17)

Fog computing is one innovation that has resulted from the fusion of edge computing with cloud computing. This evolution was made possible thanks in large part to advancements in Radio Access Network (RAN) technology with the development of 5G (Mukherjee, 2020). Additionally, this integration has also led to what is now known as “mobile or multiaccess edge computing (MEC)” (Mukherjee, 2020, pg. 219) or “edge-cloud computing” (Li et al., 2022, pg.

805). The edge-cloud architecture provides unprecedented capabilities which have opened the door for more innovative IoT applications.

Armed with a basic understanding of edge, fog, and cloud computing, one might ask how is it possible that edge computing successfully performs many of the computing tasks that used to be performed by massive cloud datacenters? The answer is through innovations in both computer architecture and computer infrastructure.

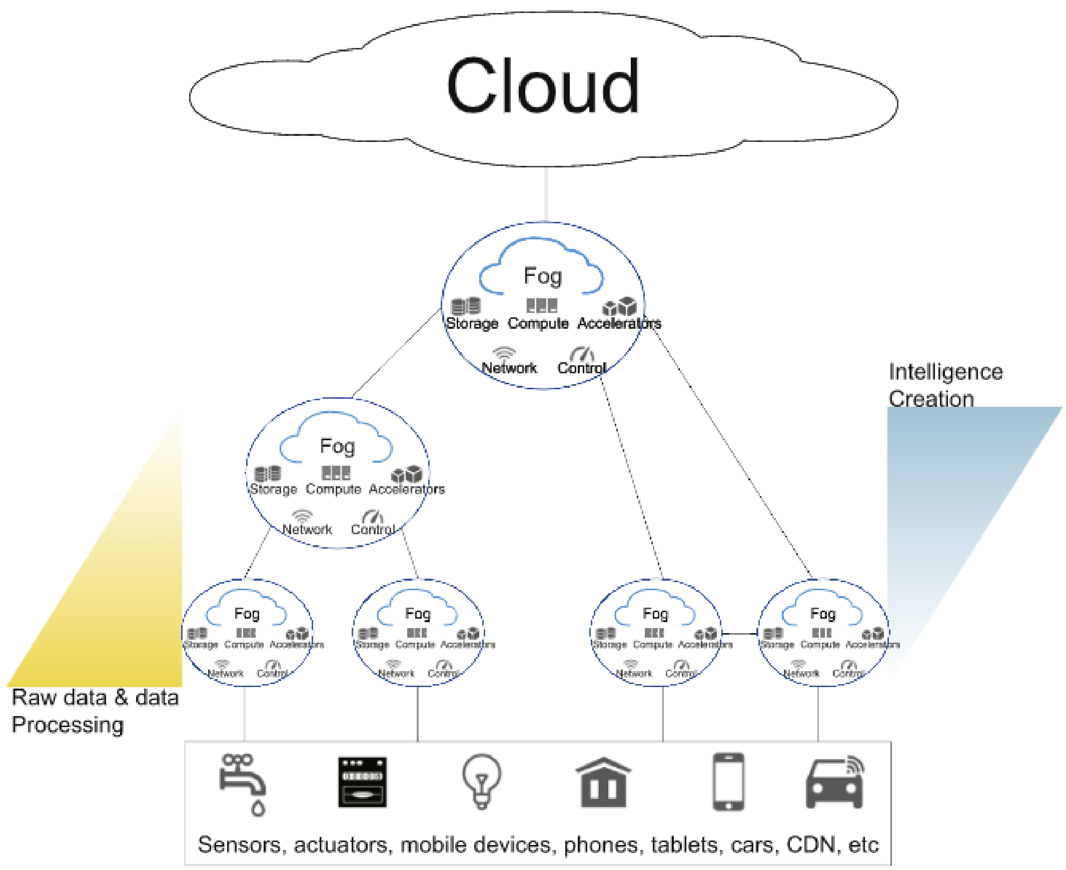


Fig. 5. A typical hierarchical architecture based on fog computing.

(Ai et al., 2018, pg. 83)

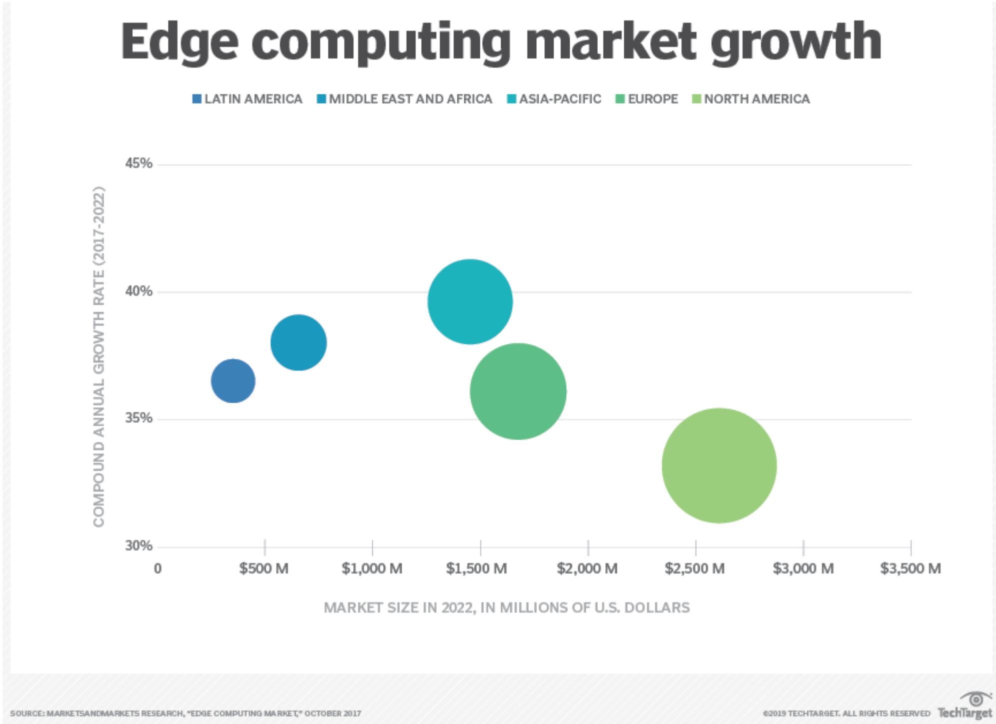
Edge-Cloud Computing Architecture

Edge computing infrastructure can be broken down into edge devices and edge servers, which combine to form the distributed IT architecture mentioned earlier (IBM Technology, 2019). Furthermore, both edge computing and cloud computing have components that facilitate “connecting, computing, storage, and communication” (Edge Computing Task Group, 2018). The main difference between to two computing structures are their size/computing power, and their location relative to where data is being generated. As previously discussed, edge computing removes some of the more repetitive, and less intensive computing operations from the cloud and assigns those responsibilities to edge computing. Those tasks include data “acquisition, processing, and synergistic combination of information” (Li et al., 2022, pg. 808). However, how is possible that edge computing can accomplish all of these tasks without the robust computing power available to cloud computing?

Edge computing accomplishes its role by utilizing “virtual technology of containers, also known as hypervisor technology” (Li et al., 2022, pg. 809) to create multiple non-hardware-intensive virtual operating systems used to run the applications edge computing supports (Li et al., 2022). This container architecture is defined by “three key components: container virtue, container isolation, and cloud-edge collaboration” (Li et al., 2022, pg. 809). Furthermore, docker technology is used in conjunction with hypervisor technology for increased platform optimization (Li et al., 2022). Specifically, “docker has the following advantages: 1. the speed of creating containers is much faster than directly creating virtual machines; 2. hardware size can be reduced; and 3. Version control is easier for operators” (Li et al., 2022, pg. 809). These advancements in computer architecture have allowed processes like “data fusion” (Li et al., 2022, pg. 808) to be performed at the edge of networks by utilizing virtual machines (VM) and associated administrative software. Furthermore, using VMs and hypervisor software allows edge computing to perform low-level data analysis and synthesis without nearly as much physical hardware as was needed before the implementation of VMs (Li et al., 2022). Lastly, without this development, it would not be possible to create an edge-cloud architecture.

How the Edge-Cloud will Affect Purchase Recommendations for a Client

Understanding how edge computing is leveraging RAN 5G to interlink with cloud computing is critical when making purchase recommendations for a client. With that said, if I were making a purchase recommendation for a client I would recommend they start investing in edge servers in the short term with the long-term goal of creating an edge-cloud infrastructure for their business. Furthermore, I would recommend starting this process immediately as “edge computing is likely to become as necessary to an organization's overall platform as cloud is” (Longbottom, 2020, para. 18). Additionally, “leaving it for a future time could result in major changes being required to existing cloud architectures and more time and money spent on trying to overlay and integrate more physical systems into a highly virtualized environment.” (Longbottom, 2020, para. 18) Lastly, an edge-cloud system will lead to improved business analytics, increased business efficiency, increased interoperability, and will help to future-proof a business.



(Bigelow, 2021)

Conclusion

In conclusion, after analyzing the architecture of RAN 5G, edge computing, fog computing, and edge-cloud computing, a common theme emerges; namely, these technologies rely upon numerous and increasingly distributed designs. Therefore, even if speculation regarding RAN 6G proves accurate, and can provide "a tenfold to fifty-fold improvement over 5G in various metrics" (Mukherjee, 2020, pg. 277) the corresponding increase in the number of IoT devices coupled with the increasing complexity of IoT applications will require the current distributed computing trend to continue. Additionally, RAN 5G technology relies on high-frequency bands that have a more limited range than their 4G predecessor, therefore it is likely that distance would become even shorter with the advent of RAN 6G (Gillis & Gerwig, 2022).

Once again, this development would only serve to exacerbate the current distributed computing architectural trend.

For reasons just stated, I believe that in the near future, state-of-the-art edge computing will consist of smaller and even more numerous edge-cloud servers that act as nodes for agile-edge-cloud-clusters that can be scaled up or down based on computing needs in a given area at a specific time. Lastly, an agile-edge-cloud-cluster architecture will provide lower latency and use even less bandwidth when performing data processing for IoT applications. Finally, an agile-edge-cloud-cluster architecture will also be a cheaper alternative to traditional cloud computing due to much lower data usage costs.

References

Ai, Y., Peng, M., & Zhang, K. (2018). Edge computing technologies for Internet of Things: a primer. *Digital Communications and Networks*, *4*(2), 77–86. https://doi.org/10.1016/j.dcan.2017.07.001

Biagi, S. (2021, April 13). *AI-enabled Robots Bring Precision to Complex Surgeries*. IoT Integrator. Retrieved August 9, 2022, from https://www.theiotintegrator.com/health/ai-enabled-robots-bring-precision-to-complex-surgeries

Bigelow, S. J. (2021, December 8). *What is edge computing? Everything you need to know*. TechTarget. Retrieved August 5, 2022, from https://www.techtarget.com/searchdatacenter/definition/edge-computing

Edge Computing Task Group. (2018, June 18). *Introduction to Edge Computing in IIoT*. Industry IoT Consortium. Retrieved August 8, 2022, from https://www.iiconsortium.org/pdf/Introduction_to_Edge_Computing_in_IIoT_2018-06-18.pdf

Gillis, A. S., & Gerwig, K. (2022, March). *What is 5G?* TechTarget. Retrieved August 7, 2022, from https://www.techtarget.com/searchnetworking/definition/5G

IBM Technology. (2019, October 1). *What is edge computing?* [Video]. YouTube. https://www.youtube.com/watch?v=cEOUeItHDdo

IDG TECHtalk. (2020, May 1). *What is edge computing?* [Video]. YouTube. https://www.youtube.com/watch?v=3hScMLH7B4o

Li, J., Gu, C., Xiang, Y., & Li, F. (2022). Edge-cloud Computing Systems for Smart Grid: State-of-the-art, Architecture, and Applications. *Journal of Modern Power Systems and Clean Energy*, *10*(4), 805–817. https://doi.org/10.35833/mpce.2021.000161

Longbottom, C. (2020, October 8). *How to implement edge computing in 5 steps*. TechTarget. Retrieved August 9, 2022, from https://www.techtarget.com/iotagenda/tip/How-to-implement-edge-computing-in-5-steps?_gl=1*1riyupt*_ga*OTEzMDI1MDYuMTY1MzYxNjc0NA..*_ga_TQKE4GS5P9*MTY2MDAzNTc1MS42OS4wLjE2NjAwMzU3NTEuMA..&_ga=2.178301601.406237832.1659642669-91302506.1653616744

Mukherjee, A. (2020). *5G New Radio: Beyond Mobile Broadband* [E-book]. Artech House. Retrieved August 6, 2022, from https://eds-p-ebscohost-com.ezproxy.umgc.edu/eds/ebookviewer/ebook/bmxlYmtfXzIzMjE5MTNfX0FO0?sid=0535eef6-030b-492f-a6fa-338facb60c1a@redis&vid=2&format=EB&rid=3

Pan, J., & McElhannon, J. (2018). Future Edge Cloud and Edge Computing for Internet of Things Applications. *IEEE Internet of Things Journal*, *5*(1), 439–449. https://doi.org/10.1109/jiot.2017.2767608

Ruparelia, N. (2016). *Cloud Computing* [E-book]. The MIT Press. Retrieved August 6, 2022, from https://eds-p-ebscohost-com.ezproxy.umgc.edu/eds/ebookviewer/ebook/bmxlYmtfXzEyMzgwMDFfX0FO0?sid=c9ea0d91-4742-47cf-97ff-c8aaefa01486@redis&vid=1&hid=http://eds.p.ebscohost.com.ezproxy.umgc.edu/&format=EB

Van Der Meulen, R. (2018, August 3). *What Edge Computing Means For Infrastructure And Operations Leaders*. Gartner. Retrieved August 7, 2022, from https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders