

Application Scenarios of Edge Computing in Conjunction with Cloud Computing and Artificial Intelligence

Zehao Li *

School of Computer and Communication Engineering, Dalian Jiaotong University, Dalian, China

* Corresponding Author Email: abrown67198@student.napavalley.edu

Abstract. The proliferation of mobile technology products and the development of wireless communication networks have given rise to numerous application scenarios that prioritize higher computing power, better privacy protection, lower latency, and lower energy consumption. Given these requirements, edge cloud computing has emerged as a critical component of computing platforms and a key enabler for communication technology. In this paper, the concept of edge computing and its synergies with cloud computing and artificial intelligence are explored. This paper outlines the various advantages that can be obtained through the collaboration of edge computing and cloud computing. Specifically, it focuses on existing and potential scenarios where this collaboration can be beneficial. These include the Internet of Things, mobile gaming, metaverse and extended reality. The potential for innovation in these fields is immense and the possibilities are vast. The paper also presents the definition of artificial intelligence and its opportunities in the field of communication networks. Particularly, a discussion of feasible models for integrating artificial intelligence with edge computing concludes the paper.

Keywords: Edge computing; Cloud computing; Artificial intelligence.

1. Introduction

Ever since Google introduced the paradigm of cloud computing back in 2006, this transformative concept has significantly impacted the field of information technology, ultimately facilitating the integration of mobile device resources and empowering users with more streamlined approaches to attaining knowledge, executing tasks, and conducting everyday life activities [1]. However, recent years have witnessed an unprecedented burst in the expansion of the Internet of Things technology, leading to a growth of demands for data storage and treatment.

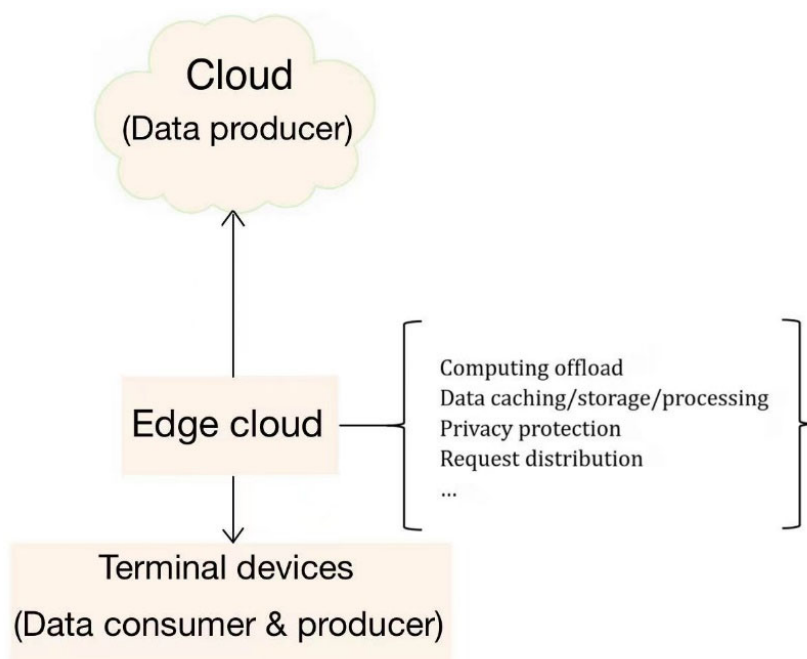


Fig. 1 Edge computing paradigm [2]

A novel computing paradigm, known as edge computing in Fig 1, has been proposed as a means of harmonizing the surplus and dispersed resources located at the edge of networks to offer users with better services. The term "edge computing" refers to innovative technological concepts that facilitate the execution of computational tasks at the proximity of the data sources. It involves both downstream data processing for cloud services and upstream data processing for Internet of Things services [2]. By undertaking computational tasks closer to the users, edge computing presents various advantages. One notable benefit of edge computing is its ability to offload computations, which reduces the burden on the network. With the growing need for real-time applications, the speed of computation is becoming increasingly critical. Edge computing reduces the latency experienced when transferring data to remote data centers, by performing computational tasks closer to where the data is generated. This results in faster response time and guarantees a seamless user experience. Moreover, edge computing offers the ability to store and cache data locally. Given the massive volume of data generated, the cost and inefficiency of centralized data centers can become obstacles. Edge computing allows for faster access to frequently used data by storing it at the edge, reducing data transfer and improving the overall system performance. This is especially valuable in scenarios where data privacy is critical because local storage enhances the security and control of sensitive information.

The remainder of this paper is organized as follows. The synergy between cloud computing and edge computing continues to be discussed in depth through various applications in section 2. And section 3 discusses artificial intelligence and feasible models for integrating with edge computing.

2. Collaborating with Cloud Computing

With the advancement in mobile technology products, many application scenarios demand higher computing power, lower latency, and lower energy consumption. In this case, edge cloud computing can serve as an important part of cloud computing and the key to 5G communication technology. In this section, numerous widespread applications that benefit from edge cloud computing will be discussed, along with related challenges.

2.1. Internet of Things (IoT)

The IoT uses a variety of technologies to connect physical electronic devices or systems by receiving and transmitting data over a network without human intervention. And the IoT platform facilitates the provision of access and management functionalities to a considerable number of devices, allowing people to conveniently and efficiently access various forms of terminal devices, and can also perform rich and complete device management in the cloud. However, the cloud IoT platform is far from the terminal devices, and the terminal devices themselves do not have strong computing power. As a consequence, for scenarios with high real-time requirements, the cloud cannot handle problems in a timely manner due to network delay, network congestion and other reasons. In this process, Edge cloud computing can help cloud computing assume the computing role, including storage, operation and transmission, because in IoT, electronic devices are not only consumers but also producers of data, and traditional cloud computing services are difficult to efficiently undertake such huge data [2]. Meanwhile, transmitting data with the cloud will lead to an increase in error rates and require more complex check codes, resulting in greater use of computing resources, especially increasing energy consumption on the mobile side. Data security will also be exposed to greater risks. However, processing data at the edge cloud can not only save more energy, but also provide safer internet service, benefiting from the fact that edge cloud devices are closer to users [2]. For example, devices such as smart home devices that are widely used in populated areas and are not frequently moved are quite suitable to be target terminals for edge cloud computing.

2.2. Games

Compared with traditional computer gaming, which highly depend on the PC performance to maintain the game quality, such as memory usage, FPS, CPU utilization, etc, cloud gaming is a

paradigm for online gaming whereby all game operations are executed on a central server infrastructure. The generated graphical outputs are subsequently compressed and conveyed to end-users through network channels. As a result, even the mobile devices like cell phones or tablets, are expected to run large games even with insufficient memory, battery life, hardware and computing power. At the same time, by offloading computing tasks to nearby edge servers, players can be guaranteed better game frame rate, image quality and acceptable latency [3]. Thanks to the advantage of the close proximity of edge servers, a balanced and optimal strategy can be provided among computational offloading, backhaul delays and additional network burdens.

2.3. Metaverse and Extended Reality (XR)

As described in Spielberg's movie Ready Player One, the Metaverse is committed to creating a brand-new social network that allows people to engage in social interactions in a 3D form of digital avatars [4]. This enables them to experience social activities that not only exist in the real world, like shopping, learning and business meetings, but also more novel activities like 3D gaming. The metaverse is also currently valued and invested by many business giants and game companies. For example, the pioneer of the social network, Facebook, has changed its name to Meta in 2021, while Apple released its new MR product, VisionPro, at the WWDC2023.

XR, including virtual reality (VR), augmented reality (AR), mixed reality (MR), etc, refers to devices that obtain digital information from other social media platforms, as well as collect biological data such as hand- and eye-tracking so as to be capable of offering users immersive experiences through multiple sensors and real-time interactions [4]. And XR devices are supposed to be an important platform for experiencing the Metaverse, such as helmet-mounted displays [4].

As is discussed above, the development of the IoT has caused a tremendous computing stress and traffic flows on the cloud. Metaverse requires a computing breakthrough to better guarantee service performance by connecting XR devices with edge computing, specifically with mobile edge computing. Due to its advantage in dealing with the privacy, latency, and energy consumption, edge computing is receiving increasing attention from business and academia. Depending upon 5G and mobile edge computing technology, the motion-to-render-to-photon latency of Metaverse service facilities can be minimized effectively less than 20ms, which is very practical [4]. Besides that, more cutting-edge researches are undergoing. For instance, paper came up with an exciting FACT algorithm to balance computational latency, network latency and system accuracy in an AR system based on mobile edge computing, which means a lot to improve the performance when meeting multiple mobile users [5].

2.4. Challenges in Application Scenarios

Facilities' Robustness: Facilities' robustness and related data security is the first issue to be concerned. Because edge cloud facilities are dispersed, their environment is more complex and they do not transmit all data to the cloud. Therefore, once damaged, edge cloud facilities are more likely to cause data loss than the central cloud.

Data Migration: When service objects are frequently migrated, moving from one edge node to another, data migration becomes a huge problem. And it is also an infrastructure challenge for businesses, owing to the slow and wasteful process.

Heterogeneous System Interconnection: In order to attain interoperability between heterogeneous edge devices and systems, it is imperative for disparate device vendors to establish pertinent standards and shared collaboration protocols [6].

Computer Resource Allocation: For a computationally intensive job, mobile devices and edge computing devices need to find the optimal strategy for resource allocation, because of the limited computing resources. However, it is generally NP-hard [7].

Steep Cost: Edge computing further increases the cost of construction, operation and maintenance of communication facilities, because communication base stations and other mobile computing nodes have to assume more computing functions, while higher power and wider channels will also increase

power supply requirements. The construction process of 5G includes the construction, operation and maintenance of millions of base stations, with a construction cost of more than 100 billion US dollars, and further improvement of edge computing to meet the needs of wider applications in the future requires new facilities and updates on the basis of existing 5G base stations, including new equipment, power supply expansion and other work. Although costs can be reduced as technology advances and economies of scale are realized, it is still a huge expense. This requires mutual cooperation between enterprises and countries to update the computing network and ensure better services.

3. Integrated with Artificial Intelligence (AI)

In this section, the paper will briefly introduce AI and how it will be combined with edge computing to influence the future development of the communication field. It will also focus on the application prospects of distributed AI and decentralized AI for autonomous driving.

3.1. AI

A detailed definition of artificial intelligence is "the system's competence in accurately comprehending external data, acquiring knowledge from it, and employing that acquired knowledge to accomplish particular objectives and tasks through adaptable adjustment" [8]. In the past decade, AI has been significantly developed and advanced in many fields, and thanks to the powerful data aggregation capabilities of big data analytics systems and cloud platforms, AI has now achieved breakthroughs in image recognition, natural language processing, etc. Meanwhile, there are broader prospects for business development. For example, since OpenAI released the natural language model ChatGPT, Microsoft has combined its GPT-4 model with the Bing search engine in an attempt to explore the future development direction of the search engine, and a number of Internet companies, such as Google, have also actively invested in research and development and have released successively their own artificial intelligence natural language models.

Machine learning (ML), as a sub-field of AI, enables machines to process large amounts of data and imitate the way humans think so as to make predictions and decisions [9]. And ML can be roughly categorized into supervised learning, unsupervised learning, and reinforcement learning.

Supervised Learning: Supervised learning algorithms aim to acquire knowledge from a precisely labeled dataset in order to effectively predict outcomes. The input data is channeled into the model, which iteratively adapts its inherent weights until the model is appropriately calibrated. It is used to build highly accurate machine learning models that can classify data or accurately predict outcomes.

Unsupervised Learning: Unlike the supervised learning, algorithms of unsupervised learning learn from the unlabeled data to discover hidden patterns or data groupings. And clustering, association and dimensionality reduction are main tasks of unsupervised learning models.

Reinforcement Learning: Motivated by the principles governing the psychology of human and animal behavior, algorithms of reinforcement learning learn as they go by trial and error. The models adhere to the paradigm of an intelligent agent engaging in interactions with the environment and taking feedback so as to reinforce the successful outcomes and seek the optimal decision-making. Nowadays, the deep neural networks have been successfully integrated into more difficult tasks such as the excellent performance of DeepMind's Alpha Go in the game of Go. For some problems where a comprehensive dataset is unavailable, deep reinforcement learning algorithms also have demonstrated encouraging outcomes [10].

3.2. Edge Computing Combined with AI

By introducing AI into edge computing, on the device side, users can reduce their dependence on mobile networks, improve the efficiency of data processing, enhance mobile computing capabilities, and better protect information security. For example, Apple introduced the Core ML framework into its chip designs in 2017. As a machine learning framework, Core ML greatly enhances the performance of mobile devices as well as image recognition, natural language processing and other

functions by supporting a wide range of models and running on the device side. Meanwhile, since it does not rely on the network, Core ML also better protects user privacy and reduces the power consumption of mobile devices.

With the growth of the computing power, ML techniques have successfully applied in solving engineering problems. In the domains of communication and electronic engineering, AI along with ML, will undoubtedly play a greater role in the future. On the basis of the Internet of Things (IoT) and the fifth generation (5G) wireless communication networks, research on the sixth generation (6G) of wireless communication networks focuses more on making applications more intelligent and enabling service coverage worldwide, and specific improvements are illustrated in Table 1 [11].

Table 1. Characteristics of 5G edge computing and 6G wide-area cloud [11]

	5G edge computing	6G wide-area cloud
Regional deployment relationship	Regional edge or private network on-premise edges	Widely distributed across device computing, network computing, regional edges and datacenters
Ownership and security domain	A single ownership and security domain	Multiple ownership and security domains
Computing resources in each location	Abundant computing resources in each location	Limited computing resources in each location
Relationship with mobile network	On the top of mobile network	Coupled with mobile network
Computing workload partition	Fixed client-server computing workload partition	Dynamic computing workload partition and distribution

These new application scenarios bring numerous new performance necessities to 6G networks, including 10–100 million devices connected with peak data rates of 1–10 TB/s; latency needed reducing to a fraction of 1 millisecond to account for real-time control applications; mobility needed to be supported exceeding 1000 km/h to accommodate ultra-high speed trains, drones and satellites; and for mission-critical applications, around six times more reliability must be achieved [9].

To better meet these goals above, combining edge computing with AI can be a promising workable method. In this system, the specific forms of integration can be generally classified into distributed AI and decentralized AI. Distributed AI system consists of the central nodes and autonomous learning processing nodes at a large scale, to maximize advantages of AI for rapid, real-time data analysis and decision-making, while decentralized AI system refers to the development of AI algorithms and models directly on local edge devices by utilizing blockchain technology to make multiple nodes reach the consensus without the central authority. These systems can play significant roles in applications such as autonomous driving.

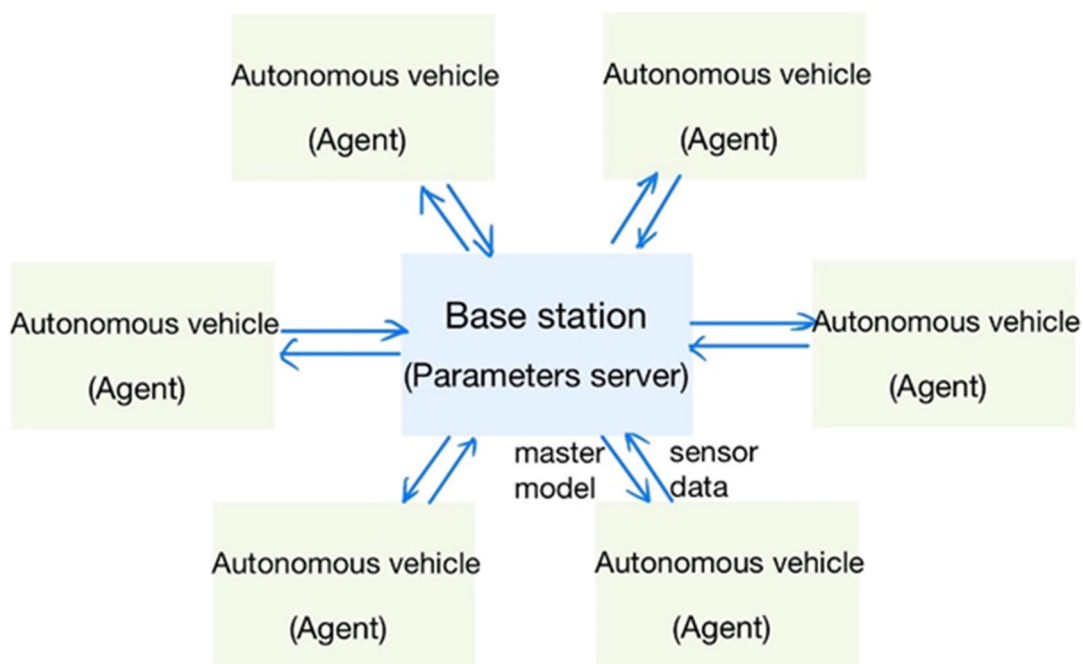


Fig. 2 Distributed AI paradigm used in autonomous driving [10]

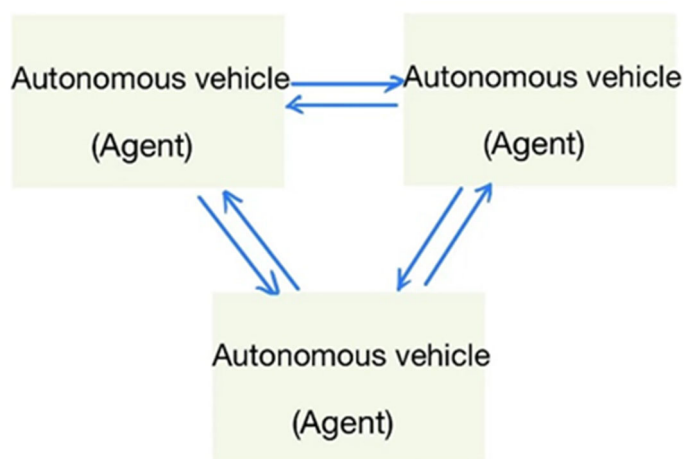


Fig 3. Decentralized AI paradigm used in autonomous driving.

Autonomous driving is the technology that allows vehicles to operate without human involvement [12]. This technology is constantly evolving and is categorized into five levels. From level 0 to level 5, they are no automation, driver assistance, partly, highly, fully automated driving and full automation. For now, level 3,4 and 5 are still in the testing phase. Because of the absence of an all-inclusive dataset, reinforcement learning algorithms can perform better than simple supervised learning algorithms [10]. And due to the complexity of real-world scenarios, the commercialization of autonomous vehicles is a gradual and risky move. Therefore, the early algorithm design by AI also needs to rely more on feasible models. In order to train to design algorithms that can control vehicles through sensor data in an orderly, safe and efficient way, using a distributed or a decentralized system can be a good strategy. Fig 2 and Fig 3 show the distributed AI paradigm and the decentralized AI paradigm used in autonomous driving respectively. A distributed system uses the base station as the center (parameter server) and the autonomous vehicles as the base points (agents), and a decentralized system refers to a system between autonomous driving cars as the basic model for AI reinforcement learning.

4. Conclusion

The paper presents the notion of edge computing, examines the progression of Internet of Things (IoT), mobile gaming, metaverse and extended reality (XR) in correlation with cloud computing in the 5G setting. Additionally, it discusses the concept and development of artificial intelligence (AI) together with various categories of machine learning. It also explores the potential integration of AI and edge computing in 6G research to facilitate the training of machine learning algorithms for autonomous driving using distributed AI and decentralized AI models. As a distributed computing model, the practice of edge computing facilitates the processing of data in proximity to the site of its origin. This has led to the development of innovative solutions that can effectively reduce signal latency and save bandwidth. And these solutions provide a great help in enhancing the speed and performance of various applications, while also ensuring better security and greater flexibility for data. Ultimately, this will lead to the creation of more efficient and intelligent systems that can make decisions faster and more accurately. However, as edge computing gains popularity, several challenges remain to be addressed to enhance its efficiency. The main obstacle pertains to the restricted computing power of edge computing. Additionally, the management of devices and the allocation of resources require refinement. Finally, complex environments pose great threats to data security. Overcoming these challenges is vital for the widespread adoption of edge computing.

In conclusion, despite the persistent technical engineering challenges, the distinctive benefits of edge computing make it full of potential. And edge computing will transform the way we interact with technology and will bring significant benefits to businesses and society as a whole.

References

- [1] M. Armbrust, A. Fox, R. Griffith, et al. A view of cloud computing. *Communications of the ACM*, 2010, 53(4): 50-58.
- [2] W. Shi, J. Cao, Q. Zhang, et al. Edge computing: vision and challenges. *IEEE Internet of Things Journal*, 2016, 3(5): 637-646.
- [3] J. Li, Q. Liang, D. Li, et al. Game-based task offloading in cloud-edge-end network. *International Conference on Computer and Communication Systems (ICCCS)*, 2023, 549-553.
- [4] Y. Wang, J. Zhao. A survey of mobile edge computing for the metaverse: architectures, applications, and challenges. *International Conference on Collaboration and Internet Computing (CIC)*, 2022, 1-9.
- [5] Q. Liu, S. Huang, J. Opadere, et al. An edge network orchestrator for mobile augmented reality. *IEEE Conference on Computer Communications*, 2018, 756-764.
- [6] Q. Luo, S. Hu, C. Li, et al. Resource scheduling in edge computing: a survey. *IEEE Communications Surveys & Tutorials*, 2021, 23(4): 2131-2165.
- [7] T. Bahreini, H. Badri, D. Grosu. Mechanisms for resource allocation and pricing in mobile edge computing systems. *IEEE Transactions on Parallel and Distributed Systems*, 2022, 33(3): 667-682.
- [8] M. Haenlein, A. Kaplan. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. *California Manage. Rev.*, 2019, 61(4): 5-14.
- [9] C. Huang, et al. Artificial intelligence enabled radio propagation for communications—part I: channel characterization and antenna-channel optimization. *IEEE Transactions on Antennas and Propagation*, 2022, 70(6): 3939-3954.
- [10] M. Spryn, A. Sharma, D. Parkar, et al. Distributed deep reinforcement learning on the cloud for autonomous driving. *IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS)*, 2018, 16-22.
- [11] Q. Li, et al. 6G cloud-native system: vision, challenges, architecture framework and enabling technologies. *IEEE Access*, 2022, 10: 96602-96625.
- [12] Atakishiyev S., Salameh M., Yao H., et al. Explainable artificial intelligence for autonomous driving: a comprehensive overview and field guide for future research directions. *Arxiv E-prints*, 2021.