Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Information Engineering Department

Master's Degree in ICT for Internet and Multimedia

# Analysis of Latency to Achieve Better Quality of Experience in The Metaverse

Author

**IREM ARTIKASLAN**

2005538

Supervisor

**LEONARDO BADIA**

Academic Year

2022/2023

Graduation Date: 27 February 2023

# DEDICATION PAGE

I would like to dedicate this page to special people who have been with me and throughout my unforgettable journey.

Thanks…

to my family who always encouraged me,

to my friends who made the life more colorful for me,

to my dearest professor, Leonardo Badia, who always supported me from the very beginning to the very end,

to Fabio Benevolo, my company tutor, who believed in me and gave me the opportunity to grow and Lorenzo Valle for his support in realization of this thesis.

*"The biggest battle is the battle against ignorance."*

Mustafa Kemal Atatürk

# ABSTRACT

The metaverse is the emerging 3D-enabled digital space that uses virtual reality, augmented reality, XR and other cutting-edge technologies to enable people to have life-like personal and business experiences online. In recent years businesses started to take the advantage of metaverse capabilities extensively to increase the brand awareness by increasing customer engagement. Aim of this thesis is to analyze the performance of different infrastructures to achieve better quality of experience (QoE) in the metaverse through a business case study and to find the optimal infrastructure setup accordingly. Findings of this thesis revealed that quality of the experience perceived by the end-user/customer is mostly dependent on the latency and the latency in the metaverse is dependent on several aspects. This study focused on the cloud-edge computing and their performance under different conditions by utilizing previous studies in the literature. Consequently, this thesis proposed that the hybrid computing infrastructure to be the most appropriate approach considering the aforementioned company's requirements.

**Key words:** *Metaverse, latency, Quality of Experience, QoE, XR, cloud computing, edge computing, hybrid computing*

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# I. INTRODUCTION

## 1.1. Metaverse

Metaverse is a term which stands for brand-new category of online social network, or arguably the next-generation internet. Although there is no consensus on the definition, it is generally acknowledged that metaverse is based on and incorporates technologies like immersive computing, edge computing, 5G, artificial intelligence (AI), and blockchain. The goal of Metaverse is to give people an immersive experience based on technologies like AR, VR, and XR[1]. Numerous industries, including entertainment, education, public relations, marketing, production, and manufacturing employ metaverse. The range of its application is expanding, and it is anticipated that all aspects of daily life including culture, society, politics, and economy will be accomplished at the same level with reality[2].

The Metaverse, which has existed in practice for a long time and has become a popular concept since 2021, is a combination of the abstract concept of "meta" and "universe" and means a space created by the convergence of virtual and augmented reality. In 1992, the book Snow Crash was the first to introduce the idea of the metaverse. Later, a number of businesses created online communities based on the idea, most notably Second Life, which was introduced in 2003. In the meta-world, people use avatars to represent themselves, communicate with one another, and virtually expand the community. Also, digital currency is being used to purchase clothing, as well as many other goods and items like guns and armor in video games. Millions of people—or their avatars— interacting in real time in a vast virtual universe.

Some industries have quickly entered the metaverse due to the Covid-19 pandemic. Young entrepreneurs are particularly drawn to fields such as medicine, recycling systems, energy efficiency, renewable energy, online systems, robots, and drone technology. Even if the Metaverse falls short of the fantastical visions described by science fiction authors, it is likely to be a new computing platform or content medium that will be worth trillions of dollars. The metaverse is evolving to be the entry point for the majority of digital experiences, a crucial part of all physical experiences, and the next great business platform[3]. Based on the global market for virtual convergence economies, international consulting firm PricewaterhouseCoopers (PwC) expects that the metaverse market will reach $1.5 trillion and 1.81% of global GDP by 2030. Additionally, it is expected that the augmented reality market will develop far faster than virtual reality, reaching $855.3 billion in 2027 [14]. The metaverse market, according to market research firm Strategy Analytics (SA), will increase from $46 billion in 2020 to $280 billion in 2025. The "Virtual Convergence Economy Development Strategy," published in 2020, estimates that by 2025, the manufacturing sector will benefit from the virtual convergence of economies by 194.5 billion dollars, education and training by 90.7 billion dollars, and logistics by 62.2 billion dollars [2].

## 1.2. Metaverse In Business

Worldwide participation to the metaverse is growing among people and organizations. Everyone will have their own avatar and by using these avatars, together with voice and facial expressions, we will be able to participate to the meetings in Metaverse. It will be possible to explore the globe and accomplish a variety of tasks, from shopping to investments. The metaverse offers companies and customers a future where people will live, work, and buy in a virtual world as a next step in this digitalization process. It is possible to make the idea formed within actuality and clarity in the metaverse, a single secure environment where the parties focus on products or services are collaborating with a single idea [3].

Many global brands now make investments in the metaverse. For instance, the process has advanced significantly as a result of Facebook purchasing the virtual reality device developer Oculus for 2.0 G$. Additionally, considerable R&D efforts continue to be made in this area by game platforms like Roblox [4]. Particularly during the Covid-19 global epidemic, the metaverse is becoming a more attractive alternate world than ever for people to invest. Small, medium-sized, or large businesses, as well as everyone, will be impacted by the metaverse [3].

In order to fulfill people's rising expectations for how they want to interact with businesses, the mobile internet has offered a new option for personalized experiences. This trend is anticipated to be driven by the metaverse, and meta technology will be crucial. As with mobile internet, the business opportunity will reflect customer behavior in the Metaverse. It is essential to keep trying and developing in order to construct the business for today's possibilities. Additionally, it offers a mechanism for business owners to swiftly develop and test items to see if they are generating enough interest in mass production. Entrepreneurs now have the ability to quickly modify and improve products in response to immediate market feedback. Virtual experiences, on the other hand, can be used to convince investors to take chances and invest in a prototype or product.

## 1.3. Customer Engagement and Branding

In order to set themselves apart from their rivals on a global scale, technological product manufacturers truly require "branding" as a critical weapon. As a result, businesses that specialize in creating cutting-edge technology products should spend money on branding and develop a successful brand management plan alongside a continuum of "technology management. The internet and digital media have given us countless chances for marketing and brand creation. Every time a product is advertised, it must increase consumer awareness of the brand and the product. A better impression is generally created in the client's mind by good product, design, content, and advertising. As a result, there are higher odds that the customer will select the product. A high percentage of consumers interacting with 3D advertising are more likely to experience a sense of presence. Users of 3D virtual worlds were more influenced by the richness of information when they were highly involved. The metaverse can be utilized for commercial purposes, including advertising. The business sector should focus on finding ways to use the best elements of the metaverse to support business growth and achieve their goal. To improve brand awareness and attract more existing customers, companies have built their own digital worlds on metaverse-like platforms. As an example, the "Gucci Garden" module, a project commissioned in 2021 by Roblox and the famous fashion brand Gucci, which is one of the pre-metaverse virtual game platforms especially accepted by the younger generation, is a very important initiative. With this project, Gucci products created for the virtual platform were able to find buyers for very high prices. This alone shows that the metaverse universe is a new economic market [5].

## 1.4. Motivation

An infrastructure sufficient to support the technologies used in Metaverse is not yet fully available. The path to the Metaverse is filled with numerous obstacles. Hardware limitations are one of the most significant obstacles. Millions of users simultaneously cannot experience an everlasting digital environment at this moment due to the limitations of present global networking and processing capabilities. If hardware, energy, and technology are sufficient, broad cultural changes will be necessary to foster the development of a true metaverse [6]. Furthermore, a virtual environment that renders high-quality graphics for each avatar at a minimum frame rate of 30 frames per second [7] has enormous processing demands and latency limits (e.g., within 1/30th of a second at most), because of this reason, large-scale storage and computation capacities are made available through the storage and computation infrastructure supported by cloud-edge-end computing [8].

Although various businesses hold varying opinions on the metaverse, its arrival cannot be denied. Building a network system that is scalable, safe, trustworthy, and provides a high quality of experience (QoE) is therefore essential for its success. The latest 5G explosion has significantly increased the possibility of creating reliable metaverse systems. The theoretical maximum throughput for 5G is between 10 and 20 Gbps. But because the sensors produce massive amounts of metadata and high-resolution video streams, metaverse has a very high bandwidth requirement. Given the necessity for scalability, 5G could not be able to meet the bandwidth needs of the metaverse [9]. A practical difficulty is ensuring low latency when consumers are spread over geographically dispersed places. The security, usability, and financial components of the metaverse are crucial to the success of this new Internet generation in addition to the increasing demand for network support [1].

To improve experience in the Metaverse, the KPI's must be analyzed to reach the optimal performance. Since this thesis stresses the importance of Metaverse for business, customer engagement and branding, QoE/QoS plays an important role here. In order to maximize QoE, the latency should be minimized. In this paper, latency aspect of Metaverse has been chosen to be deeply investigated.

## 1.5. Outline of the paper

The introductory chapter, Chapter 1 introduces the reader to the metaverse, speaks of its application fields and its importance. Chapter 2 provides background information about metaverse enabling technologies and related concepts to prepare the reader to fully understand the context of the study. Chapter 3 highlights the relevant literature in order to situate the forthcoming research within the extant theoretical paradigms. Chapter 4 proposes a business case to be analyzed together with the comparison of different infrastructure setups, followed by a Conclusion in Chapter 5.

# II. BACKGROUND

## 2.1. Metaverse Architecture

Metaverse is a self-sustaining, hyper spatiotemporal, and 3D immersive virtual shared space, created by the convergence of physically persistent virtual space and virtually enhanced physical reality. In other words, the metaverse is a synthetic world made up of user-controlled avatars, digital objects, virtual environments, and other computer-generated elements where people (represented by avatars) can interact, collaborate, and socialize with one another using any smart device and their virtual identities [10]. The development of AI-XR Metaverse applications will be significantly facilitated by advances in a variety of technologies, including virtual reality (VR), augmented reality (AR), mixed reality (MR), extended reality (XR), artificial intelligence (AI), and 5G/6G communication.

## 2.2. Technologies

In order to understand the metaverse, it is important to be familiar with the related concepts that are explained in this part.

**Virtual Reality (VR):** Virtual reality is a computer-simulated experience that entirely replaces the user's perspective of the real world with one that is either similar to or completely different. The user's senses are deceived into believing they are in a different environment by this device. The term for this is "sense of presence" [11]. By wearing a VR headset, users can join a virtual (computer-simulated) environment and entirely block out the outside world, creating an immersive experience. To give the user the impression that the virtual world is genuine, high fidelity user interaction is essential for achieving immersion in VR. Facebook Oculus and HTC VIVE VR headsets are the most commonly used examples of VR headsets, and they are essential to enter the digital world.

**Augmented Reality (AR):** AR takes a distinct approach to real-world environments; it incorporates digital inputs and virtual features into the real world to improve it. With augmented reality (AR), users can display digital content (text, images, and audio) onto the real world to create an immersive experience. In contrast to VR, augmented reality (AR) can be achieved without specialized hardware (like a headset) by using smartphones, implants, glasses, or contact lenses to overlay digital content on top of the actual world [12].

**Mixed Reality (MR):** The concept of mixed reality encompasses both the actual and virtual worlds to varying degrees. Real world and computer components are combined in MR. The user often wears a head-mounted display. Using cutting-edge sensing and image technology, the user can interact with and manipulate both real-world and virtual objects and surroundings [11]. When the real and virtual worlds come together to create new settings and experiences, physical and digital items co-exist and engage in real-time interactions. This phenomenon is known as mixed reality (MR) (it is an enhanced form of AR). An example of an MR headset is the Microsoft HoloLens [12].

**Extended Reality (XR):** XR is an umbrella term that includes VR, AR, and MR. It addresses every potential future reality that could result from these technologies. By 2022, it is expected that the XR market will reach $209 billion [12]. Designers can test out design ideas and look at design solutions in mixed-reality settings thanks to (XR) applications. Virtual reality and augmented reality (XR) technologies have the potential to improve design outcomes and the design process.

Figure 1 illustrates the reality-virtuality continuum and provides an overview of the XR ecosystem and how VR, AR, and MR relate to the physical and digital worlds.

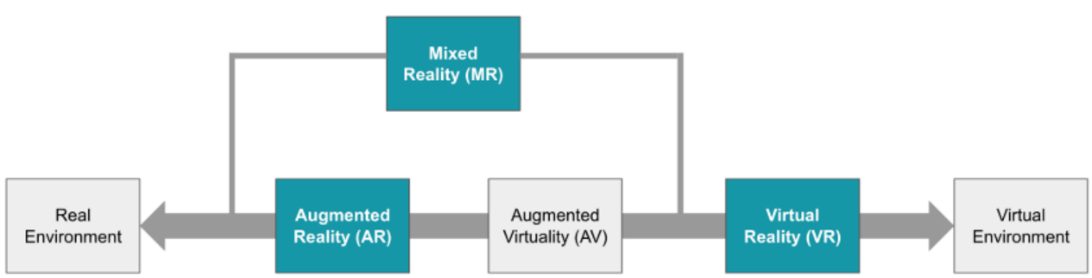


**Figure 1:** Reality–Virtuality continuum [11]

The most important aspect of XR, VR, and AR is to realize immersive first-person experiences. The Metaverse takes this to the next level by enabling huge groups of people to share an immersive first-person experience while yet feeling a strong sense of mutual presence.

### 2.2.1. Metaverse Enabling Technologies

**Interactivity:** As miniaturized sensors, embedded technology, and XR technologies advance, XR gadgets like helmet-mounted displays (HMDs) are anticipated to become the primary entry point into the metaverse [13]. MR offers a transition experience between VR and AR, while VR offers immersive experiences in a virtual world, AR offers true presence experiences with virtual holograms, images, and movies in the real world, and VR provides immersive experiences in virtual worlds. The indoor smart devices help the wearable XR devices execute fine-grained human-specific information perception as well as ubiquitous sensing for objects and surroundings (e.g., cameras) [10].

**Artificial Intelligence (AI):** Artificial intelligence algorithms assist robots in learning from enormous amounts of data, continually adapting to different contexts, and showing human-like behaviors to assist people with daily tasks. In order to keep the metaverse's degree of self-management high, the use of AI-enabled technologies like as machine learning (ML), deep learning (DL), and federated learning (FL) is crucial. The movement and features of both real world and virtual world items must be extremely precise since the transactions and interactions done in the metaverse are captured using XR devices. Only AI-enabled algorithms can make this possible [14].

**Digital Twins (DT):** Digital Twin paradigm has recently come into existence to allow for the depiction of a digital mirror for a physical entity. The term "digital twin" refers to the digital clone of physical items and systems. Due to the continuous enhanced processing and real-time data collection, the twins evolve synchronously during the physical entity's lifetime and can evolve independently in the metaverse. This can only be accomplished utilizing DT techniques supported by AI that intelligently clone the data and do data analysis [14].

**Networking:** In the Metaverse, the network plays a crucial part in ensuring high transmission capacity, accuracy, and the shortest possible delay for the enormous amounts of diverse data. By the virtue of advancements in next-generation networks (NGN), especially in 6G communication network, the user experience for immersive metaverse services and applications will dramatically improve [14]. Network latency is important to the quality of experience in fact low latency is very important for preventing motion sickness. A practical difficulty is ensuring minimal latency when consumers are in geographically dispersed places. The success of this new generation of the Internet depends not only on the increased need for network support but also on the security, accessibility, and financial aspects of the metaverse [1].

**Ubiquitous Computing:** The goal of ubiquitous computing, often known as ubicomp, is to provide consumers access to computing anywhere and at any time. With ubicomp, human users can freely engage with their avatars and experience real-time immersive metaverse services via ubiquitous smart items and network connectivity in the environment rather than using specialized equipment (like a laptop) [10]. In ubicomp, the cloud-edge-end computing orchestrates heterogeneous edge computing infrastructures (closer to end users/devices) and highly scalable cloud infrastructures (with powerful processing and storage capacity) for increased user quality-of-experience (QoE) [8].

**Blockchain:** Decentralized architecture should be used to build the metaverse in order to reduce centralization issues including SPoF (single point failure), low transparency, and controlled by a small number of entities [15]. Blockchain is a distributed ledger that possesses the characteristics of decentralization, immutability, transparency, and auditability. Data is organized into blocks with hash chains. NFT stands for irreplaceable and indivisible tokens, which can support ownership provenance and asset identification in the blockchain.

## 2.3. QoE, QoS, Performance, Latency

The concepts explained below will make the reader understand better the context of this research and which concepts are under investigation.

There are only few studies in the literature that looked into how to assess the Quality of Experience (QoE) in the Metaverse since it is a novel technology. However, the QoE definition published by the Qualinet White Paper can be deemed acceptable also for Metaverse applications as the beginning point for the study, similar to other multimedia-based apps: “The QoE is the degree of delight or annoyance of the user of an application or service which involves an immersive media experience” [16]. The primary elements that could affect the users' quality of experience (QoE) with a given application or service are identified in this White Paper as the human and system influence factors (IFs).

**Human influence factors:** Takes into account both static (such as sex and age) and dynamic (such as affective state) data about the user.

**System influence factors:** The hardware, software, and multimedia content are all covered by the system IFs. Supporting the Metaverse applications requires a strong network infrastructure as a baseline, especially when high-resolution multimedia material is involved [17].

Network quality is more difficult to understand and measure than one may imagine. To begin with, there is no consensus about network quality. It is assessed in a variety of ways and in different ways depending on the network technology. It is often impossible to compare these metrics. But the issues do not end there. Bandwidth, latency, packet loss, and fluctuations are just a few of the numerous aspects of network quality, making it challenging to define what "improved network quality" actually implies [18].

There are numerous methods for assessing network quality, which vary depending on:

- Different technologies (4G, 5G, Wi-Fi, fiber, cable, etc., and their generations and related bands of operation)

- OSI levels and the nodes hosting the functions that are associated with each layer.

- The potential end user's equipment and network hardware.

**Challenges:**

- Knowing the user and their equipment and using indirect measurements is necessary to assess network performance from the perspective of the end user.
- Guarantees of Quality of Service (QoS) based on presumed performance expectations might not quite correspond to Quality of Experience (QoE).
- Since network quality can be measured along many different axes, it might be challenging to define what "better" actually implies [19].
- Applications have complicated contextual characteristics, and a given specific application heavily influences network configuration for optimum performance and heavily related with resource allocation policies of the operators [20].

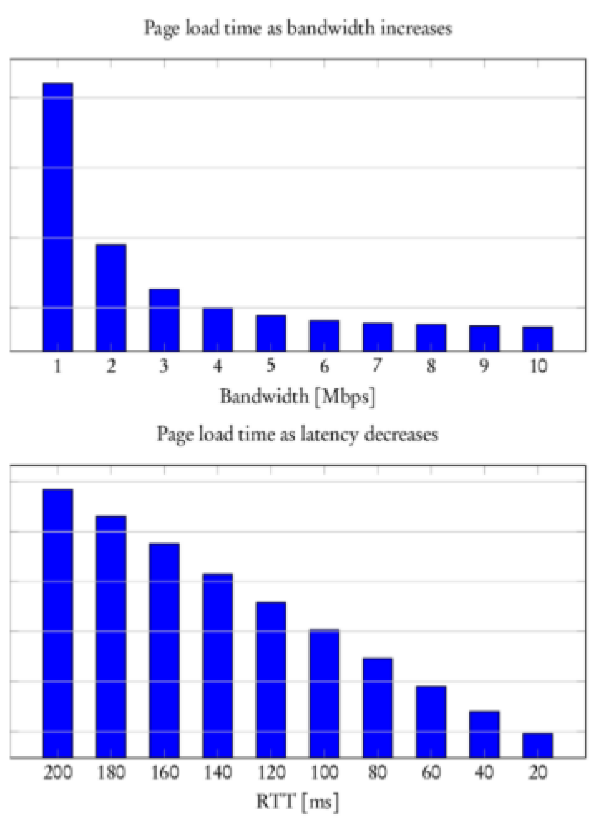


**Figure 2:** Correlation of Page Load Time (PLT) with bandwith and latency [18]

According to High Performance Browser Networking, a website's Page Load Time (PLT) is a function of bandwidth (top graph) and latency (bottom graph) as indicated in the Figure 2. These graphs show that boosting bandwidth above 8 Mbps yields almost little benefits. On the other hand, decreasing the latency, also called jitter, linearly decreases PLT. It is simple to optimize a network in a way that increases throughput (Mbps), but it might result in worse user experience related to higher latency or packet loss. There is no certainty that increasing bandwidth will reduce latency unless bandwidth is increased simultaneously at every location, from beginning to finish. Insufficient bandwidth results in queues, which result in latency, which may result in packet loss. “But not all packet loss is caused by latency and not all latency is caused by insufficient bandwidth” [18]. Only when the network's quality is enhanced across the board, including its statistical components, we can say with certainty that its quality is "better". You may, for instance, increase throughput, packet loss, and average latency while lowering quality of experience (by increasing jitter). In general, the application being utilized determines if a network has "higher" quality. In this sense, "better" refers to superior application results, such as online pages that load more quickly, video conferencing that are lag-free, and streaming in higher definition [18].

The creation of multimedia packets involves an incremental encoding process that makes use of the inherent spatial and temporal redundancy found in the raw data. Error correlation has a significant impact on performance thus on experienced quality, in [21], it is considered a transmission over a Markov channel where it is set for average error burst size as well as error probability. This makes it possible to derive a number of performance measurements fully analytically using Markov analysis.

The methodology covered in this letter [22] uses a Markov approach to assess the delay terms' statistical data with the aim of creating a framework that brings together several, previously unexplored areas is the main objective. They examine the impact of error correlation in the channel, and they demonstrate that when the channel is somewhat correlated, the delivery delay may actually reduce with an increasing arrival rate. Therefore, error correlation may occasionally indicate a general reduction in the total latency. To accurately estimate the delay in real-time multimedia services over wireless, these factors are remarkable.

A constant stream of data packets is typically taken into account in traditional investigations. A carefully crafted model is also needed since video transmission presents the non-trivial difficulty of the packets' varying length and, more significantly, their incremental encoding. The framework suggested in [23] can be utilized as a useful tool to comprehend how error control methods used for wireless video transmission behave and ultimately develop design principles for such systems. Its modular design makes it possible to quantitatively compare various packet formats and error control strategies for every relevant scenario.

Recently, as another point of view in the academic world, Age of Information is being discussed as a more effective performance indicator than throughput or delay to assess the effectiveness of medium access strategies, particularly when used in distant sensing applications and more generally the Internet of Things [24].

As indicated above, there is not only one absolute solution to increase the Quality of Experience of the end user because of some several reasons. In the light of this information, this study will mainly focus finding ways to decrease the latency to obtain better Quality of Experience which the business will benefit from, to establish stronger customer engagement through Metaverse.

# III. LITERATURE

To overcome the latency issue and achieve better Quality of Experience, studies mainly focuses on computation and communication algorithms and scenarios on the technologies used in realization of the Metaverse. Although, there are also other aspects that can effect quality of experience which are called influence factors (IF) including acceptability, cybersickness, quality and presence [17]. In this chapter, the literature review takes place as a state-of-art, first the studies related to QoE evaluation will be explained then studies related to latency, more specifically cloud computing, edge computing and hybrid computing, will be reviewed.

## 3.1. QoE studies

There are studies discusses the QoE and ways of measuring it. In [25], the QoE of adaptive PC(point cloud) streaming was examined with various network setups. However, there are currently only a few acknowledged flaws in the present objective quality assessment methods [26], which drives the need for more effective algorithms. The quality of experience for VR broadcasting is complicated. According to [17], the quality of experience (QoE) of VR streaming is correlated with a number of characteristics, including presence, usability, and cyber-sickness. The study [27] posited that the user's QoE may be significantly impacted by the social and economic aspects of the Metaverse. Users' interactions and behaviors with other individuals are considered social characteristics, as is their standing in the Metaverse. Economic considerations include the need for resources to buy items and employment in the Metaverse.

In [28], the network architecture of the Metaverse is addressed, and it is found that offloading computations (such as multimedia rendering) for Metaverse apps to the edge network greatly improves performance and the quality of experience (QoE) for the end user. It is anticipated that multimedia applications would consume the majority of the bandwidth in future communication networks. Therefore, scheduling systems that can handle multiple forms of multimedia communication (conversation, interactive, streaming video, etc.) and provide sufficient Quality of Experience in accordance with application needs are required. It is proposed an implementation scenario of a multimedia-aware scheduler in [29] which manages two priority classes, for GBR(Guaranteed Bit Rate) and NGBR(Non-Guaranteed Bit Rate) traffic.

The idea to measure the QoE in study [30] states that, a video application on a mobile network may be delivered via various channels, producing varied QoE, which is then monitored at the end user and reported back to the network, sometimes with fine granularity to monitor user mobility and channel changes. As a result, the mobile operator is responsible for directing the service to maintain the desired QoE and to maximize the use of network resources. The massive data from new applications significantly increases the processing latency of conventional communication networks, requiring a rethink of the classical information theory (CIT) driven communication networks in order to achieve the user-centric QoE metrics necessary for the successful implementation of Metaverse [31].

When we dive into IoT, a lot of work has been put into network planning, conceptualization, and execution [32], yet there are also a number of issues with experience design, as well as an apparent lack of standards for assessing the effectiveness and usability of IoT devices [33].

A higher emphasis is placed on the Quality of Service (QoS) as a result of the invisibility of smart systems, whether it be through the invisibility of interactions or the distributed nature of technology [34]. Study [35] states that there is no guarantee on the achievable QoS, but the performance metrics can be improved.

The only way to ensure portability, scalability, and fairness—which lead to overall user satisfaction only with the inclusion of QoS-aware efficient protocols—is through a strong integration of various access strategies. In [36], the authors address realistic techniques to successfully provide a sufficient QoS in heterogeneous wireless networks as they analyze routing solutions to give coverage extension in those networks. They argue that periodic AP(access point) advertising help network management understand the quality that can be sustained. They also address route advertisement selection methods and offer a utility-based method for modeling the QoS of routing paths that is easily adaptable to diverse goals.

Furthermore, they present a backward utility framework that could further increase the realism of the model by accounting for dynamic changes in the given QoS.

## 3.2. Latency Studies

The majority of computations in the cloud computing paradigm take place in the centralized cloud, where data and requests are also processed. Although the user experience may be compromised by longer latency (such as long tail latency), such a computer paradigm may experience it. Numerous studies have examined the energy-performance trade-off of cloud offloading in a mobile-cloud context [37], [38]. Cloud computing is arguably now the standard computing platform for large data processing in business and academics. One of the main promises of cloud computing is that data would eventually be handled in the cloud, whether it currently exists there or is being sent there. However, due to privacy concerns and the high cost of data transfer, stakeholders don't frequently share the data that they own with one another. Therefore, the likelihood of cooperation amongst many stakeholders is low [39]. It is also possible for the logical model to include the edge, a physically small data center that connects the cloud and end user with data processing capability. The concept of the collaborative edge has been put out in [40], which links the edges of numerous stakeholders who are geographically dispersed.

In the age of smart living, the IoT-cloud-based augmented reality framework for home automation system that is proposed in [41] is a novel and generic framework that allows bi-directional augmented reality data flow by providing low latency home appliance control and sensor data collecting. The proposed layered and modular approach makes this system more generic and customizable. In contrast to conventional systems, where users must control devices using a list or console, this AR-based platform offers users an extra layer of convenience by providing a camera-based AR interface to point at any household item and control it via the interactive on display control panel. From object detection to device control, the system is intended to have as little latency as possible.

Early datacenter deployment in the cloud was limited, with locations far from end users and high end-to-end communication latency. Later, cloud datacenters, became more widely dispersed geographically, and the networks' bandwidth kept growing, lowering end-user latency. A research on extensive global client-to-cloud latency measurements toward 189 datacenters from all major cloud providers is conducted in article [42] , the authors give a thorough assessment of cloud reachability in this article. They claim that in year 2009 when this article was published, their findings show that the majority of the world's population can readily access many latency-critical applications, such as cloud gaming, with the existing cloud coverage.

The timely delivery of resource-intensive and latency-sensitive services (such as industrial automation, augmented reality) over distributed computing networks (such as mobile edge computing) is gaining increasing attention, so the authors introduce a novel queuing system that is capable of tracking data packets' lifetime and formalize the optimal cloud network control problem with strict deadline constraints in [43]. They focus on the crucial goal of delivering next-generation real-time services ahead of related deadlines on a per-packet basis while lowering total cloud network resource cost, which is also prompted by the inadequacy of average delay performance assurances provided by prior studies.

Some IoT applications may need instantaneous responses, may contain sensitive information, or may generate vast amounts of data that could put a strain on networks. These applications cannot be supported by cloud computing due to its inefficiency [39].

Edge computing brings computational and memory capability closer to the point of use. Edge computing lowers the computational load on the initial cloud computing center, relieves the demand on network bandwidth, and increases the efficiency of data processing by moving some or all calculation work to the network edge devices [44].

A common network for edge computing is the Content Delivery Network (CDN). A CDN is made up of geographically dispersed proxy servers and the data centers that host them. It tries to deliver high performance by dispersing the service spatially in relation to end consumers [44]. For the purpose of forecasting CDN service performance, a general AI-defined attention network was used in [44], the model predicts CDN performance via server-side monitoring data, in contrast to earlier studies that primarily concentrate on the client-side user QoE and it is capable of providing precise performance forecasts for all CDN cache groups while preserving a minimal training overhead. An analysis that makes recommendations for the placement of general-purpose edge computing resources across the Internet is presented in [45] and found that network operators are excellent candidates for market dominance in edge computing since they are heavily present in the network. However, cloud providers have already considerably impacted ISP networks, leaving network operators with little room for deployment. In their previous works [42], [46] , it is also revealed that the latencies in Europe and Oceania are similar to those in the US, also demonstrate that latencies to the cloud are much greater in Asia, Latin America, and Africa, making those countries more desirable as deployment locations.

When Internet content providers (ICPs) choose which CDN to use for their content distribution, one key consideration is the statistical latency between the set of servers in the CDN and the set of objective user [47], the statistical latency can be used by Internet service providers (ISPs) to enhance network performance, including identifying connection faults and optimizing routing schemes. Measuring the statistical latencies can benefit these scenarios, in [48] proposed the DMS(DNS-based statistical latency Measurement platform at Scale) framework, a statistical delay measurement system which can lower the relative error by 33% for statistical latency prediction and 18.5% for real-time end-to-end latency prediction.

With the advent of 5G networks, mobile edge computing (MEC) promises to significantly reduce network latency (down to 1 ms) by placing apps closer to consumers at the network edge, opening the door for web AR performance enhancement, the authors of research [49] propose a web AR service to make up for the lack of processing capabilities and to alleviate the new limitations brought on by cloud computing, especially with regard to latency and bandwidth providing a framework with MEC.

A powerful computing infrastructure is proposed in [50] using the MEC-based URLLC digital twin architecture, which looks into task offloading and task caching strategies on neighboring edge servers. Regarding to reliability and latency, the suggested technique can enhance the digital twin's quality-of-experience for metaverse applications. The deployment of Base Stations (BSs) and MEC Points of Presence (PoPs) are determined mathematically in article [51] utilizing innovative point processes that take into account both the population density and the minimum distances between BSs. The model is used to generate 5G gNodes B and MEC deployments that meet the strictest 5G latency constraint of Ultra-Reliable and Low Latency Communications (URLLC) slices and can subsequently support future Augmented Reality/Virtual Reality (AR/VR) and low latency streaming services. The model is applied in real urban, industrial, and rural scenarios.

In paper [52], an opportunistic approach, a distributed edge cloud platform with a disaggregated strategy is proposed. By the third-party companies such as tel-co providers, network operators or specialized edge providers, edge resources are immediately deployed, configured, and enrolled in a cloud management platform after being leased from existing bare-metal cloud providers, giving customers uniform access to this segmented resource pool.

In addition, there are developed systems aiming to decrease the latency. Marvel is a mobile augmented reality (MAR) system that operates on typical mobile devices and achieves a notation display service with imperceptible latency (100 ms) and low energy consumption. Marvel uses a mobile device's local inertial sensors primarily for object recognition and tracking while computing local optical flow and offloading images only when necessary, in contrast to conventional MAR systems that recognize objects using image-based computations carried out in the cloud [53].

In paper [54], an edge-cloud based and low-latency, mobile augmented reality system called Jaguar with flexible object tracking is designed, constructed and tested. Jaguar's ability to cut end-to-end latency to 33 ms is proved through a prototype implementation of the suggested comprehensive solution.

An innovative three-tier distributed software system called L3BOU lowers average end-to-end latency and cloud-edge bandwidth in the backhaul network for 360-degree video streaming applications. By utilizing edge-based, efficient upscaling strategies, the L3BOU architecture achieves low bandwidth and low latency [55].

In study [56], a consistent and comparable method of calculating the latency in a video see-through AR system was developed. By encoding the time in the image and decoding the time following camera feedback, the latency is estimated. It has the ability to totally automatically extract latencies.

In study [57], multi-user AR applications' communication and processing difficulties are investigated and discovered that pre-made AR apps have poor virtual item placement between users and across time as well as excessive communication latency. In addition to a quantitative approach to calculating these positioning changes, the ways for effective data exchanges amongst AR users to reduce latency while preserving accurate placing of the virtual objects are presented.

Applications for augmented reality demand high levels of computing and have latency requirements between 15 and 20 milliseconds. Fog computing meets these needs by bringing the computational resources closer to the augmented reality devices, resulting in lower latency and on-demand computing capacity [58]. Cloud infrastructure and fog computing operate independently of one another. Fog computing uses devices with computational power for data processing, such as routers, base stations, smartphones with multiple cores, etc. This computing paradigm is situated between the cloud servers and IoT devices. To deliver services more quickly and effectively, fog computing also makes use of free computing resources close to the users. During peak hours, these machines are actively processing data; otherwise, they are idle. However, for more complicated calculations, fog computing can still use cloud servers. A potential paradigm known as fog computing architecture proposes the execution of tasks across hierarchical processing levels by evenly distributing them to satisfy application requirements [59]. For instance, high-frequency operations with minimal latency requirements can be carried out close to end devices, but huge data analysis tasks can be carried out at the cloud layer. This hierarchical architecture allows for reduced latency program execution while also utilizing less network capacity[58].

Authors of [60] developed a hybrid fog-edge computing architecture for Metaverse applications that makes advantage of edge devices to carry out the necessary computational load for demanding activities like virtual world collision detection and computation of 3D physics. A virtual entity's associated computational chores are carried out at the end-device of that physical entity. In comparison to the legacy cloud-based Metaverse applications, simulation results indicated that the suggested architecture can cut visualization latency by 50%.

Simple offloading to other computers has been tried, but it has shown to be impractical because of the high communication latency and poor user experience. In paper [61], the authors suggest a novel Mobile Edge Computing framework for Augmented Reality applications (MEC-AR). This framework is envisioned as having three layers: the end user, the mobile edge, and the cloud. The investigations and computer simulations reveal that the MEC-AR framework is more effective than the other traditional schemes as a result of the cooperation between the cloud and edge layer. A greater Quality of Experience (QoE), including excellent image resolution, extended battery life, convenient portability, and real-time performance, can be achieved by drastically reducing both processing delay and CPU energy usage.

In [62], a two-tier architecture for data classification is proposed where the reference edge servers(ES) and individual devices both have decision-making skills, but with varying degrees of precision. Investigation is being done on a coordinated strategy that combines offloading the most important data to the ES while still applying a domain classification to the vast majority of the data for local processing at the (mobile devices) MDs. The study demonstrated that the two recommended approaches can work in concert to achieve nearly 100% accuracy with minimal offloading and a sensible selection of local classification domains. This supports the use of such an architecture for effective decision-making in environments that require a lot of data.

Despite the fact that edge and cloud devices process data more quickly than mobile devices, simultaneous transmission of massive data streams may overwhelm the local wireless network, increasing the overall delay. In a scenario where mobile users share the same network resource but lack a priori knowledge of the wireless links, a game-theoretic framework for distributed decision-making is proposed in[63]. The analysis resulted with this finding that that a Bayesian scenario, i.e., one with incomplete information, really delivers better utility (or lower social cost) than one in which consumers have complete knowledge. This is a result of the game's competitive nature, which encourages selfish rational users to attempt computational job offloading even when doing so is not advantageous to the network as a whole.

# IV. CASE STUDY

## 4.1. Business Case

Vertiv Co which is an Ohio, USA based international company, provider of equipment and services for data centers launched its mobile application XR App in 2021 and web based Virtual Showroom in 2022. According to TechRepublic, 2020, 91% of business organizations are already leveraging or planning to adopt VR or AR technology and according to Gartner, 2022, by 2026, 30% of B2B sales cycles will be managed through digital sales rooms, which will then be used to manage the customer life cycle and buyers increasingly want a seller-free experience and to drive their own unique buying journeys. They are motivated by the intensive use of technological advancements in the industry in the latest years to digitalize their customer journey through unique, meaningful, and relevant immersive digital experiences, to enhance the brand perception and improve the demand generation outcomes. Vertiv XR is publicly available, free of charge, for anyone to immerse in the exploration of their solutions in Augmented Reality internationally, also the Virtual Showroom in Metaverse world revolves around their customers' needs and built upon their integrated portfolio, enabling a new way for their communities to get to know and interact with them lively.


**Figure 3:** Reception of Vertiv Virtual Showroom where the customers are welcomed

### 4.1.1. The XR App

The XR app was built for 2 main purposes: (1) to bridge the imagination gap and provide a more immersive and engaging brand discovery and product exploration in the early stages of the relationship between their prospects and their brand, and (2) to equip their marketing, sales, product offering, leadership, and partners with cutting-edge technology that can provide them with the competitive advantage they need to overcome their rivals.

### 4.1.2. The Virtual Showroom

The exploration of the products takes place by entering a 3D virtual world. The customer, in proximity to a product, has the possibility to walk around it, to interact with it from the simple callout that displays textual information / drawings or videos to animations that show its operation. Other effects such as aerial view or X-ray view facilitate the focus on the most important product components.

The experience within the virtual world is as human as possible, each visitor is represented by their own avatar which will enable the possibility of interacting with all the other participants via voice, chat, or gesture. The event host has the option of teleporting guests to the desired area or temporarily activating the guided mode in which guests will “forced" follow the host for introduction and presentation.

Salespeople, marketing experts, offering managers create their own meetings or events through the Vertiv Virtual Showroom planner that will define the details of the event from title to schedule. It is also possible to customize the environment based on the event by the possibility of co-branding or selecting some types of content to be displayed (videos, presentations or rollups) that will be embedded within physical objects in the environment.

**CUSTOMER JOURNEY**

Unaware | Aware | Interested | Committed | Purchased

| *Area* | *XR App* | *Virtual Showroom* |
|---|---|---|
| *Awareness* | Deliver an impactful brand discovery experience, engage, and delight audiences, position themselves as leaders in their space, stay top-of-mind for their prospects. | Enable users to get to know the brand and discover capabilities in a human-centric manner, through synchronized virtual experiences in an immersive world built around their needs that brings the meaningful physical interactions from events, meetings, or customer sessions into the digital space. |
| *Interest/ Education* | Reinvent the immersive experience designed to boost interest in their capabilities, digital journey that supports through leadership, product launches, demand creation, reachability where and when the users choose. | Make the education experience memorable, when it comes to products and services, through life-sized 3D models that provide users with the chance to interact with them like they would do in the physical world. |
| *Decision/ Purchase* | Leverage augmented reality to enable digital relationship building, transform unknown users into engaged prospects and enhance sales engagement, immersive mutual experiences designed to amplify the Halo effect & emotionally convince their audiences, reveal previously unseen system features that can close the deal for the company. | Ease the users' paths towards a purchase decision with multi-sensory, shared experiences hosted in a tailored environment that will enable user to redefine the project definition and providing engaged users with the reassurance that their solutions and their integration are the answers the customers seek. |
| *Post-sale/ Advocacy* | Use the extended reality capabilities of XR and Virtual Showroom to keep the customer engaged after the sale, showcase project progress, introduce complementary products/components/ services for cross and upselling, allow instant share of experience with their colleagues. | |

**Table 1:** Detailed expectations and outcomes that the company and its customers expected to have with utilization of XR App and Virtual Showroom

**ADDED VALUES**

- Immense brand equity & fitness benefits for long-term growth plus a strong impression that will certainly result in companies lining up for on-demand demos & Marketing Qualified Leads (MQLs) as a result of these sessions in Virtual Showroom.

- Win rates improvement by delivering the most immersive project definition out there and leveraging the hallo effect to influence and speed-up decision making.

- Increase the impact they are having in any markets, with product launches that inspire & amaze, then immediately take advantage of the momentum in one virtual environment, at a fraction of the costs.

- Provide sales with next-generation tools that come with the competitive advantages they need to stand out from the crowd and deliver a spotless customer experience even after the sale was made.

To satisfy the needs of this business case, this study proposes the use of hybrid computation to minimize the latency and achieve better QoE since the main target is increasing the customer experience to increase customer engagement with the brand, products and services. To this end the company needs a continuous experience in the metaverse with minimum interruptions as possible to maintain the customers' attention and involvement to be able to interact with them and provide them a successful experience. In order to manage this, the company required to have a good infrastructure background and optimized ways to decrease the latency as much as possible. The proposal is to use hybrid computing both involving cloud and edge computation to benefit the advantages of both. This hybrid approach is recommended within consideration of this app's wide and geographically dispersed and spread usage environment through continents.

This thesis conducted to find the best approach to fit the needs of this business case. In the rest of this section, there is comparison of the two studies previously carried out to show why and how hybrid approach performs better rather than using cloud-only or edge-only computation for latency-sensitive applications. There are two major considerations; latency and coverage.

Article [64] suggests that one way to define the interactive response time T can be as below to understand the latency experienced by the user,

$$T = t_{\text{client}} + \overbrace{t_{\text{access}} + t_{\text{isp}} + t_{\text{transit}} + t_{\text{datacenter}}}^{t_{\text{network}}} + t_{\text{server}}$$

- $t_{client}$ playout delay; this is the time used by the client to 1) receive and play the video and 2) provide action information. It is solely the responsibility of the client's hardware.

- $t_{server}$ processing delay; this describes the length of time required by the server to process the information received from the client, create the necessary video information, and send the information back to the client. This latency varies from 10 ms to more than 30 ms. The processing delay is the provider's responsibility.

Only changing the hardware will allow playout and processing delays to be decreased. The playout and processing delays for this study are optimistically estimated to be 20 ms; nevertheless, this estimate can vary in reality. The 80 ms threshold network latency for interactive applications is identified by deducting the 20 ms playout and processing delay from the intended latency 100 ms defined in [65]. The network makes up the final portion of the overall delay. 100 ms latency can change according to need and demand of each interactive application.

- $t_{access}$; the time it takes for data to travel from the client's device to the first router connected to the Internet.

- $t_{isp}$; is the of time it takes for a signal to travel from the access router to the peering point linking the ISP network to the transit network's next hop.

- $t_{transit}$; the time it takes from the first peering point to the datacenter's front-end server.

- $t_{datacenter}$; is the time it takes for data to travel between the client hosting server and the datacenter's front-end server.

## 4.2. Cloud-only Infrastructure

Measurement experiments have undertaken to assess the performance and latency of cloud gaming services on existing cloud infrastructures in the US in order to ascertain the capability of the current cloud to deliver the on-demand gaming service [64].

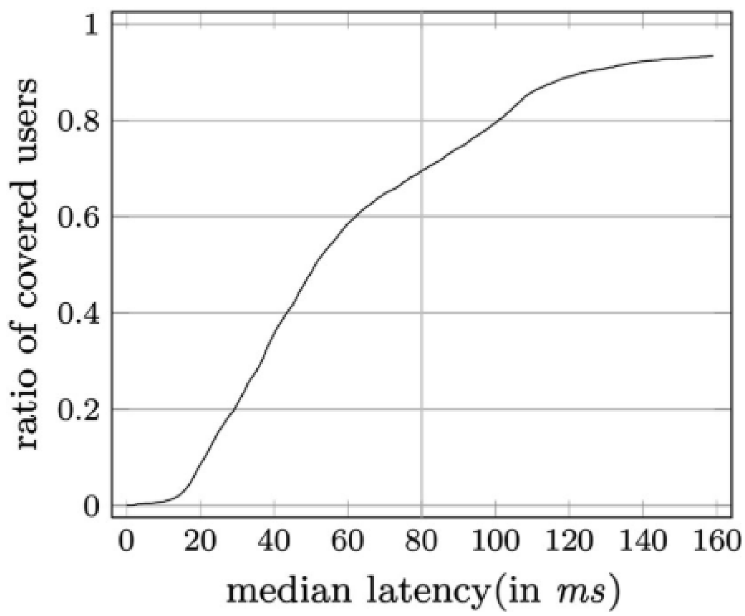


**Figure 4:** Population covered by cloud infrastructure as a function of the median latency [64]

Figure 4 shows, for a particular latency objective, the proportion of covered end users with at least one network connection to each 3 chosen datacenters. The network latency barrier of 80 ms yielding in a 70% coverage. Nearly 10% of potential customers are practically unreachable.

Two strategies designed for deciding the location of datacenters: Latency-based and Region-based.

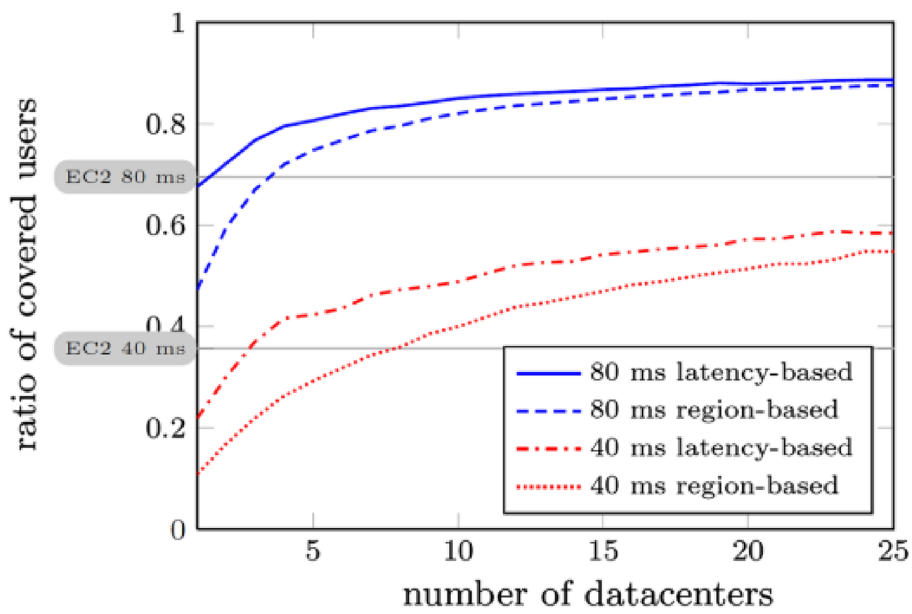


**Figure 5:** Coverage vs. the number of deployed datacenters [64]

Figure 5 shows the proportion of users covered as a function of the target response time for two target network latencies: 80 ms, which allows for reasonable response times for action games, and 40 ms for even more demanding games. Users in that region may profit from the deployment of more datacenters in the Southern US. While geographically dispersed across several locations, including the Southern US, a larger cloud infrastructure is still unable to meet the 80 ms target for a sizeable portion of the end-user population. Consequently, a larger deployment might not be enough to give all end users an acceptable level of delay [64].

Although the end-user coverage offered by employing current cloud datacenters is intolerable, there is a chance that new cloud operators will enter the market. A large number of smaller datacenters can be used as an alternative to establishing a small number of large datacenters. Cloud providers should carefully assess if it is economically advantageous to create a new datacenter because a large datacenter is typically more cost-efficient than a small datacenter.

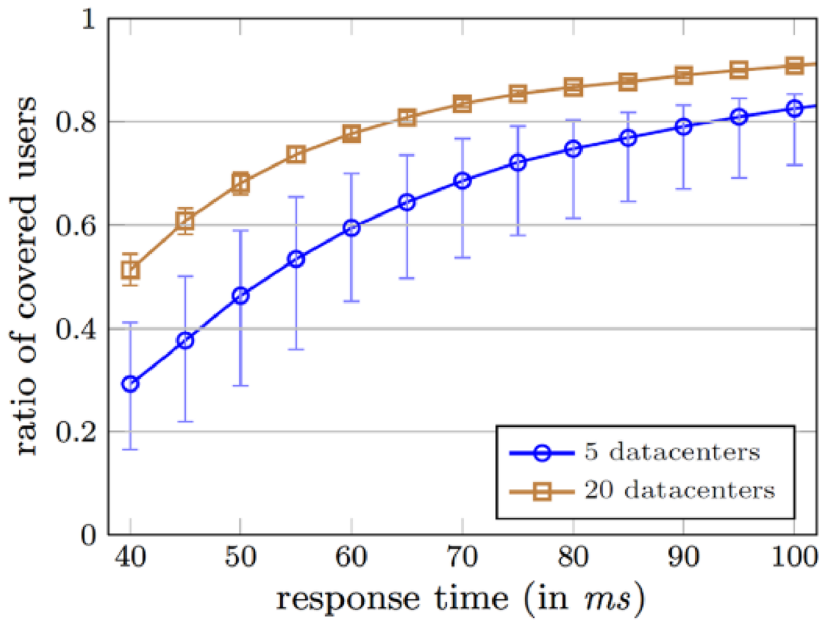


**Figure 6:** User coverage for a region-based datacenter location strategy (average with min and max from every possible set of locations) [64]

It has been noted that performance gaps between a deployment of 5 and 20 datacenters might be quite large. Moreover, even if it is anticipated that a location strategy based on regions will provide adequate coverage, five datacenters do not ensure optimal performance. 80% coverage can typically be attained for 80 ms with a carefully considered deployment of five datacenters. A bad 0.6 coverage ratio can be caused by a poorly planned five-datacenter implementation. In comparison, a deployment of 20 data centers shows negligible variations in the coverage ratio.

## 4.3. Edge-only Infrastructure

A pure edge-server deployment using CDN servers, which are co-located at ISPs, is an alternative to using datacenters. Despite being close to end users, edge servers can only provide services to a certain number of end users.

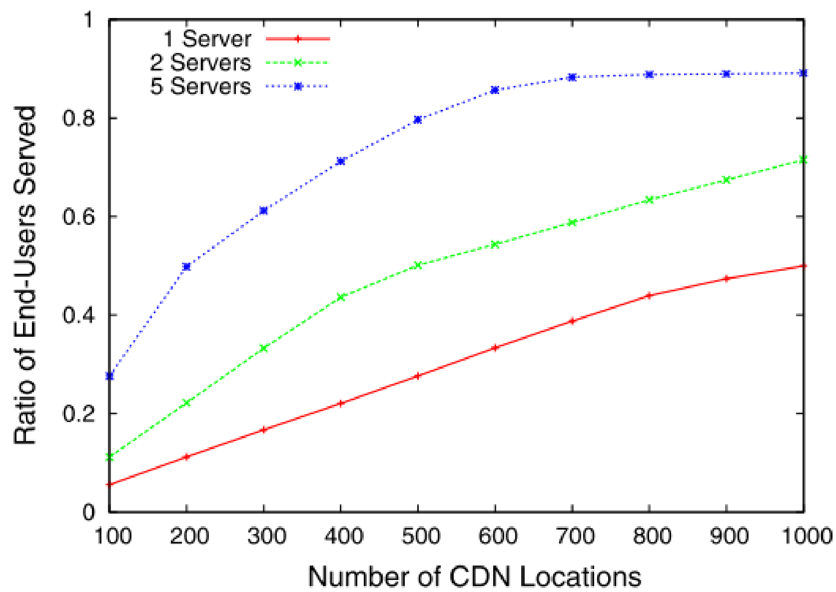


**Figure 7**: Ratio of end-users served versus the number of CDN locations for the 80 ms latency target. The number of servers at each CDN locations varies, and each server hosts 5 out of 100 apps. There are 1,500 clients [64]

According to the simulations, a significant number of edge sites are required to serve 50% of end users with an edge-only deployment. It is found that at least 1,000 edge servers are needed to serve half of the end users, as shown in Figure 7, when choosing edge sites and allocating non-static content randomly (200 edge sites with 5 servers each, or 500 edge sites with 2 servers each) [64].

A significant number of edge sites are required, according to the simulations, in order to serve 50% of end users with an edge-only deployment. As shown in Fig. 6, while choosing edge sites and allocating apps randomly, at least 1,000 edge servers are needed to serve half of the end users (200 edge sites with 5 servers each or 500 edge sites with 2 servers each). Only 70% of end users can be served in a deployment made up of 400 edge sites, each with five servers, which is comparable to a deployment that uses the cloud solely. A considerable portion of the target group is still unable to use the app despite having 2,000 servers allocated for 1,500 end users. So, a hybrid infrastructure that combines the use of edge servers with current cloud infrastructure is under investigation.

### 4.4. Hybrid Infrastructure

The present cloud is not well adapted to meet the latency requirements, as was described above, and an edge-only deployment requires a considerable number of edge sites and edge servers. Consequently, the effectiveness of a hybrid infrastructure will be discussed in this section. The term "smart edge" will be called to an edge server using our hybrid infrastructure.

According to the findings of the measurements, an infrastructure that connects CDN servers and current cloud datacenters is needed. Compared to cloud data centers that are located far away, CDN servers might provide end users lower latency because they are closer to them. Although the inclusion of CDN servers can significantly improve existing cloud architecture, there are still numerous issues that must be resolved. Determining the best smart edge configuration to optimize user coverage is a one challenge because the facility locating problem is an NP-hard problem, and this is an example of it. In order to find a solution, numerous heuristics are investigated [64].

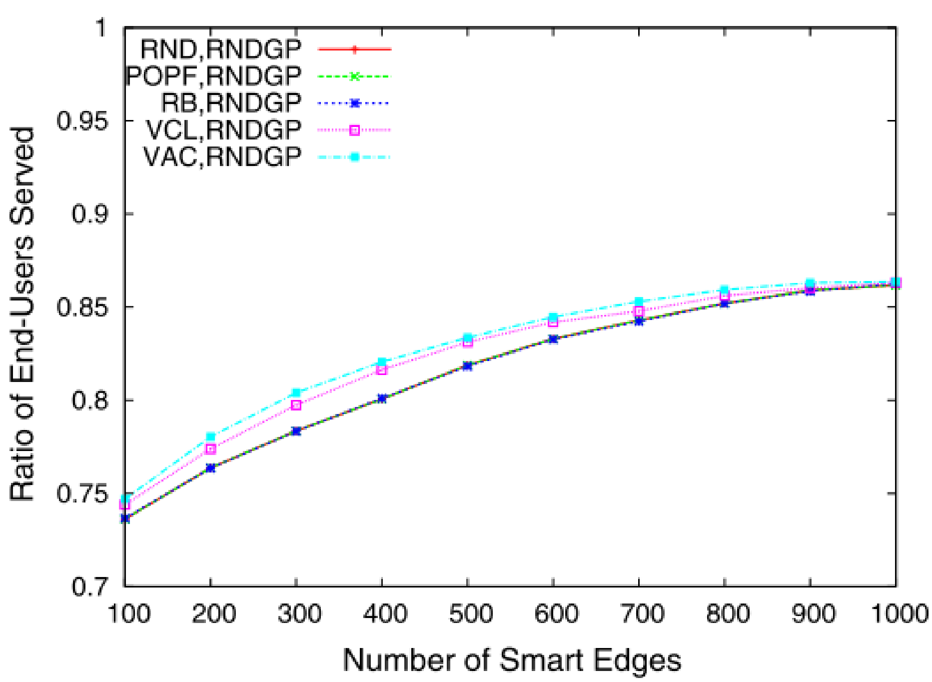


**Figure 8:** End-user ratio for various intelligent edge-selection algorithms. 1,500 customers, 5 applications per smart edge, and 100 applications overall [64]

As a result, Figure 8 shows that with 1,500 clients, it is possible to serve 86% of end users with 1,000 smart edges, if only 5 of the 100 accessible apps are randomly assigned to each smart edge. This shows that a significant portion of the client population is within 80 ms of either a datacenter or a smart edge.

An extensive measurement study is used to show that the current cloud infrastructure is inadequate to handle the demands of a new class of multimedia apps that are latency-sensitive. The ability of the cloud to handle interactive apps is specifically evaluated, and it is discovered that only 70% of end users can achieve the necessary 80 ms latency. Then, an edge-only deployment is investigated, where all end users are catered to by edge servers. In order to get a 90% ratio of served users, it is discovered that an excessive number of edge servers are necessary. Therefore, it is suggested to deploy edge servers, which are co-located at content distribution network (CDN) locations that are close to the end users, in order to enable latency-sensitive applications. Overall, by using a specific algorithm called "voting-based mechanisms for site selection", 90 % of end-users can be served. Therefore, the viability of interactive applications by enhancing the cloud with edge servers increases the performance significantly.

Another research to show benefits of hybrid computation is undertaken by authors of [66], they used more than 8.5k RIPE Atlas nodes to simulate an edge-cloud integrated services scenario and thoroughly examine how latency varies for consumers when they send queries to edge servers and cloud data centers. They used Akamai servers as edge servers, Ripe Atlas nodes as end-user representations, and key computing cloud providers as cloud locations. They calculated the roundtrip time (RTT) from 8,456 vantage points to 69 cloud locations and 6,341 edge servers. The words vantage point and end-user are used interchangeably in the remaining sections of the study. Five pings are sent from a given location to each cloud location and edge server as part of each latency assessment. For the measurement, the representative delay is the median RTT of five measurements.

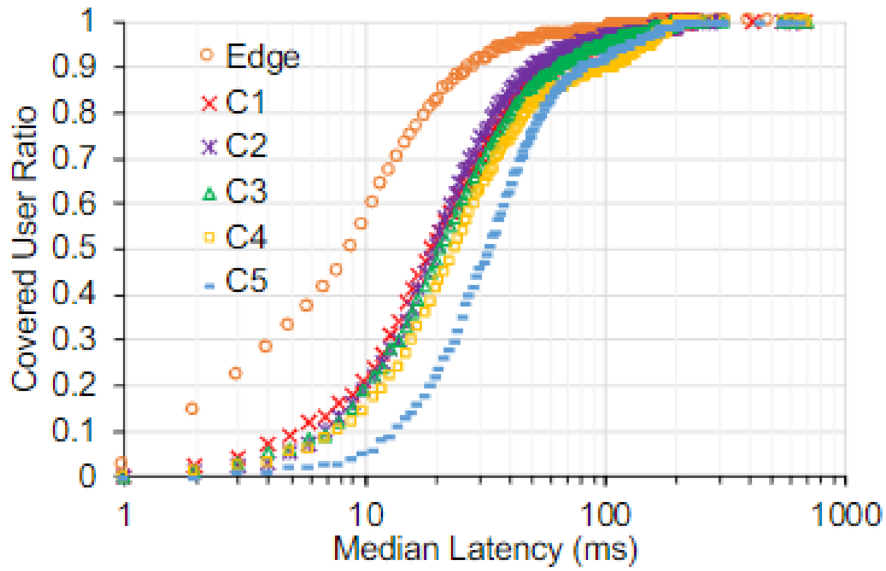


**Figure 9:** User latency coverage (CDF) [66]

Figure 9 compares the user's latency to the edge and several cloud providers denoted as C1-C5. With a latency of 20 ms or less, 82% of end users can locate a nearby edge server, compared to 55% with a latency of less than 10 ms. However, for specific cloud providers, the user coverage ratio varies for 10 ms and 20 ms latency ranges, respectively, between 3% and 21% and 22% and 52%.

There may be regional variations in latency characteristics, such as between Europe and Asia. Therefore, it is crucial to examine the regional variation in observed latency.

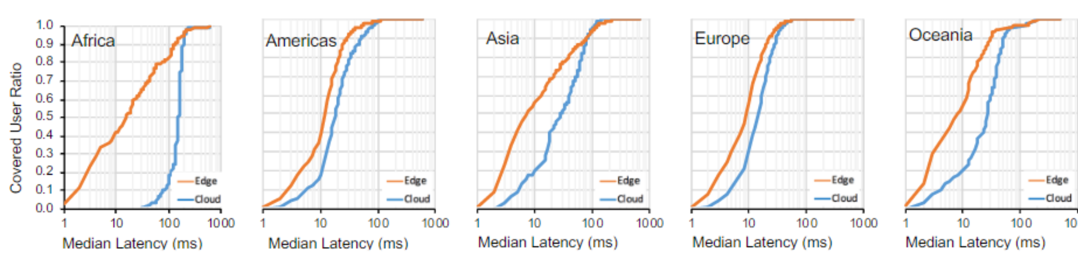


**Figure 10:** Continent-specific user coverage for edge servers and combined cloud providers [66]

Figure 10 compares user coverage between edge servers and all cloud sites throughout each continent. Due to the lack of a cloud datacenter in Africa, there is a significant disparity in user coverage. Similar trend is seen in Oceania, where each cloud provider may have up to two datacenters. Comparing edge servers to cloud, Americas and Europe have the best coverage.

Regional latency analysis demonstrates that cloud locations can match the performance of edge servers when data centers are plentiful, such as in West Europe.

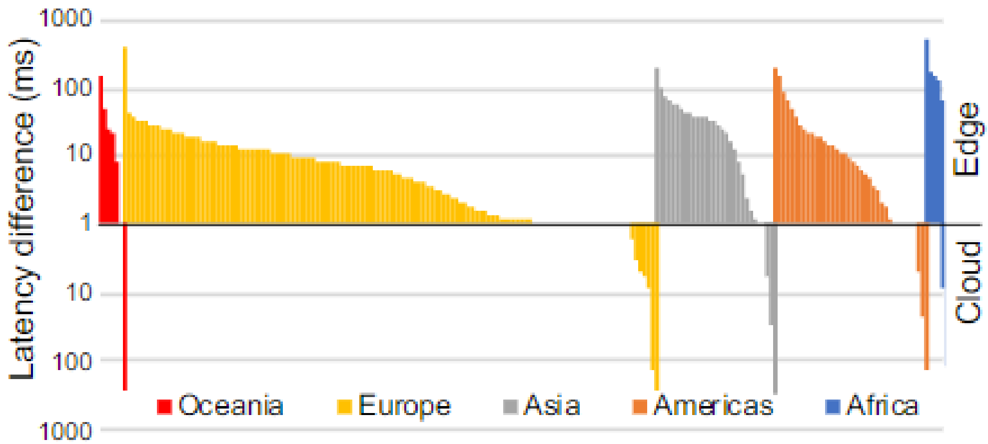


**Figure 11:** Comparison between edge servers and the best-performing cloud provider's latency. The end users are divided up by continent on the x-axis and are ranked according to the latency difference [66]

Figure 11 contrasts the edge servers' latency differences with those of cloud service provider C1, which is the best in terms of user latency. The end users are divided up by continent on the x-axis and are ranked according to the latency differences. The log-scaled delay difference is displayed on the y-axis. The research shows that end user latency to edge servers is significantly lower than end user latency to cloud locations. However, there are a few end users on each continent for whom the latency to the C1 cloud is significantly lower than the latency to the nearest edge server. This discrepancy may result from the absence of edge servers (i.e., Akamai servers) close to the end users. With the exception of Western Europe, where the difference is considerably lower, it is found that the average latency difference between edge servers and the combined cloud in Europe is roughly 10 ms, because so many cloud centers are found in Western Europe, including Frankfurt, London, Paris, and Amsterdam, end users there have a lower difference in latency between the cloud and the edge. The necessity for edge servers in those regions to enhance the quality of service for end users is heightened by the fact that some continents lack cloud provider coverage.

Overall, it has been found that edge servers are able to provide more end consumers with lower latency than cloud providers. More than 95% of end users are better served by the edge servers, often by orders of magnitude, even though the cloud can offer comparable latency in some regions (such as West Europe). These results demonstrate that edge servers are suitable for applications that require low latency.

End-user latency-based distribution to edge servers follows a power-law, necessitating non-uniform server deployment to prevent hot spots and improve end-user experience. Cloud datacenters hold thousands of servers that are installed in one place. As a result, cloud datacenters appear to have limitless resource capacity. Although edge servers significantly reduce latency compared to the cloud, capacity restrictions on edge servers may prohibit user requests from being fulfilled, particularly in densely populated areas.

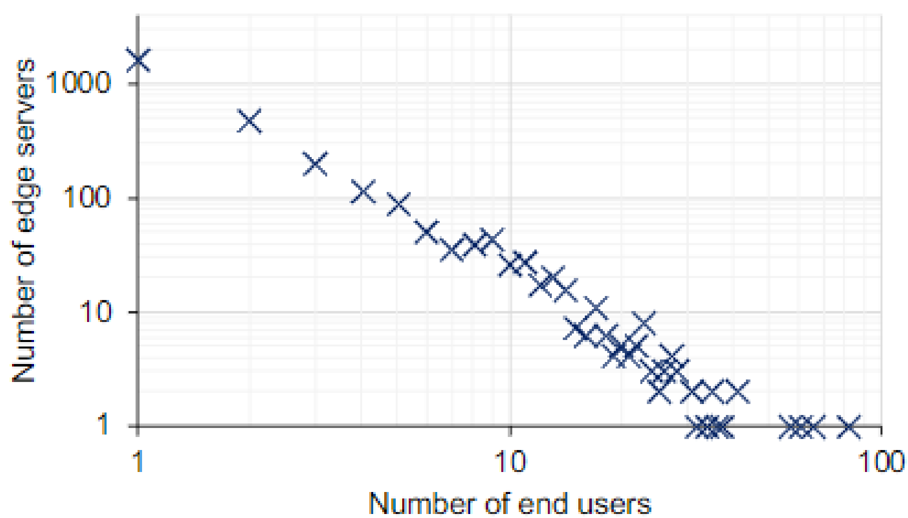


**Figure 12:** Distribution of users to edge servers [66]

When each end user is allocated to the edge server that is closest to them, the distribution of edge servers to end users is shown in Figure 12. Only 2,855 of the 6,341 edge servers are closest to at least one end user, with the remaining servers having greater latencies. When the closest edge server does not have the resources to handle the user workload, certain end users may be forwarded to a more distant edge server (or a cloud location).

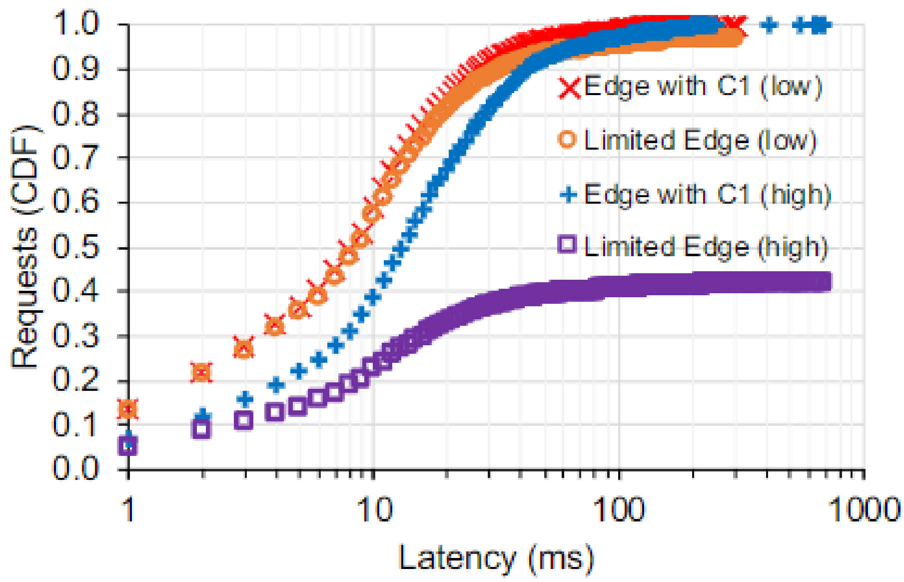


**Figure 13:** Coverage of user latency for both light and heavy workloads [66]

Figure 13 illustrates user latency coverage when an edge is supported by the cloud (i.e., Edge with C1) and when it is not supported by the cloud (i.e., Limited Edge), results in a constrained capacity. Under low workload conditions, C1's assistance can fulfil 59% of user requests with a latency of less than 10 ms, but as the workload increases, the ratio drops below 40%. The percentage of users who are covered when there is no cloud support falls short of 100% because some users cannot be scheduled to any edge servers nearby. This suggests that one feasible solution to the resource constraints of edge servers is to support them with resource-rich cloud locations.

# V. CONCLUSION

As Metaverse being one of the most intriguing research fields in academy and also in business lately, this thesis covered the metaverse enabling technologies, the challenges in implementation, its significance in business, line of sight for measuring the quality of experience and in this manner analyzed the performance and proposed methods for decreasing the latency.

Based on the findings of this study, considering the interactive and globally dispersed end-users of the company's Metaverse application, Virtual Showroom, it is suggested to use hybrid approach. The hybrid infrastructure setup includes the usage of edge servers and cloud datacenters together to achieve maximum coverage and minimum latency which will eventually result with a better Quality of Experience perceived by the end users. On the other hand, in the light of the previously conducted research, present in the literature section of this thesis, it is recommended to use CDN's for the XR App given that it contains only static content within the application.

This work revealed that enhancing the cloud with edge servers increases the performance significantly of the interactive applications. By reviewing the information collected for this thesis and the methodologies employed, from the performance perspective, this hybrid approach can be improved by determining the best smart edge configuration. The facility locating problem is an NP-hard problem therefore as a future work it is suggested to develop heuristic algorithms to locate cloud and edge servers in order to achieve optimal coverage and optimal latency.

# REFERENCES


[1] R. Cheng, N. Wu, S. Chen, and B. Han, “Reality Check of Metaverse: A First Look at Commercial Social Virtual Reality Platforms,” in *Proceedings - 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW 2022*, 2022, pp. 141–148. doi: 10.1109/VRW55335.2022.00040.

[2] J. Lee, I. Yeo, and H. Lee, “Metaverse Current Status and Prospects: Focusing on Metaverse Field Cases,” in *Proceedings - 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science, BCD 2022*, 2022, pp. 332–336. doi: 10.1109/BCD54882.2022.9900579.

[3] A. Devrim Yemenici and A. Devrim Yemenici Düzce, “Journal of Metaverse Review Article Entrepreneurship in The World of Metaverse: Virtual or Real? Entrepreneurship in The World of Metaverse: Virtual or Real?” [Online]. Available: https://www.researchgate.net/publication/362862020

[4] H. Ning *et al.*, “A Survey on Metaverse: the State-of-the-art, Technologies, Applications, and Challenges,” 2021, [Online]. Available: http://arxiv.org/abs/2111.09673

[5] Z. AYAZ and B. ERSÖZ, “Advertising Towards The Metaverse Universe,” *Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, Jun. 2022, doi: 10.26745/ahbvuibfd.1092245.

[6] D. Owens, A. Mitchell, D. Khazanchi, and I. Zigurs, “An Empirical Investigation of Virtual World Projects and Metaverse Technology Capabilities,” *Data Base Adv. Inf. Syst.*, vol. 42, no. 1, pp. 74–101, 2011, doi: 10.1145/1952712.1952717.

[7] S. Kumar *et al.*, “Second life and the new generation of virtual worlds,” *Computer (Long. Beach. Calif).*, vol. 41, no. 9, pp. 46–53, 2008, doi: 10.1109/MC.2008.398.

[8] C. Kai, H. Zhou, Y. Yi, and W. Huang, “Collaborative Cloud-Edge-End Task Offloading in Mobile-Edge Computing Networks with Limited Communication Capability,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 624–634, 2021,

doi: 10.1109/TCCN.2020.3018159.

[9] L.-H. Lee *et al.*, “All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda,” vol. 14, no. 8, pp. 1–66, 2021, [Online]. Available: http://arxiv.org/abs/2110.05352

[10] Y. Wang *et al.*, “A Survey on Metaverse: Fundamentals, Security, and Privacy,” *IEEE Commun. Surv. Tutorials*, 2022, doi: 10.1109/COMST.2022.3202047.

[11] S. Doolani *et al.*, “A Review of Extended Reality (XR) Technologies for Manufacturing Training,” *Technologies*, vol. 8, no. 4, p. 77, Dec. 2020, doi: 10.3390/technologies8040077.

[12] A. Qayyum *et al.*, “Secure and Trustworthy Artificial Intelligence-Extended Reality (AI-XR) for Metaverses,” Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.13289

[13] S. Takenoshita and H. Yasuhara, *Surgery and Operating Room Innovation*. 2021. doi: 10.1007/978-981-15-8979-9.

[14] M. Aloqaily, O. Bouachir, F. Karray, I. Al Ridhawi, and A. El Saddik, “Integrating Digital Twin and Advanced Intelligent Technologies to Realize the Metaverse,” *IEEE Consum. Electron. Mag.*, no. June, pp. 1–8, 2022, doi: 10.1109/MCE.2022.3212570.

[15] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, and E. Dutkiewicz, “MetaChain: A Novel Blockchain-based Framework for Metaverse Applications,” *IEEE Veh. Technol. Conf.*, vol. 2022-June, no. i, 2022, doi: 10.1109/VTC2022-Spring54318.2022.9860983.

[16] A. Perkis and C. Timmerer, “QUALINET White Paper on Definitions of Immersive Media Experience (IMEx),” 2020.

[17] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, “A subjective study on QoE of 360 video for VR communication,” *2017 IEEE 19th Int. Work. Multimed. Signal Process. MMSP 2017*, vol. 2017-Janua, pp. 1–6, 2017, doi: 10.1109/MMSP.2017.8122249.

[18] W. S. P. Group, "HetNet OpenSchema QoS / QoE Score Enabling Intelligent User-Oriented and Data Driven Traffic Steering," no. June, 2021.

[19] L. Badia, R. L. Aguiart, M. Zorzi, C. F. Ricerche, R. De Villejust, and S. Antipolis, "Wireless access architectures for video applications: the approach proposed in the MEDIEVAL project," pp. 991–996.

[20] I. Ahmed and L. Badia, "Analysis of management policies for multicast transmission of scalable video content in next generation networks," *2013 9th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2013*, pp. 1206–1211, 2013, doi: 10.1109/IWCMC.2013.6583728.

[21] L. Badia and A. V. Guglielmi, "A Markov analysis of automatic repeat request for video traffic transmission," *Proceeding IEEE Int. Symp. a World Wireless, Mob. Multimed. Networks 2014, WoWMoM 2014*, 2014, doi: 10.1109/WoWMoM.2014.6918965.

[22] L. Badia, "On the impact of correlated arrivals and errors on ARQ delay terms," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 334–338, 2009, doi: 10.1109/TCOMM.2009.02.070074.

[23] L. Badia, N. Baldo, M. Levorato, and M. Zorzi, "A Markov framework for error control techniques based on selective retransmission in video transmission over wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 488–500, 2010, doi: 10.1109/JSAC.2010.100419.

[24] L. Badia, A. Zanella, and M. Zorzi, "Game Theoretic Analysis of Age of Information for Slotted ALOHA Access With Capture," *INFOCOM WKSHPS 2022 - IEEE Conf. Comput. Commun. Work.*, 2022, doi: 10.1109/INFOCOMWKSHPS54753.2022.9797974.

[25] J. Van Der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming," *2020 12th Int. Conf. Qual. Multimed. Exp. QoMEX 2020*, pp. 3–8, 2020, doi: 10.1109/QoMEX48832.2020.9123081.

[26] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point cloud rendering after

coding: Impacts on subjective and objective quality," *IEEE Trans. Multimed.*, vol. 23, pp. 4049–4064, 2021, doi: 10.1109/TMM.2020.3037481.

[27] S. Porcu, A. Floris, and L. Atzori, "Quality of Experience in the Metaverse: An Initial Analysis on Quality Dimensions and Assessment," pp. 1–4, 2022, doi: 10.1109/qomex55416.2022.9900897.

[28] W. Y. B. Lim *et al.*, "Realizing the Metaverse with Edge Intelligence: A Match Made in Heaven," *IEEE Wirel. Commun.*, pp. 1–9, 2022, doi: 10.1109/MWC.018.2100716.

[29] I. Ahmed, L. Badia, N. Baldo, and M. Miozzo, "Design of a unified multimedia-aware framework for resource allocation in LTE femtocells," *MobiWac'11 - Proc. 9th ACM Int. Symp. Mobil. Manag. Wirel. Access, Co-located with MSWiM'11*, pp. 159–162, 2011, doi: 10.1145/2069131.2069160.

[30] D. Munaretto, L. Badia, T. Melia, and M. Zorzi, "Online QoE computation for efficient video delivery over cellular networks," *MMTC E-Letter*, vol. 7, no. 3, pp. 13–16, 2012, [Online]. Available: http://www.dei.unipd.it/~badia/papers/2012_E-Letter.pdf

[31] H. Du, B. Ma, D. Niyato, and J. Kang, "Rethinking Quality of Experience for Metaverse Services: A Consumer-based Economics Perspective," pp. 1–7, 2022, [Online]. Available: http://arxiv.org/abs/2208.01076

[32] C. Perera, C. H. Liu, and S. Jayawardena, "The Emerging Internet of Things Marketplace from an Industrial Perspective: A Survey," *IEEE Trans. Emerg. Top. Comput.*, vol. 3, no. 4, pp. 585–598, 2015, doi: 10.1109/TETC.2015.2390034.

[33] M. O. Thomas, B. A. Onyimbo, and R. Logeswaran, "Usability Evaluation Criteria for Internet of Things," *Int. J. Inf. Technol. Comput. Sci.*, vol. 8, no. 12, pp. 10–18, 2016, doi: 10.5815/ijitcs.2016.12.02.

[34] M. Varela, L. Skorin-Kapov, and T. Ebrahimi, "Quality of Service Versus Quality of Experience," *T-Labs Ser. Telecommun. Serv.*, no. June 2016, pp. 85–96, 2014, doi: 10.1007/978-3-319-02681-7_6.

[35] L. Badia and M. Zorzi, "On utility-based radio resource management with and

without service guarantees," *ACM MSWiM 2004 - Proc. Seventh ACM Symp. Model. Anal. Simul. Wirel. Mob. Syst.*, pp. 244–251, 2004, doi: 10.1145/1023663.1023707.

[36] L. Badia, M. Miozzo, M. Rossi, and M. Zorzi, "Routing schemes in heterogeneous wireless networks based on access advertisement and backward utilities for QoS support [Quality of service based routing algorithms for heterogeneous networks]," *IEEE Commun. Mag.*, vol. 45, no. 2, pp. 67–73, 2007, doi: 10.1109/MCOM.2007.313397.

[37] S. Kosta, A. Aucinas, P. Hui, R. Mortier, and X. Zhang, "ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading," *Proc. - IEEE INFOCOM*, pp. 945–953, 2012, doi: 10.1109/INFCOM.2012.6195845.

[38] B. G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti, "CloneCloud: Elastic execution between mobile device and cloud," *EuroSys'11 - Proc. EuroSys 2011 Conf.*, pp. 301–314, 2011, doi: 10.1145/1966445.1966473.

[39] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016, doi: 10.1109/JIOT.2016.2579198.

[40] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," *MCC'12 - Proc. 1st ACM Mob. Cloud Comput. Work.*, pp. 13–15, 2012, doi: 10.1145/2342509.2342513.

[41] A. Mishra, S. Karmakar, A. Bose, and A. Dutta, "Design and development of IoT-based latency-optimized augmented reality framework in home automation and telemetry for smart lifestyle," *J. Reliab. Intell. Environ.*, vol. 6, no. 3, pp. 169–187, 2020, doi: 10.1007/s40860-020-00106-1.

[42] L. Corneo and M. Eder, "Surrounded by the Clouds," pp. 295–304, 2009.

[43] Y. Cai, J. Llorca, A. M. Tulino, and A. F. Molisch, "Optimal Cloud Network Control with Strict Latency Constraints," *IEEE Int. Conf. Commun.*, pp. 1–6, 2021, doi: 10.1109/ICC42927.2021.9500573.

[44] J. Li *et al.*, "A general AI-defined attention network for predicting CDN performance," *Futur. Gener. Comput. Syst.*, vol. 100, pp. 759–769, 2019, doi:

10.1016/j.future.2019.05.067.

[45] L. Corneo *et al.*, "(How Much) Can Edge Computing Change Network Latency?," *2021 IFIP Netw. Conf. IFIP Netw. 2021*, 2021, doi: 10.23919/IFIPNetworking52078.2021.9472847.

[46] N. Mohan, L. Corneo, A. Zavodovski, S. Bayhan, W. Wong, and J. Kangasharju, "Pruning Edge Research with Latency Shears," *HotNets 2020 - Proc. 19th ACM Work. Hot Top. Networks*, pp. 182–189, 2020, doi: 10.1145/3422604.3425943.

[47] G. Min and M. Ould-Khaoua, "A performance model for wormhole-switched interconnection networks under self-similar traffic," *IEEE Trans. Comput.*, vol. 53, no. 5, pp. 601–613, 2004, doi: 10.1109/TC.2004.1275299.

[48] X. Zhang, G. Min, Q. Fan, H. Yin, D. Oliver Wu, and Z. Ma, "A Light-Weight Statistical Latency Measurement Platform at Scale," *IEEE Trans. Ind. Informatics*, vol. 18, no. 2, pp. 1186–1196, 2022, doi: 10.1109/TII.2021.3098796.

[49] X. Qiao, P. Ren, S. Dustdar, and J. Chen, "A New Era for Web AR with Mobile Edge Computing," *IEEE Internet Comput.*, vol. 22, no. 4, pp. 46–55, 2018, doi: 10.1109/MIC.2018.043051464.

[50] D. Van Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, "Edge Intelligence-Based Ultra-Reliable and Low-Latency Communications for Digital Twin-Enabled Metaverse," *IEEE Wirel. Commun. Lett.*, vol. 11, no. 8, pp. 1733–1737, 2022, doi: 10.1109/LWC.2022.3179207.

[51] J. Martin-Perez, L. Cominardi, C. J. Bernardos, A. De La Oliva, and A. Azcorra, "Modeling Mobile Edge Computing Deployments for Low Latency Multimedia Services," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 464–474, 2019, doi: 10.1109/TBC.2019.2901406.

[52] E. Huedo, R. S. Montero, R. Moreno-Vozmediano, C. Vázquez, V. Holer, and I. M. Llorente, "Opportunistic Deployment of Distributed Edge Clouds for Latency-Critical Applications," *J. Grid Comput.*, vol. 19, no. 1, 2021, doi: 10.1007/s10723-021-09545-3.

[53] K. Chen, T. Li, H. S. Kim, D. E. Culler, and R. H. Katz, "MARVEL: Enabling

mobile augmented reality with low energy and low latency," *SenSys 2018 - Proc. 16th Conf. Embed. Networked Sens. Syst.*, pp. 292–304, 2018, doi: 10.1145/3274783.3274834.

[54] W. Zhang, B. Han, and P. Hui, "Jaguar: Low latency mobile augmented reality with flexible tracking," *MM 2018 - Proc. 2018 ACM Multimed. Conf.*, pp. 355–363, 2018, doi: 10.1145/3240508.3240561.

[55] A. Sarkar, J. Murray, M. Dasari, M. Zink, and K. Nahrstedt, "L3BOU: Low Latency, Low Bandwidth, Optimized Super-Resolution Backhaul for 360-Degree Video Streaming," *Proc. - 23rd IEEE Int. Symp. Multimedia, ISM 2021*, pp. 138–147, 2021, doi: 10.1109/ISM52913.2021.00031.

[56] T. Sielhorst, W. Sa, A. Khamene, F. Sauer, and N. Navab, "Measurement of absolute latency for video see through augmented reality," *2007 6th IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR*, pp. 215–220, 2007, doi: 10.1109/ISMAR.2007.4538850.

[57] X. Ran, C. Slocum, Y. Z. Tsai, K. Apicharttrisorn, M. Gorlatova, and J. Chen, "Multi-user augmented reality with communication efficient and spatially consistent virtual objects," *Conex. 2020 - Proc. 16th Int. Conf. Emerg. Netw. Exp. Technol.*, pp. 386–398, 2020, doi: 10.1145/3386367.3431312.

[58] S. M. Salman, T. A. Sitompul, A. V. Papadopoulos, and T. Nolte, "Fog Computing for Augmented Reality: Trends, Challenges and Opportunities," *Proc. - 2020 IEEE Int. Conf. Fog Comput. ICFC 2020*, pp. 56–63, 2020, doi: 10.1109/ICFC49376.2020.00017.

[59] I. S. Association, "IEEE 1934-2018 standard for adoption of openfog reference architecture for fog computing," in *IEEE 1934-2018 standard for adoption of openfog reference architecture for fog computing*, 2018.

[60] S. Dhelim, T. Kechadi, L. Chen, N. Aung, H. Ning, and L. Atzori, "Edge-enabled Metaverse: The Convergence of Metaverse and Mobile Edge Computing," vol. 14, no. 8, pp. 1–8, 2022, [Online]. Available: http://arxiv.org/abs/2205.02764

[61] A. Younis, B. Qiu, and D. Pompili, "Latency-aware Hybrid Edge Cloud Framework

for Mobile Augmented Reality Applications," *Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Networks Work.*, 2020, doi: 10.1109/SECON48991.2020.9158429.

[62] F. S. Abkenar, L. Badia, and M. Levorato, "Selective Data Offloading in Edge Computing for Two-Tier Classification With Local Domain Partitions" , Proceedings IEEE PerCom, 2023.

[63] A. V. Guglielmi, M. Levorato, and L. Badia, "A Bayesian game theoretic approach to task offloading in edge and cloud computing," *2018 IEEE Int. Conf. Commun. Work. ICC Work. 2018 - Proc.*, pp. 1–6, 2018, doi: 10.1109/ICCW.2018.8403695.

[64] S. Choy, B. Wong, G. Simon, and C. Rosenberg, "A hybrid edge-cloud architecture for reducing on-demand gaming latency," *Multimed. Syst.*, vol. 20, no. 5, pp. 503–519, 2014, doi: 10.1007/s00530-014-0367-z.

[65] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hoßfeld, "An evaluation of QoE in cloud gaming based on subjective tests," *Proc. - 2011 5th Int. Conf. Innov. Mob. Internet Serv. Ubiquitous Comput. IMIS 2011*, pp. 330–335, 2011, doi: 10.1109/IMIS.2011.92.

[66] B. Charyyev, E. Arslan, and M. H. Gunes, "Latency Comparison of Cloud Datacenters and Edge Servers," *2020 IEEE Glob. Commun. Conf. GLOBECOM 2020 - Proc.*, 2020, doi: 10.1109/GLOBECOM42002.2020.9322406.