2022

Edge Computing Research Paper

BIG DATA MANAGEMENT SYSTEMS AND TOOLS

Group 6: Herbert Gnanaraja, Farhan Hasan, Rayna Ilieva, Vladimir Miskovic, Justine Spero, Melodie YunJu Song,

Contents

Objective

The objective of this project is to research and report on Edge Computing as applied primarily to the upcoming Metaverse Virtual Reality/Augmented Reality infrastructure. The Metaverse will combine both big data and a virtual reality/augmented reality world representation. For example, users go shopping in a digital store for digital goods using a VR headset. We discuss the scalability, reliability, speed, efficiency, and privacy/Security concerns in each section below.

Introduction

Cloud computing enables networking, databases, storage and analysis through the internet on computing hardware of data servers in data centers. Cloud computing can service a wide variety of clients in different locations simultaneously.

Edge computing is a variation of the cloud computing model but focuses more on speed by closing in the distance between the source of the data and the computing resources. Cloud computing and edge computing are not in competition with one another, they will co-exist and complement each other depending on the scenario. Edge computing is an extension of cloud computing. In areas where processing of large data and more computationally heavy data processing is needed, cloud computing will take on those tasks. (Azure.com, 2022). A full comparison of cloud computing and edge computing has been included in Appendix A.

Scalability

Scalability is a feature that is crucial for the success of any technology implemented in the big data realm. Edge computing is typically scaled vertically, where small computational resource pools are located near the source of the data. These are usually cloudlets or Fogs and the system needs to determine the ideal tradeoff between computing latency and transmission latency. The incoming data needs to be evaluated if it should be performed locally, or to Fog/Cloudlet servers or if needed be offloaded to the cloud servers.

The scalability of edge computing heavily relies on the algorithm that decides how and which computational resource to offload the data. The offloading process is very challenging because most of the data that edge computing will handle are from the Internet of Things (IoT). In an ecosystem of IoTs the number of individual IoT devices are high and also the heterogeneity of the devices are high, hence the heterogeneity of the data is also high. Furthermore, each IoT may have their own stringent security requirements which could introduce further latency.

There are several directions edge computing is being optimized for scalability. One of them is the use of mobile edge computing (MEC) and a selective offloading scheme. As the edge servers are aware of their geographical location and there are multiple of them spread out, they can be grouped together into virtual clusters and organized into a hierarchical structure to handle the computational workloads. The solution is a three-layered integration architecture which consists of the user plane, edge computing plane and the cloud computing plane. (Maheshwari et. Al, 2018)

At the user plane the IoT devices are clustered together virtually, at the edge computing plane the geo-distributed edge servers can be organized in a hierarchical structure to optimize the expenditure of resources. At the cloud computing plane any data which requires processing of large amounts of data is offloaded.

Let us shift our focus to the Metaverse, as we discussed before regarding the IoT ecosystem and the real-time data that needs instant processing, the issues with scaling are the same for the Metaverse. Currently the Metaverse might require one device that generates visual and audio data but further down the line there will be additional sensory devices used by the customer for the full experience. This could include motion sensors, temperature sensors and multiple other haptic feedbacks. All these will need edge servers physically close to most of the customers. Figure 1 illustrates the Metaverse IoT ecosystem using edge servers.

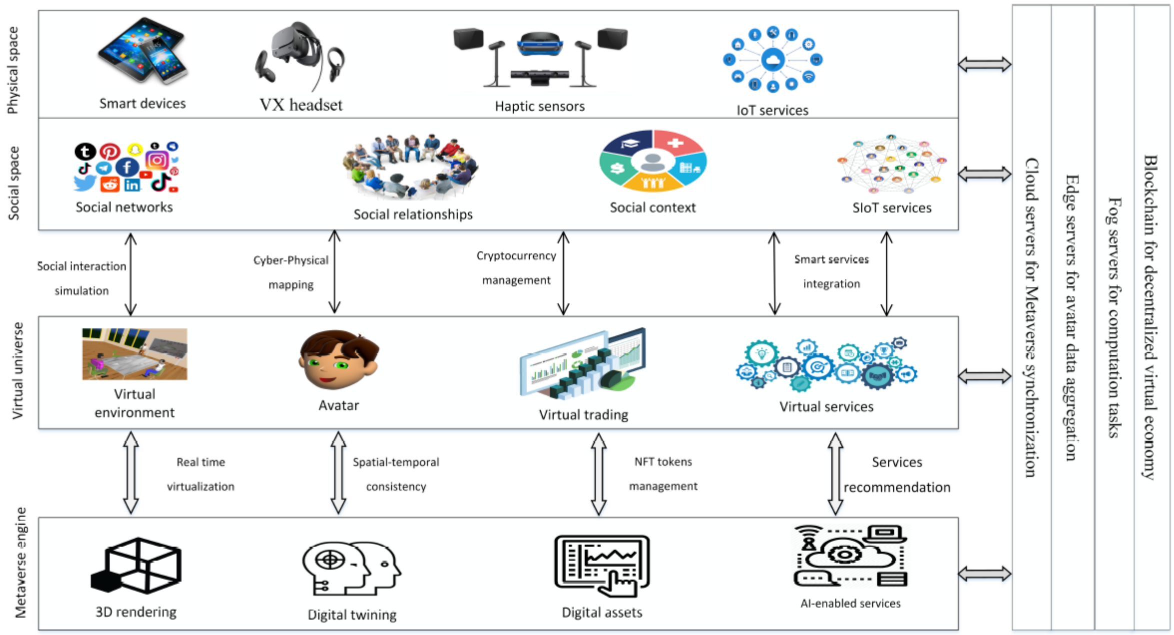


Figure 1: the Metaverse IoT ecosystem, Dhelim et. Al, 2022

Edge computing can add multiple edge devices as the ecosystem keeps expanding and scaling as needed. So, organizations do not need to commit to large scale operations and can incrementally scale by adding endpoint hardware and edge devices. This also cuts cost in the form of resource management as the organization needs to manage only the additional devices (Pratt, 2022).

For most IoT devices fog computing plays a huge role in scalability cloud computing. The idea is to introduce an intermediate layer between final devices and central cloud and is geared towards low latency, geo-distribution, location awareness and mobility support. End to end delays, traffic congestion from simultaneous off-loading and processing of enormous amounts of data are the issues that arise from large physical distances hence Fog computing acts as an extension to the Cloud computing (Bellavista, 2020).

There are many challenges to scaling with edge computing, like any other emerging technology. Scaling always presents itself as a huge challenge but as the technology is implemented more and more companies find ways to scale in innovative ways.

One of these challenges is a convenient form factor that can house all the required components to handle the workload but small enough to be installed in edge locations. The main engine behind the processing power are GPUs and they are well known to require a lot of power and cooling. The more cooling it requires, usually in the form of fans, the bigger these form factors tend to be. One upside is that most of these GPUs are geared for gaming or other use cases, so there needs to be more PCIe cards that are designed and built specifically for edge computing. Smaller FPGA cards are a good alternative that doesn't require much space and cooling.

Artificial intelligence will play a crucial role in optimizing analytics and automation. It could make the scalability process cheaper in terms of resources and hardware. To tackle the large quantity of edge devices one solution is to use data to actually figure out which physical locations will need the extra computing power.

One double edged sword in edge computing scalability by different vendors is the heterogeneity, each company is tackling the device form factor, firmware and data in different ways hence there is no standard that has been set in stone which works for everyone (Burt, 2021).

Reliability

We can look at the reliability of edge computing by asking of it that it perform three related tasks (sourced from Module 1 Part 2: Reliable and Scalable Data-Intensive Applications*)*:

- That the application or service remains running and performs as expected.
- That the service be accessible from any location (where location proximity helps lower latency).
- That there be minimal downtimes (performance is close to real-time)

Application performance

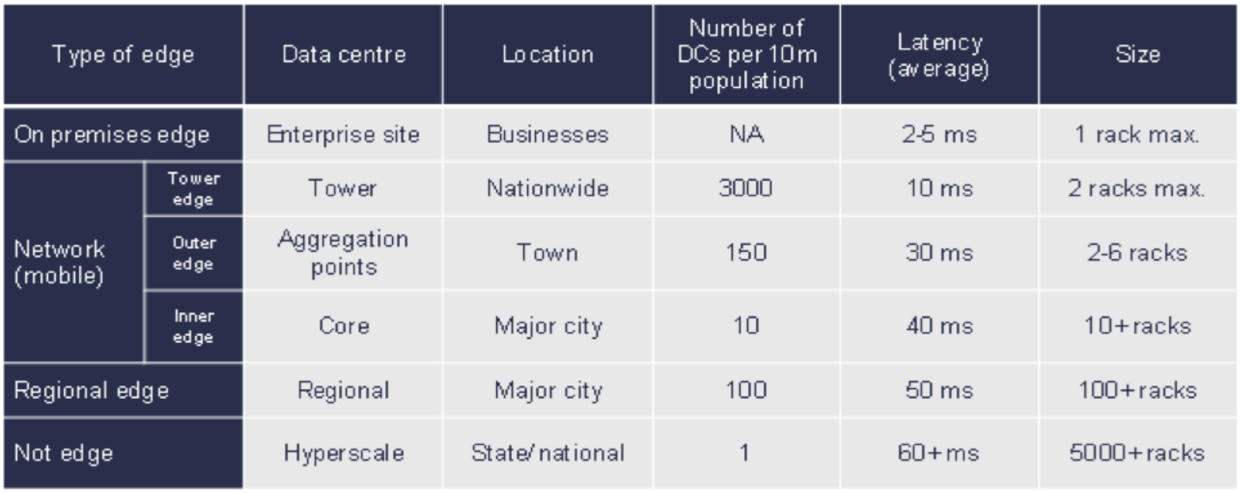
Applications running on IoT devices need to be able to communicate quickly and reliably with central cloud servers. Latency issues arise whenever data is sent across wider networks. A fundamental feature of edge computing networks is to allow IoT devices at the edges of the network to process data. Despite onboarding some computational tasks from centralized servers, data still needs to be sent to and from centralized cloud servers, and therefore latency will always exist but may be reduced (Al-Dulaimy et. Al, 2020).

Services that operate with low and consistent latency are well suited to edge computing implementations (Shi W. et. al, 2016) suggests that health emergencies and public safety are examples of such implementations. In those cases, it is more efficient to offload decision making to the edge of the cloud and IoT devices located there. See Table 1 and Table 2 for the description of the types of edge data centres, example of these data centres, locations, processing latency and size.

Location-awareness

Another feature of edge computing that allows for more varied applicability and greater reliability is location-awareness. Traditionally, large cloud server farms are dislocated from any region. Massive server farms of thousands of servers can be located arbitrarily vis-a-vis the customers they serve. Edge computing brings smaller, more regional server farms closer to end customers. This allows for decreased latency of data pushes and pulls between the IoT edge device and central servers. Additionally, this may open avenues for increased functionality for software and applications running on devices at the edge of the cloud (Varghese B, et.al 2019). Data can be collected and processed at the source that generated them, rather than being sent through the cloud (Shi W. et. al, 2016).

Location-aware applications have been increasing in popularity and have encouraged more frequent implementations of edge computing concepts by service providers. In this way, latency can be minimized by taking advantage of the smaller distance between edge servers and users. Likewise, increasingly more devices in the IoT are not only consuming data, but also generating it (Varghese B, et.al 2019). The importance of offloading some computation to the devices at the edge cannot be overstated. Gigabytes of data can be generated by single IoT sources, which would represent a considerable load to the available network bandwidth. Processing the data locally helps alleviate the burden (Van Rijmenam M, 2019). Edge nodes can be thought of as being dedicated servers located at the edge of the cloud layer, or similarly IoT devices that perform computations more typical of cloud servers (Al-Dulaimy et. Al 2020). Interestingly, an advantage of performing tasks on edge nodes is the minimization of energy consumption. (Varghese B, et.al 2019).



Type of edge		Data centre	Location	Number of DCs per 10m population	Latency (average)	Size
On premises edge		Enterprise site	Businesses	NA	2-5 ms	1 rack max.
Network (mobile)	Tower edge	Tower	Nationwide	3000	10 ms	2 racks max.
	Outer edge	Aggregation points	Town	150	30 ms	2-6 racks
	Inner edge	Core	Major city	10	40 ms	10+racks
Regional edge		Regional	Major city	100	50 ms	100+racks
Not edge		Hyperscale	State/national	1	60+ms	5000+racks

Table 1. Types of edge data centres, Stlpartners.com, 2017.

Table 2. Description of the types of edge data centres, Stipartners.com, 2017.

Type	**Description**
On premises	Servers located closest to customers. Data sent to and from these servers experiences the lowest latency in transmission.
Regional edge	These data centers represent the largest edge computing data centers. They are located near their customer/client locations. Latency is minimized while still taking advantage of edge computing benefits.
Network (mobile)	Telecommunications infrastructure owned by various types of telecom companies. These range from tower edges (physical locations within a region) to inner edge (data centers routing multiple subnetworks).
Hyperscale/not edge	Large data centers comprising thousands of servers. These locations are not considered to be part of edge computing. They can more generally be considered to be part of traditional "cloud computing" models.

Downtimes/Failures

While handling service failure is a feature of all cloud computing nodes implementing with reliability in mind, protocols and network topologies continue to be researched in the case of edge computing. Edge nodes that experience failure need to be able to communicate their failure with the rest of the network. Moreover, the dynamicity inherent in edge computing networks compels the discovery of new communication protocols (Al-Dulaimy et. al 2020).

Fortunately, edge computing is uniquely suited to take advantage of the improvements in cellular network technologies that increase the frequency at which devices communicate wirelessly. 5G technologies show promise in being integrated with edge computing concepts to facilitate low-latency and high-volume data flows (Yu W, Liang F, 2022).

Speed and Efficiency

By its definition and design, edge computing eliminates the need to move data from endpoints to the cloud and back again. Decreasing that travel shaves time off the entire process; this time savings can be measured in seconds, sometimes even milliseconds. That might not seem like much, but travel time -- known as latency -- is a critical consideration in a connected world where real-time decision-making capabilities are necessary for proper functioning of the endpoint devices. For example, autonomous vehicles, industrial and manufacturing IoT deployments and medical use cases all require machines to analyze data and return instructions nearly instantaneously in order to function safely.

Latency reduction

Edge computing can lower latency on a reliable network, bringing workloads and applications closer to digital interactions, which in turn can result in better experiences overall. Solutions such as dynamic connections and data center/colocation services accomplish this by connecting and migrating workloads as close as possible to data acquisition and analysis.

Enhanced scalability
Dynamic network connections enable network provisioning from the customer premises or the cloud to edge locations, allowing businesses to deploy and manage workloads in an agile development model. Enterprises can connect their applications and workloads and scale their network up and down quickly, on-demand.

Provide real-time data provisioning – Edge services can connect to thousands of data centers and on-net locations to run distributed IT workloads close to the edge of the network. Data provisioning lets enterprises store valuable data that is created from remote digital interactions and not miss out on opportunities to improve products and services. Storage-as-a-service has the power to ingest data, reduce data, and make smart decisions about how and where to move data, and send simultaneous copies to multiple locations (Tucker, 2020).

Edge computing offers higher performance than centralized computing because all of the time sensitive computing can be performed at the edge node itself rather than at some distant location after being transmitted across networks. Since high-speed transmissions are required only occasionally when transmitting large amounts of data back up to the central location for processing, transmission may occur 10X less frequently than typically required by conventional systems (Nanoprecise.io, 2022).

Efficiency
Edge Computing takes place when the solution requires low latency, reduced backhaul, and data localization. The Cloud approach takes place when scalability and mobility are the essential conditions required by the business needs.

Privacy/Security Concerns

The benefits of edge computing such as scalability, efficiency, latency reduction and computational resource allocation that allow for smart homes, quality control and logistics in manufacturing, imaging and surveillance also bring about one of its most vulnerable challenges - security (Zeyu et al, 2020). In theory, edge computing promises enhanced security and privacy because data is processed and stored closer to the data source and data is not sent to the cloud. In practice, complex and heterogeneous network environments, such as the end-user, their devices' computational capacity, and system and software security design makes it difficult to implement much classical security and privacy protection for them (ibid).

In this section, we illustrate the contemporary challenges of edge computing for privacy and security, known attacks and mitigation techniques. Three broad genres of privacy issues, namely context-aware privacy, location-aware privacy, and decentralized privacy concerns have been extensively discussed and researched (Gao et al, 2020). Briefly, context-aware privacy concerns stem from the context-dependent rules that are made based on a user's "situation". Privacy rules are made dependent on the information provided by the user (e.g., geolocation, time, type of device, identity). By default, when distributing data processing to devices on the *edge* of the network, individual nodes such as core routers, regional servers, and WAN switches become part of the computing infrastructure, but the users' data in the edge computing environment are not under the control of users. Context-based issues lead to 6 types of attacks (see Table 3), the most popular among them are background knowledge attacks which occur when the edge infrastructure deals with large context-based information: these types of attacks can proliferate into other types of attacks and its adversarial are hard to model and predict, making attacks hard to defeat. Past approaches to mitigate context-aware concerns include differential privacy preserving schemes (e.g., Privacy-Preserving Content-Based Publish (PCP)), k-anonymity-based privacy preservation, and t-closeness, these approaches perform better on stable networks compared to other dynamic or customizable networks. More promising are Markov Decision Processes (MDP), which offers high-level privacy protection for mobiles in edge-computing environments (Gao, 2021).

Table 3. Edge-computing privacy issues and known attacks, adapted from Gao, 2021.

	Context-aware privacy	Location-aware privacy	Decentralized privacy
Background knowledge attack	Known cases	Known	Known
Collusion attack	Known cases	Known	Known
Linkage attack	Known cases	Rare	Known
Structural attack	Known cases	Rare	No known occurrence
Forgery attack	Rare	Known	No known occurrence
Eavesdropping attack	Known	Known	Known
Sybil attack	Known	Rare	No known occurrence

The second type of privacy concern is location-aware privacy: edge computing technology enhances services and applications that require real-time monitoring (e.g., weather alerts, disaster warning, emergency rescue) through retrieving and processing accurate geolocation in volumes. The denser the number of devices/users, the more edge nodes are deployed and connected to the end devices. The drawback is that the user's current specific location as well as the past locations stored and processed within the edge node are exposed, leading to vulnerabilities in wireless distribution systems. Exposing user locations lead to background knowledge attacks, collusion attacks, forgery attacks, eavesdropping attacks, to name a few. In particular, collusion attacks happen when different adversaries acquire different background knowledge of the user and share information with each other to launch a joint attack (Tian et al, 2020, source). Spatial cloaking, a technology in which users are allowed to *cloak* their location when sending information to edge nodes, such as those adopted by trusted third-party privacy applications, and network k-anonymity (Xu et al, 2019), are used to prevent location-aware attacks.

The third type of privacy concern rests in the decentralized design of edge computing, where inherently a system's master key can be in every end user device, posing decentralized privacy concerns in real-world application. For example, malicious "linkage attacks" from uncertified servers and devices are known to disrupt entire networks and re-establish false edge nodes to rapidly flux the network with data volume and data sources, and ultimately collecting user data that were theoretically decentralized from multiple data sources as linkages from the false users reach a majority threshold. To prevent attacks, there are third-party tools like end-device identity authentication protocols that differentiate certified servers from uncertified servers are used to prevent intrusion; and decentralized federated learning where users can establish local learning parameters from a local model which updates to a trusted third party; further, new developments in blockchain technology and federated learning allow conditional privacy models that enable frequent change of anonymous certificates can also prevent edge computing linkage attacks (Lu et al, 2020). Eavesdropping attacks can be a concern related to decentralization, this is a form of attack whereby the unauthorized attacker uses the internet or electromagnetic wave to steal sensitive information by intercepting during data transmission.

To date, the terminal, data, network, and systems risks in edge computing and the security challenges involved can be typologized into the following:

- Malicious terminal penetration attack: referring to the edge computing terminals that access IoT's sensing devices, which often are too dynamic to perform distributed authorization and behavior control. These uncontrolled edge environments, if accessed by unauthorized sensing devices, may be exploited by hackers to create system paralysis or result in malicious data leakage (Yahuza et al, 2020).
- Sensitive data stealing and tampering: During data exchange and data sharing with the edge computing network, hackers may infiltrate, hijack, and mutilate data and data exchange through network attacks. Traditional single-point attacks are now multipoint attacks that are harder to detect and defend (Xu et al, 2020). Researchers at Rutgers University show that VR/AR headsets that capture and record speech-associated facial dynamics to steal biometrics (voice prints) can be easily used to steal credit card data and passwords elsewhere (Shi et al, 2021).
- Ubiquitous transmission intrusion: Ubiquitous transmission networks such as 4G and 5G networks and mesh networks are risk windows in edge computing, attacks on the transmission network lead to paralysis and server collapse because devices require data continuity to provide accurate service. Transmission intrusions are a key vulnerability to edge computing because counterattack and detection evasion techniques evolve quickly (Li et al, 2019). Commonly known malicious attacks we are familiar with include Denial of Service (DoS), snooping, port scanning, remote login, etc (Singh et al, 2021).
- Attack transmission of all time domain and interconnection services: in edge computing, attacks can quickly spread to all connected devices up to the millions, at the least, it threatens the security of businesses and users that rely on these devices for logistical planning, and at worst, jeopardize critical operations that require accuracy and timeliness in data processing and data transmission such as an emergency response system. Examples are infiltrations of smart-home models, attackers can monitor power and water

utilization of smart-home devices and predict when the buildings are empty and good for robbery (Zeyu et al, 2020).

Beyond the mitigating technologies mentioned in the aforementioned paragraphs, regulatory developments in Multi-Access Edge computing (MEC) are currently setting the standard for edge computing in the industry to address challenges in data privacy and security, access control, attack mitigation, and anomaly detection (ETSI, 2018). Though not a requirement for devices, MECs are edge computing technologies that apply principles that improve security: such as requiring (1) Network Functions Visualization (NFV) alignment, whereby enhancements must be made to include virtualized platforms; (2) mobility support: User devices, specifically consumer-oriented services must maintain mobility functions that ensure the continuity, mobility of application, and mobility of application-specific user-related information; (3) deployment independence: the technology must account for deployment at various locations in the edge computing environment (e.g., radio node, aggregate point, edge of data centers). The ETSI which establishes these edge computing outline two requirements: first, The MEC system must provide secure environments for the user, the network operator, third-party provider, the application developer, the content provider, and the platform vendor; and second, “the MEC platform shall only provide a MEC application with the information for which the application is authorized (ETSI, 2018, p23).

Security and privacy concerns in edge computing is a concern at the forefront of any Metaverse development. Programmers conceptualized that Metaverse will replace the flat-surface (internet, webpages, 2D social media) and become a wholly interactive universe that is the virtual representation of the world we live in, with add-ons (Dhelmin, et al., 2015). At present, Metaverse evokes images of virtual social spaces in Avatars, immersive gaming, and convenient services like virtual shopping. These services require the convergence of blockchain and AI technologies to provide real-time, no latency 3D rendering, token management, spatial-temporal consistency and AI-enhanced recommendations, all of which a distributed Metaverse through edge computing technology can achieve. In the Metaverse, privacy concerns in the “2D” world are exacerbated and its reach widened. The two most obvious concerns are user safety and information privacy. Regarding user safety, hackers can acquire others’ background knowledge, geolocation, and user behaviors to perform cyber-bullying, identity theft, and sexual harassment on targeted or diffused groups of users. Related to information privacy, any information that anonymous users search, input or retrieve can still be corroborated with other contextual information to guess the real identity of users, exposing users to forgery attacks, eavesdropping attacks, and Sybil attacks. Due to the above concerns with edge computing, technology firms developing applications such as AR/VR and interconnected IoTs have been heavily invested in the R&D of approaches to protect user security and privacy, and adherence to, and development to enhance industry standards (such as the MEC) is a first step to bringing the Metaverse to life.

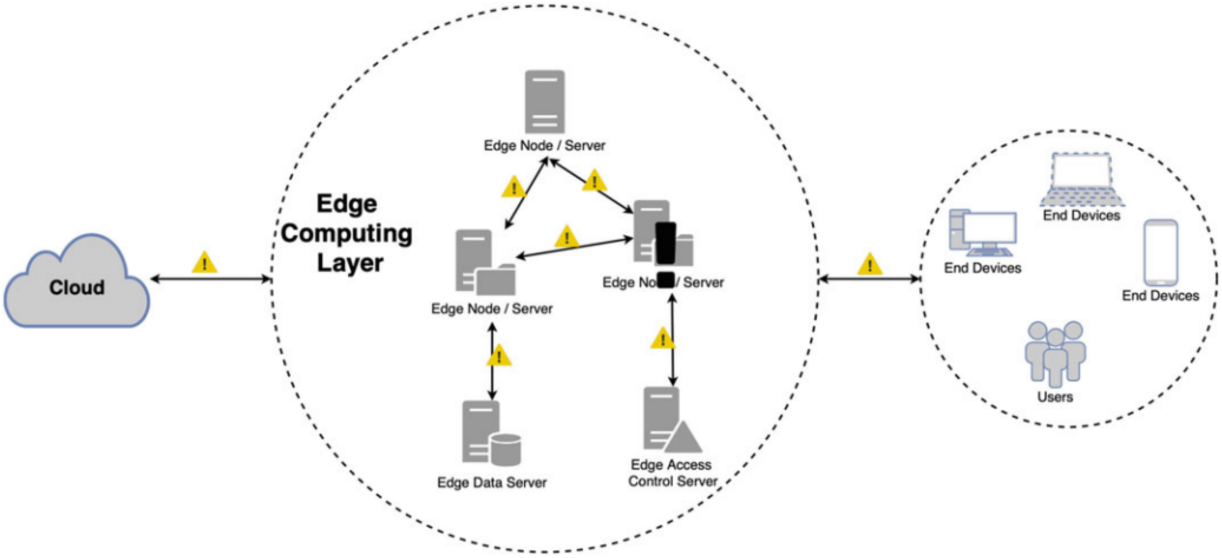


Figure 2. Known vulnerabilities and attack windows in edge computing, Longxiang et.al, 2021.

Conclusion

The vast amounts of data collected and stored as part of the Metaverse necessitates the use of Edge Computing. The Metaverse will require constant processing and updating of data and need a quick, reliable, and efficient means of dealing with this data. In the Metaverse, privacy concerns in the “2D” world are exacerbated and its reach widened. The two most obvious concerns are user safety and information privacy. The Metaverse is still in its infancy and its needs will likely evolve as it is implemented, and this poses an opportunity for the expansion of our capacity in data processing, potentially even beyond the current limits of Edge Computing.

References

Al-Dulaimy, Auday & Sharma, Yogesh & Gokan Khan, Michel & Taheri, Javid. (2020). Introduction to edge computing. 24 Edge computing: models, technologies and applications Retrieved online from: https://www.researchgate.net/publication/344218125_Introduction_to_edge_computing/citation/download

Ashtari, Hossein. 2022. Spriceworks.com Edge Computing vs. Cloud Computing: 10 Key Comparisons. Retrieved online from: https://www.spiceworks.com/tech/cloud/articles/edge-vs-cloud-computing/#:~:text=cloud%20computing%201%20Speed%20%26%20Agility.%20Edge%20solutions,Performance.%20...%204%20Reliability.%20...%205%20Security.%20

Bellavista, Paolo. 2020. Edge Computing for Extreme Reliability and Scalability. University of Bologna. Retrieved online from: http://amsdottorato.unibo.it/9433/5/Tesi-Scotece.pdf

B. Li, T. Chen, G.B. Giannakis, Secure mobile edge computing in IoT via collaborative online learning. IEEE Trans. Signal Process. 67(23), 5922–5935 (2019)

Burt, Jeffrey. 2021. The New Stack. The Challenge of Sacaling the Intelligent Edge. Retrieved online from https://thenewstack.io/the-challenge-of-scaling-the-intelligent-edge/

Craven, Connor. 2019. What's the Difference Between Edge Computing and MEC? SDxCentral. Retrieved online from: https://www.sdxcentral.com/edge/definitions/whats-the-difference-between-edge-computing-and-mec/

Dhelim, Sahraoui and Kechadi, Tahar and Chen, Liming and Aung, Nyothiri and Ning, Huansheng and Atzori, Luigi. 2022. Edge-enabled Metaverse: The Convergence of Metaverse and Mobile Edge Computing. Published by: arXiv. Retrieved online from: https://arxiv.org/abs/2205.02764

ETSI GS MEC002. 2018. V2.1.1 Multi-access Edge Computing (MEC) Phase 2: Use Cases and Requirements. Retrieved online from https://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/02.01.01_60/gs_MEC002v020101p.pdf

Finnegan M. Boeing 787s to create half a terabyte of data per flight, says Virgin Atlantic, Computer World, UK. Last accessed July 2019. https://www.computerworlduk.com/data/boeing-787s-create-half-terabyte-of-data-per-flight-says-virgin-atlantic%2D3433595/.

Gao, L., Luan, T. H., Gu, B., Qu, Y., & Xiang, Y. (2021). Privacy-Preserving in Edge Computing. Springer.
K. Cao, Y. Liu, G. Meng and Q. Sun, "An Overview on Edge Computing Research," in IEEE Access, vol. 8, pp. 85714-85728, 2020, doi: 10.1109/ACCESS.2020.2991734.

Longxiang Gao, Tom H. Luan, Bruce Gu, Youyang Qu, Yong Xiang. 2021. Privacy – Preserving in Edge Computing Retrieved online from: https://doi.org/10.1007/978-981-16-2199-4

Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2019). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Transactions on Industrial Informatics, 16(6), 4177-4186.

M. Yahuza et al., "Systematic Review on Security and Privacy Requirements in Edge Computing: State of the Art and Future Research Opportunities," in IEEE Access, vol. 8, pp. 76541-76567, 2020, doi: 10.1109/ACCESS.2020.2989456.

Maheshwari, Sumit & Raychaudhuri, Dipankar & Seskar, Ivan & Bronzino, Francesco. (2018). Scalability and Performance Evaluation of Edge Cloud Systems for Latency Constrained Applications. 10.1109/SEC.2018.00028.

Microsoft. Azure.com. 2022. What is Cloud Computing? A beginners guide. Retrieved online from: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing/#:~:text=Simply%20put%2C%20cloud%20computing%20is,resources%2C%20and%20economies%20of%20scale.

Mohanan, Remya. 2022. What is Edge Computing? Components, Examples and Best Practices. Retrieved online from: https://www.spiceworks.com/tech/edge-computing/articles/what-is-edge-computing/

Nanoprecise.io. 2022. Edmonton, Canada. Edge Computing in Industrial IoT. Retrieved online from: https://nanoprecise.io/edge-computing-in-industrial-iot/

Pathak, Amrita. Geekflare.com. What is Edge Computing and What are its Applications? 2021. Retrieved online from: https://geekflare.com/edge-computing-and-its-applications/

Pratt, Kary K. 2022. Techtarget.com. Top 5 benefits of edge computing for businesses. https://www.techtarget.com/iotagenda/tip/Top-5-benefits-of-edge-computing-for-businesses

Q. Xu, Z. Su, K. Zhang, P. Li, Intelligent cache pollution attacks detection for edge computing enabled mobile social networks. IEEE Trans. Emerg. Top. Comput. Intell. 4(3), 241–252 (2020)

Shi W, Cao J, Zhang Q,Youhuizi L, and Lanyu X. Edge computing: vision and challenges. IEEE Internet of Things Journal. 2016;3(5):637–646.

Van Rijmenam M. 2013. Self-driving cars will create 2 petabytes of data: what are the big data opportunities for the car industry? DataFloq. Retrieved online from: https://datafloq.com/read/self-driving-cars-create-2-petabytes-data-annually/172.

STL Partners. 2017. Edge Data Centres: What and where? Retrieved online from: https://stlpartners.com/articles/edge-computing/edge-data-centres/

Singh, A., Chatterjee, K., & Satapathy, S. C. (2021). An edge based hybrid intrusion detection framework for mobile edge computing. Complex & Intelligent Systems, 1-28.

Shi, Cong., Xu, Xiangyu., Zhang, Tianfang., Walker, Payton., Qu, Yi., Liu, Jian., Saxena, Nitesh., Chen, Yingying., Yu, Jiadi. MobiCom '21: Proceedings of the 27th Annual International Conference on Mobile Computing and Networking. October 2021. Face-Mic: inferring live speech and speaker identity via subtle facial dynamics captured by AR/VR motion sensors. Pages 478–490. Retrieved online from: https://doi.org/10.1145/3447993.3483272

Tucker, Randy. Lumen. 2020. Edge Computing Transforms Businesses With Low Latency, Real-Time Data And Compute Provisioning Retrieved online from: https://blog.lumen.com/edge-computing-transforms-businesses-with-low-latency-real-time-data-and-compute-provisioning/

Tian, Z., Wang, Y., Sun, Y., & Qiu, J. (2020). Location privacy challenges in mobile edge computing: Classification and exploration. IEEE Network, 34(2), 52-56.

Varghese B, Wang N, Barbhuiya S, Kilpatrick P, and Nikolopoulos D. Chal-lenges and opportunities in edge computing. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud); 2016. pp. 20–26

W. Wang, Q. Zhang, Privacy preservation for context sensing on smartphone. IEEE/ACM Trans. Network. 24(6), 3235–3247 (2016)

X. Xu, B. Tang, G. Jiang, X. Liu, Y. Xue, Y. Yuan, Privacy-aware data offloading for mobile devices in edge computing, in 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (2019), pp. 170–175

Yu W, Liang F, He X, et al. A survey on the edge computing for the Internet of things. IEEE Access. 2018;6:6900–6919.

Zeyu, H., Geming, X., Zhaohang, W., & Sen, Y. (2020). Survey on Edge Computing Security. Paper presented at the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE).

Appendix A: Comparison with other types of computing

Edge computing:

Edge computing makes the processing and the storage of data instantaneous by bringing computing systems as close as possible to the device, application, or component that collects or generates data. The computation takes place on the edges.

Cloud computing:

Cloud computing involves the use of hosted services, such as servers, data storage, networking, and software over the internet where the data is stored on physical servers maintained by a Cloud service provider. The computation takes place on the cloud.

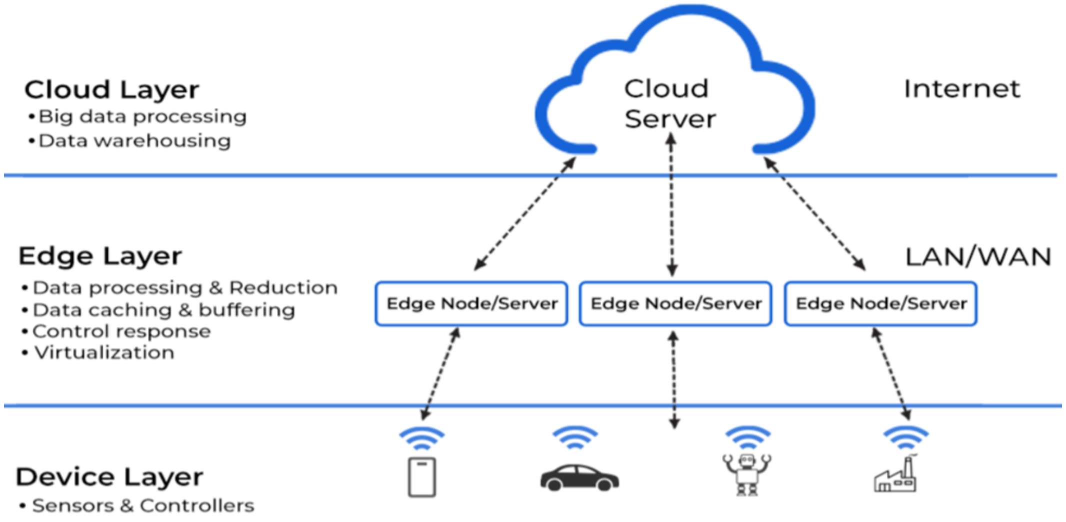


Figure 3: Edge Computing Architecture, Mohanan, 2022.

The major difference between Edge and Cloud computing solutions is where the computations are executed. With Cloud computing, the computation takes place on the Cloud. With the Edge computing approach, the computation takes place on the Edge Layer, closer to the Device layer. (Ashtari, 2022). The charts in the pages that follow have been sourced from Hossein, 2022.

1. Speed & Agility

Edge Computing	**Cloud Computing**
Edge solutions bring their analytical and computational powers as close to the data source as possible.	While traditional Cloud computing setups are unlikely to match the speed of an expertly configured Edge computing network, Cloud computers have their way of exuding agility.
This increases responsiveness and boosts the throughput of the applications hosted on Edge computers. In fact, for certain applications, a well-designed and sufficiently capable Edge platform would be able to outperform Cloud-based systems.	Cloud computing services are generally on-demand for starters and can be accessed through self-service. This means that even vast volumes of computing resources are just a few clicks away and can be deployed by an organization in a matter of minutes.
Edge computing is far more ideal than Cloud platforms for applications that require minimal response times for safe and efficient operations.	Secondly, Cloud platforms give enterprises easy access to various technologies, allowing for agile innovation and the speedy creation of new applications. Any enterprise can access cutting-Edge infrastructure services, incredible computing power, and near-unlimited storage at a moment’s notice.
Machines can leverage Edge computing to mimic the perception speed of a human being, which is immensely helpful for applications such as augmented reality (AR) and autonomous vehicles.	The Cloud gives organizations the freedom to test new ideas, experiment with data, and differentiate user experiences.

2. Scalability

Edge Computing	**Cloud Computing**
In an Edge computing ecosystem, scalability must account for device heterogeneity. This is because different devices come with varying performance levels and energy considerations.	Scalability is one of the key benefits of Cloud computing services. Organizations can easily scale up data storage, network, and processing capabilities using an existing Cloud computing subscription or in-house infrastructure.
Additionally, Edge networks operate in more dynamic conditions when compared to Cloud computers. This means that an Edge network would require reliable infrastructure to ensure robust connections for timely scaling of resources.	Scaling is typically quick and easy and brings with it zero downtime or disruption. Especially in the case of third-party Cloud services, all the infrastructure is in place, and scaling up is as simple as a few extra authorizations by the client.
Finally, security measures on the network can introduce latency in node-to-node communication, decelerating scaling operations.	

3. Productivity & Performance

Edge Computing	**Cloud Computing**
In an Edge network, computing resources are placed in close physical proximity to end-users. This means that client data is processed using analytical tools and AI-powered solutions within a few milliseconds.	Cloud computing removes the need for ‘racking and stacking’, such as setting up hardware and patching software associated with on-site datacenters. This enhances the productivity of IT personnel, allowing them to focus on higher-value tasks.
As such, operational efficiency—one of the critical advantages of this system—is enhanced. This leads to heightened productivity and performance for clients with the proper use case.	Cloud computing vendors also improve organizational performance, boost economies of scale, and minimize network latency for their clients by regularly adopting the latest computing hardware and software.

	Finally, organizations do not have to worry about over-provisioning or falling short of resources due to fluctuating demand levels. By always ensuring the perfect amount of resources, Cloud platforms help ensure near-perfect productivity and performance.

4. Reliability

Edge Computing	**Cloud Computing**
Failover management is crucial for Edge computing services.	Cloud computing is often more reliable than Edge computing.
In a correctly configured Edge network, losing a few nodes does not prevent users from accessing a service at total efficiency. Edge computing vendors also implement redundant infrastructure to ensure recovery from failures and impeccable business continuity.	Due to its centralized nature, data backup, business continuity, and disaster recovery are easier and less expensive in the case of Cloud computing.
Further, systems can be implemented to alert users in case of component failure, thus allowing IT personnel to respond rapidly. However, an Edge computing network is inherently less reliable than a Cloud platform due to its decentralized nature.	Copies of critical data are stored in multiple sites accessed automatically if the closest location is inaccessible. Large Cloud platforms often can continue operations without a hitch, even if an entire data center goes down.
Finally, a key advantage of Edge computing is its ability to operate without access to the internet. This is because Edge computers often rely on LAN connectivity to transmit and process information and only use the internet for transferring data to the Cloud for storage and analytics.	However, Cloud computing needs a strong internet connection on both the server-side and the client-side to operate reliably. Without internet connectivity, the Cloud server cannot communicate with connected endpoints, thus bringing operations to a halt unless continuity measures exist.

5. Security

Edge Computing	**Cloud Computing**
The distributed nature of Edge computing systems has led to a shift in the cybersecurity paradigm typically associated with Cloud computing. This is because Edge computers can transmit data directly between nodes without first communicating with the Cloud.	Cloud computing platforms are inherently more secure due to vendors' and organizations' centralized implementation of cutting-Edge cybersecurity measures.
Such an arrangement calls for Cloud-independent encryption mechanisms that operate on even the most resource-constrained Edge devices. However, this might negatively affect the cybersecurity posture of Edge computers vis-à-vis Cloud networks. As rightly said, a chain is only as strong as its weakest link.	Cloud providers often implement advanced technologies, policies, and controls that enhance their general cybersecurity posture.
However, by restricting the transmission of sensitive data to the Cloud, Edge computing enhances privacy as data is less likely to be intercepted while in motion.	Protecting data is also easier in the case of Cloud platforms due to the widespread adoption of end-to-end encryption protocols.
	Finally, cybersecurity experts implement measures that safeguard Cloud-based infrastructure and applications from potential threats and guide client companies on doing the same.

When to use Edge/Cloud Computing?

Edge Computing is preferable to Cloud Computing in instances such as basic data visualization, basic data analytics, and short-term data historian features. Edge Computing is also best used for data caching, buffering, and streaming. In addition, Edge Computing is advantageous to use for data pre-processing, cleansing, filtering and optimization. Finally, Edge Computing is used for Device-to-Device communications along with some data aggregation:

When it comes to Cloud computing, it is primarily used in areas such as complex analytics, big data mining, business logic sources, and ML rules. Cloud computing also gives an advantage for specific requirements such as advanced visualizations and reporting of data, along with long term data warehousing.

Fog Computing vs. Edge Computing

Fog Computing:

- Fog Node decides whether to process the data from multiple data sources using its own resources or send to the cloud
- It implies distribution of communication, computation and storage resources and services on or close to devices in the control of end-user
- Fog computing pushes intelligence down to the local area network level of the network architecture.
- Fog works with the cloud
- Fog has a hierarchical and flat architecture with several layers that form a network

Edge Computing:

- Each individual edge component plays its role to process data locally rather than sending them towards the cloud
- Edge computing is typically referred to the location where services are instantiated
- Edge computing places the intelligence and power of the edge gateway into the device such as programmable automation controllers
- Edge is defined by the exclusion of cloud
- Edge is often limited to separated nodes that do not form a network

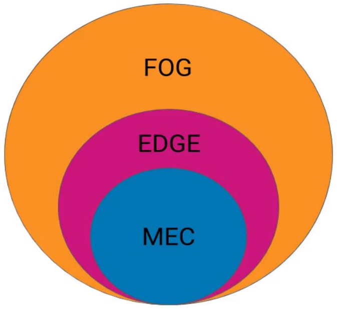


Figure 4. Fog computing is more general than edge computing, and MEC is a more defined set of standards that must be met for a technology to be considered edge computing, Craven, 2019.