

Edge Computing for IoT

Balqees Talal Hasan and Ali Kadhum Idrees

Abstract Over the past few years, The idea of edge computing has seen substantial expansion in both academic and industrial circles. This computing approach has garnered attention due to its integrating role in advancing various state-of-the-art technologies such as Internet of Things (IoT) , 5G, artificial intelligence, and augmented reality. In this chapter, we introduce computing paradigms for IoT, offering an overview of the current cutting-edge computing approaches that can be used with IoT. Furthermore, we go deeper into edge computing paradigms, specifically focusing on cloudlet and mobile edge computing. After that, we investigate the architecture of edge computing-based IoT, its advantages, and the technologies that make Edge computing-based IoT possible, including artificial intelligence and lightweight virtualization. Additionally, we review real-life case studies of how edge computing is applied in IoT-based Intelligent Systems, including areas like healthcare, manufacturing, agriculture, and transportation. Finally, we discuss current research obstacles and outline potential future directions for further investigation in this domain.

1 Introduction

In as early as 1966, an insightful prediction emerged from Karl Steinbuch, a pioneer in German computer science. He predicted that within a few decades, computers would be a necessary component of almost every industrial product. The term "pervasive computing" was first introduced by W. Mark in 1999. It means integrating computers into everyday objects so seamlessly that they become a natural and unno-

Balqees Talal Hasan,
Dept. of Computer and Information Engineering, Nineveh University, Mosul, Iraq , e-mail: balqees.hasan@uoninevah.edu.iq

Ali Kadhum Idrees,
Dept. of Information Networks, University of Babylon, Babylon, Iraq, e-mail: ali.idrees@uobabylon.edu.iq

ticed part of the environment, with people interacting with them effortlessly. In the same year (1999), The term "Internet of Things" was coined by Kevin Ashton at a presentation at Procter & Gamble (P&G) [2].

IoT is a new paradigm for attaching various physical objects to the Internet so they can interact and make informed decisions. The technologies that fall under this paradigm include pervasive computing, RFID, communication technologies, sensor networks, and Internet protocols. In IoT, physical things have the ability to intelligently collaborate and establish connections with the Internet, operating autonomously and introducing innovative applications. These applications span a variety of industries, such as manufacturing, transportation, healthcare, industrial automation, and emergency response [10, 33].

IoT has become permeated our daily lives, providing crucial measuring and data-gathering capabilities that influence our decision-making. Numerous sensors and gadgets run continuously, producing data and enabling vital communication over complex networks. It is challenging to execute complicated computations on most of the IoT devices due to their limited CPU and energy resources. In general, IoT devices collect data and transmit it to more robust processing centers for analysis [35]. The data is subjected to extra processing and analysis at these centers [36]. In order to lighten the load on resources and avoid overuse of them, edge computing has become prevalent as a novel way to address IoT and local computing requirements [9]. In edge computing, small servers that are located closer to users are used to process data. In close proximity to the devices of the consumers, these edge servers are capable of doing complex tasks and storing enormous volumes of data. As a result of their proximity to users, processing and storage at the network edge becomes faster and more efficient [26]. In 2022, the worldwide edge computing market was worth USD 11.24 billion. Experts predict that it will experience significant expansion, with an expected yearly growth rate of 37.9% from 2023 to 2030 [6]. Edge computing is different from the usual cloud computing approach. Instead of processing and storing data in centralized data centers far from users, edge computing involves positioning resources closer to users, specifically at the "edge" of the network. This means there are multiple computing nodes spread throughout the network, which reduces the burden on the central data center and makes data exchange much faster, as there is less delay in sending and receiving messages [9]. Edge computing allows for the intelligent collection, analysis, computation, and processing of data at every IoT network edge. This implies that data can be filtered, processed, and used close to the devices or data sources, where it is generated. Edge computing makes everything faster and more effective by pushing smart services to the edge of the network. Making decisions and processing data locally can also help deal with significant limitations in networks and resources, and it can address concerns related to security and privacy too [14, 34].

Here is how this chapter is organized: Section 2 provides a comprehensive explanation of computing paradigms for IoT. Moving on to Section 3, a detailed introduction to edge computing paradigms is presented. Section 4 outlines the architecture of edge computing-based IoT. In Section 5, the focus shifts to illustrate the advantages of edge computing-based IoT. The enabling technologies for edge computing-based

IoT are introduced in Section 6. Section 7, the chapter reviews edge computing in IoT-Based Intelligent Systems. Section 8 illustrates the challenges and future research directions for edge computing-based IoT. Finally, Section 9 concludes the chapter.

2 Computing Paradigms for IoT

This section describes the fundamental concepts underlying the three major computing paradigms and how they are integrated with IoT: cloud computing, edge computing, and fog computing, figure 1 shows the architecture of the 3 tiers computing paradigms 1.

2.1 Cloud Computing

As mobile hardware evolves and improves, it will always face limitations in terms of available resources compared to stationary hardware. Regarding devices that people wear or carry for extended periods, prioritizing improvements in weight, size, and battery life takes precedence over enhancing computational power. This is a fundamental aspect of mobility rather than a transient restriction imposed by modern mobile electronics. Therefore, there will always be trade-offs when using computational power on mobile devices. The resource limitations of mobile devices can be solved simply and effectively by using cloud computing. With this approach, a mobile device can execute an application that requires a lot of resources on a robust remote server or a cluster of servers, allowing users to interact with the application through a lightweight client interface over the Internet [3].

The cloud computing paradigm gives end users on-demand services by utilizing a pool of computing resources. These resources include computing power, storage, and more, and they are all immediately available at any time [8].

IoT and the cloud have had separate evolutionary processes. However, its integration has produced a number of benefits for both parties. On the one hand, IoT can greatly profit from the boundless capabilities of cloud computing to overcome its own technological limitations, such as storage, processing power, and energy requirements. The cloud can take advantage of IoT through expanding its range of applications to handle real-world objects in a more distributed and adaptable fashion, thus providing new services in various real-life situations [4].

2.2 Edge Computing

Typically, the architecture of the cloud is used to manage the massive amount of data that IoT devices produce. However, cloud computing encounters various challenges

such as lengthy transmission time, increased bandwidth requirements, and latency between IoT devices and the cloud. The concept of edge computing has emerged to overcome these difficulties. This approach enhances scalability, latency, and privacy factors while enabling real-time predictions by processing data at the source [5, 41]. In an extension of cloud computing, edge computing places computer services at the edge of the network, where they are more accessible to end users. Edge computing shifts services, computational data, and applications out from cloud servers and toward the edge of a network. This enables content providers and application developers to provide users with services that are located nearby. Edge computing is unique in that it may be used for a variety of applications thanks to its high bandwidth, very low latency, and fast access to network data [8, 42].

In the world of IoT, both Edge computing and Cloud computing offer major advantages due to their distinct characteristics, such as their capacity to execute complex computations and store large amounts of data. However, when it comes to IoT, edge computing outperforms cloud computing, despite having somewhat limited compute capability and storage capabilities. In particular, IoT demands fast responses rather than powerful computing and massive storage. Edge computing fulfills the requirements of IoT applications by offering satisfactory computing capability, sufficient storage, and fast response times. Edge computing, on the other hand, can also leverage IoT to expand its framework and adapt to the dynamic and distributed nature of edge computing nodes. These edge nodes can serve as providers and may consist of either IoT devices or devices with some residual computational capabilities [9].

2.3 Fog Computing

CISCO introduced the concept of fog computing in January 2014 [11]. This computing paradigm offers numerous advantages across various fields, especially the IoT [12, 37]. According to Antunes, a senior official in charge of advancing corporate strategy at Cisco, edge computing is a division of fog computing. He explains that fog computing primarily focuses on managing the location of data generation and storage. In essence, edge computing involves processing data in proximity to its source of origin [38]. Fog computing, on the other hand, leverages edge processing and the necessary network connections to transfer data from the edge to the endpoint [11]. The fog computing system was not designed to replace cloud computing; instead, its development aimed to fill any service gaps present in cloud computing. Fog computing emphasizes on bringing abilities of cloud computing closer to the edge of the network so that users can access communication and software services faster. This approach works well for offering cloud solutions for highly mobile technologies like vehicular ad hoc networks (VANET) and the IoT [13]. Fog computing serves the endpoints or edges of the network of interconnected devices. It prioritizes the analysis of time-sensitive data near the sources, sending only the selected and abridged data to the cloud [11, 39].

The concept of "fog as a service" (FaaS) is a new service possibility brought about

by the integration of fog computing and IoT. According to this concept, a service provider creates a network of fog nodes throughout the area covered by its service, operating as a landlord to numerous tenants from diverse businesses. Each fog node provides storage, networking, and local computing capabilities. Through FaaS, customers can access services using innovative business models. Unlike clouds, which are often managed by huge businesses with sizable data centers, FaaS enables both small and large businesses to create and manage public or private computing, storage, and control services at different scales, meeting the requirements of various clients [12, 40].

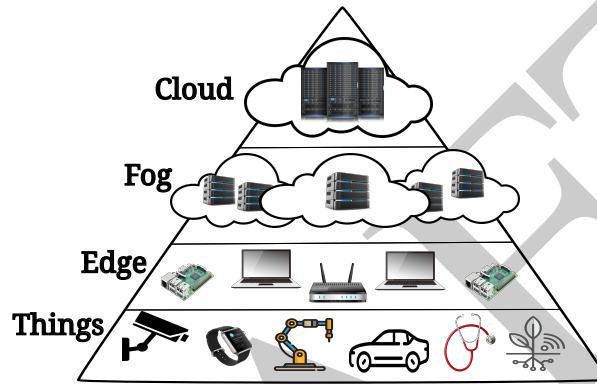


Fig. 1 3-Tiers Architecture of Computing Paradigms

3 Edge Computing Paradigms

Edge computing emerged due to the evolution of cloud computing, and it provides different computing advantages. Several paradigms for operating at the edge of the network have been established throughout the growth of edge computing, including cloudlet and mobile edge computing. The two primary edge computing paradigms are introduced in this section.

3.1 Cloudlet

In 2009, Satyanarayanan and his team first proposed cloudlet computing as a remedy for the problems that can arise while using traditional cloud computing. These restrictions cover things like delay, jitter, and congestion [3]. Cloudlets, as shown in figure 2, are essentially small data centers, often referred to as miniature clouds, which are frequently just a hop away from a user device [1]. Instead of relying on a far-off

"cloud," a nearby cloudlet with abundant resources can be used to alleviate the limited resources of a mobile device. To fulfill the requirement for real-time interactive responses, a solution is to establish a wireless connection with a nearby cloudlet that provides high-bandwidth, one-hop, and low-latency wireless access. In this situation, the mobile device serves as a lightweight client, with the cloudlet in close proximity handling the majority of the complex computational operations[3]. Cloudlet relies on technologies like Wi-Fi, making it reliant on a strong internet connection[19]. Owners of the network infrastructure, such as Nokia and AT & T, can enable cloudlets to be positioned in closer proximity to mobile devices, in hardware with smaller footprints than the cloud computing's massive data centers. As a result of their smaller footprint, cloudlets have less computational power than typical clouds, but they still have advantages over them including lower latency and energy usage [1]. A "data center in a box" is how cloudlets are like. It operates autonomously, with minimal power consumption, and only needs an Internet connection and access control for setup. This administrative simplicity correlates with a device-based computing architecture, enabling easy implementation in a variety of business premises such as doctor's offices or coffee shops. From an internal viewpoint, A cloudlet can be perceived as a cluster of computers equipped with multiple cores, fast internal connections, and a high-bandwidth wireless LAN. For safe implementation in unsupervised areas, the cloudlet can be housed in a protective casing designed to resist tampering with third-party remote monitoring of hardware integrity [3]. To avoid any serious implications in the event of loss or malfunction, it is crucial to emphasize that the cloudlet should only store transient data and code, such as cached copies. In critical circumstances like military operations, disaster-affected areas, and even cyberattacks, cloudlets are essential. In these circumstances, the cloudlet, which is the middle layer in a 3-tier architecture consisting of the mobile, cloudlet, and cloud, is required to fill in and provide vital services, making up for the cloud's unavailability. Cloudlets also have the benefit of reducing the dangers linked to multi-hop networks, such as the possibility of DoS attacks[2].

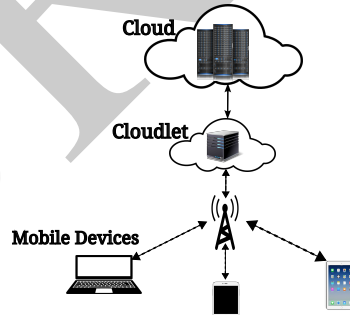


Fig. 2 Cloudle Architecture

3.2 Mobile Edge Computing

The idea behind a cloudlet is to strategically place powerful computers so that they can provide neighboring user equipment (UE) with computing and storage capabilities. Similar to WiFi hotspots, cloudlets deliver cloud services to mobile customers as opposed to offering internet connectivity. The fact that mobile UEs predominantly use WiFi connections to access cloudlets has a potential disadvantage because it requires users to alternate between the mobile network and WiFi anytime they need cloudlet services[16].

Another option that enables cloud computing at the edge is mobile edge computing (MEC), which first announced in 2014 by the European Telecommunications Standard Institute (ETSI). The MEC platform is characterized as a system that provides IT and cloud-computing functionalities within the Radio Access Network (RAN), positioned in close proximity to mobile subscribers [17]. MEC is defined by the ETSI as having low latency, local processing and storage resources, network awareness, and better service quality supplied by mobile carriers. MEC makes mobile end-user services more accessible by providing compute and storage resources. These resources are intended to be deployed on mobile networks near end users. MEC resources can be used in a variety of locations, including radio access networks (RANs), indoor and outdoor base stations, and access points, which connect user equipment to Mobile Network Operators' (MNOs') core networks [18].

Within the radio access network, MEC provides cloud computing capabilities, as depicted in Figure 3. Rather than routing mobile traffic from the core network to end-users, MEC creates a direct link between users and the nearest edge network empowered with cloud services. By deploying MEC at the base station, it enhances computation, mitigates bottlenecks, and reduces the risk of system failure [19]. MEC is implemented on a virtualized platform that takes advantage of the most recent advancements in information-centric networks (ICN), network functions virtualization (NFV), and software-defined networks (SDN). A single-edge device with NFV at its core can provide computational services to numerous mobile devices by producing several virtual machines (VMs). These VMs can handle several tasks or perform diverse network operations at the same time [17].

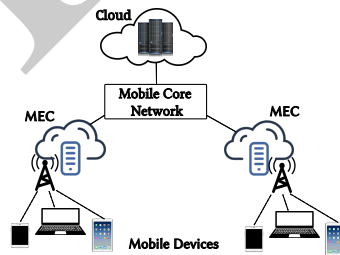


Fig. 3 Mobile Edge Computing Architecture

4 Architecture of Edge Computing-Based IoT

In the context of IoT, edge computing is primarily focused on its implementation across various IoT scenarios, aiming to minimize decision-making latency and network traffic[27]. The edge computing-based IoT architecture, as depicted in figure 4, consists of three distinct layers: IoT, edge, and cloud, all of these are built on top of existing edge computing reference designs. Our primary focus is to define the specific functions allocated to each layer and explore the communication mechanisms established among these layers[27, 29].

1. **IoT layer** : The IoT layer encompasses a broad spectrum of devices and equipment, such as smart cars, robots, smart machinery, handheld terminals, instruments and meters, and other physical goods. These objects are tasked with overseeing the functioning of services, activities, or equipment. Furthermore, The IoT layer consists of actuators, sensors, controllers, and gateways constructed expressly for IoT contexts, which enable the administration of computational resources within IoT devices [27].
2. **Edge layer** : The main purpose of this layer is to receive, process, and send streams of data from the Device Layer. It offers real-time services like intelligent computing, security and privacy protection, and data analysis. Based on the equipment's ability to process data, Three further sub-layers are separated from the Edge Layer: the near-edge layer, the mid-edge layer, and the far-edge layer [29].
 - a) **Far-Edge Layer (Edge controller layer)** : In this layer, data is collected from the IoT layer by edge controllers and subsequently undergoes initial threshold assessment or data filtering. After that, the edge layer or cloud layer directs the control flow back to the IoT layer. After IoT device data has been collected, it is preprocessed to determine thresholds or perform data filtering. Consequently, the edge controllers in this layer must incorporate algorithm libraries tailored to the environment's configuration to consistently improve the strategy's efficiency. Additionally, these edge controllers should convey the control flow back to the IoT layer via the programmable logic controller (PLC) control or action control module after receiving decisions from the edge controller layer or upper layers [27].
 - b) **Mid-Edge Layer (Edge gateway layer)** : This layer is often made up of edge gateways, which can connect to wired networks like industrial ethernet or wireless networks like 5G to receive data from the edge controller layer. Furthermore, the layer enables diverse processing capabilities and caches the accumulated data. Moreover, The edge gateways in this layer play a crucial role in shifting control from the upper layers, such as the cloud layer or edge server layer, to the edge controller layer. Simultaneously, they monitor the equipment in both the edge gateway layer and the edge controller layer[27]. The mid-edge layer has more storage and processing power than the far-edge layer, which can only carry out basic threshold judgment or data filtering. As a result, it can handle IoT layer data in a more thorough manner [29].

- c) **Near-Edge Layer (Edge server layer)** : The edge server layer is equipped with robust edge servers. Within this layer, advanced and crucial data processing takes place. The edge servers leverage dedicated networks to gather data from the edge gateway layer and generate directional decision instructions based on this collected information. Additionally, platform administration and business application management features are anticipated for the edge servers in the edge server layer [27].
3. **Cloud layer** : This layer primarily focuses on in-depth data mining and seeks to allocate resources optimally on a big scale, across a whole organization, a region, or even the entire country. Data from the edge layer is sent to the cloud layer through the use of the public network. Additionally, the edge layer has the ability to receive feedback from cloud layer-provided business applications, services, and model implementations [27].

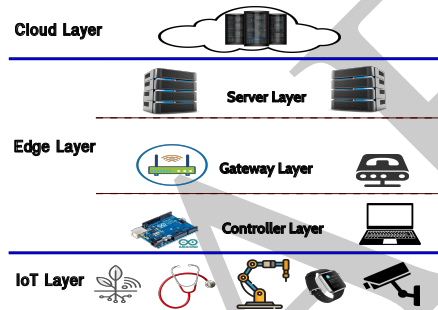


Fig. 4 Edge Computing-based IoT Architecture

5 Advantages of Edge Computing-based IoT

Edge computing plays a vital role as a computing paradigm for IoT devices, involving the utilization of cloud centers located near the IoT devices for tasks such as filtering, preprocessing, and aggregating IoT data [27]. The primary advantages of edge computing include:

- A) **Low Latency**: The close proximity and low latency of edge computing provide a solution to the response delay faced by User Equipment (UEs) while accessing typical cloud services. Edge computing can drastically reduce response time, which includes communication, processing, and propagation delays. Cloud computing typically results in an end-to-end latency of more than 80ms (or 160ms for response delay), making it unsuitable for time-sensitive applications such as remote surgery and virtual reality (VR), which require near-instantaneous replies

within 1ms. Edge computing, on the other hand, benefits UEs by reducing total end-to-end delay and reaction delay due to their close proximity to edge servers. This enhancement enables faster and more efficient interactions for time-critical applications, meeting the requirements for tactile speed and responsiveness[26].

- B) **Energy Saving** : IoT devices often have limited energy supply due to their size and intended usage scenarios, yet they are expected to conduct complicated activities that are frequently power-intensive. It is difficult to design a cost-effective system to properly power numerous distributed IoT devices since regular battery charging or discharging is not always practicable or possible. However, edge computing offers a solution by enabling IoT devices to offload power-consuming computation tasks to edge servers. This not only substantially lowers energy use but also enhances processing efficiency, enabling billions of IoT devices to function optimally [30].
- C) **Security and Privacy** : among the most important features of cloud platform services is enhancing data security and privacy. Customers of these services can obtain centralized data security solutions from these providers, but any compromise of the centralizedly held data may have severe consequences. In contrast, edge computing has the benefit of allowing local deployment of customized security solutions. With this approach, less data transport is necessary because the majority of processing can be done at the network edge. As a result, there is a lower chance of data leakage during transmission, and less data is stored on the cloud platform, lowering the security and privacy risks [27].
- D) **Location Awareness**: Edge servers with location awareness can acquire and handle data generated by User Equipment (UEs) based on their geographical locations. As a result, personalized and location-specific services can be offered to UEs, allowing edge servers to collect data directly from nearby sources without sending it to the cloud. This allows for more efficient and targeted service provisioning customized to specific UE needs[26].
- E) **Reduce Operational Expenses** : Transmitting data directly to the cloud platform incurs substantial operational expenses due to the demands for data transmission, sufficient bandwidth, and low latency. Edge computing, on the other hand, has the advantage of minimizing data uploading volume, resulting in less data transmission, lower bandwidth consumption, and lower latency. As a result, edge computing reduces operational costs when compared to direct data transfer to the cloud platform [27].
- F) **Network Context Awareness**: Edge servers are able to understand the network context through network context awareness. This includes User Equipment (UE) information, such as allocated bandwidth and user locations, as well as real-time network conditions, such as traffic load in a network cell and radio access network specifics. With this invaluable knowledge, edge servers are better equipped to adapt and accommodate to the various UEs and network conditions, which leads to an optimum use of network resources. As a result, edge servers can effectively handle a large amount of traffic, improving network performance. Additionally, the availability of fine-grained information enables the development of services

that are specifically customized to the needs of various traffic flows and individual users [26].

6 Enabling Edge computing-based IoT Technologies

Edge computing-based IoT can be implemented with the integration of several enabling technologies. This section illustrates the relevant enabling technologies by using artificial intelligence and lightweight virtualization as examples.

6.1 Edge Intelligence

As the need for intelligent edge devices has grown, the industry has responded with innovation and the adoption of intelligent edge architectures. These innovative architectures support real-time, mission-critical applications that work with a wide variety of devices. Any machine can qualify as intelligent if it mimics human behaviors and skills including perception, attention, thinking, and decision-making. Machine learning has gained a lot of traction as a field of advancement in artificial intelligence. This has led to a surge in the presence of intelligent devices, fueled primarily by advancements in deep learning techniques [5].

Deep Neural Networks (DNNs) have received substantial attention in the Machine Learning era because of their unrivaled performance across different use cases such as computer vision, natural language processing, and image processing [32]. Notably, deep learning has even outperformed human players in complex games like Atari Games and the game of Go. The integration of deep learning and edge computing holds promise for addressing challenges and opening up new possibilities for applications. On one hand, edge computing applications greatly benefit from the powerful processing capabilities of deep learning, enabling them to handle intricate scenarios like video analytics and transportation control. On the other hand, edge computing offers specialized hardware foundations and platforms, such as the lightweight Nvidia Jetson TX2 development kit, to effectively support deep learning operations at the edge[30]. Many techniques have been introduced to improve the performance of deep learning when performed on edge computing devices, such as :

- A) **Model design:** When machine learning researchers design DNN models for resource-constrained devices, they commonly emphasize creating models with fewer parameters in order to minimize memory usage and execution latency, while still maintaining high accuracy. Several techniques are employed to achieve this, including MobileNets, SSD, YOLO, and SqueezeNet. These methods are aimed at optimizing DNN models for efficient performance on such devices[31].
- B) **Run-Time Optimizations:** Depending on the particular requirements of the application, suitable run-time optimizations can be employed to minimize the quantity of samples that need to undergo processing. For instance, in object

detection applications, a high-resolution image can be divided into smaller images (tiling), and a selection criterion can be used to choose images with high activity regions. This approach allows the design of DNNs that can handle smaller inputs, resulting in improved computational and latency efficiency.

- C) **Hardware:** In the pursuit of accelerating deep learning inference, hardware manufacturers are adopting various strategies. These include utilizing already existing hardware like CPUs and GPUs, as well as developing custom Application-Specific Integrated Circuits (ASICs) dedicated to deep learning tasks, like Google's Tensor Processing Unit (TPU). Additionally, there are novel custom ASICs like ShiDianNao, which prioritize efficient memory access to minimize latency and energy consumption. FPGA-based DNN accelerators also show promise, as FPGAs can deliver fast computation while remaining reconfigurable[31].

6.2 Lightweight Virtualization

Virtualization technologies are widely employed in cloud computing due to their effective method of harnessing the cloud's capabilities by partitioning a physical host into smaller, more manageable virtual components. By leveraging these technologies, cloud computing services become more user-friendly and economically efficient. Hypervisors such as VirtualBox and VMware are frequently used in cloud computing hardware virtualization. However, this approach has limitations such as increased resource cost, longer startup times, and larger attack surfaces. To solve these limitations, lightweight virtualization technologies such as Unikernels and Containers have evolved and are currently used in both cloud and edge computing. These lightweight virtualization technologies offer fast deployment and high efficiency, effectively overcoming the limitations posed by traditional hypervisor-based virtualization[20]. Considering that the computational capabilities of edge computing devices are less potent than data centers, the adoption of emerging lightweight virtualization technologies offers numerous advantages. These benefits encompass swift initialization, minimal overhead, high instance density, and commendable energy efficiency, making them well-suited for the edge computing environment [21]. Lightweight virtualization technology is critical in edge computing because it allows the deployment of resource management, orchestration, and isolation services without the need to account for different hardware configurations. This technology has brought about a significant transformation in software development and deployment practices. Container-based virtualization can be regarded as a lightweight alternative to the traditional Hypervisor-based virtualization [20]. Container-based virtualization offers a different level of abstraction in terms of isolation and virtualization when compared to hypervisors, as illustrated in figure5. Hypervisors virtualize hardware and device drivers, resulting in increased overhead. Containers, on the other hand, isolate processes at the OS level [21]. Containers allow independent applications to be isolated with their own virtual network interfaces, process spaces, and file

systems because they share the same host machine's operating system kernel. Containers allow for a higher number of virtualized instances with lower image volumes, all executing on a single machine thanks to the shared kernel feature [20].

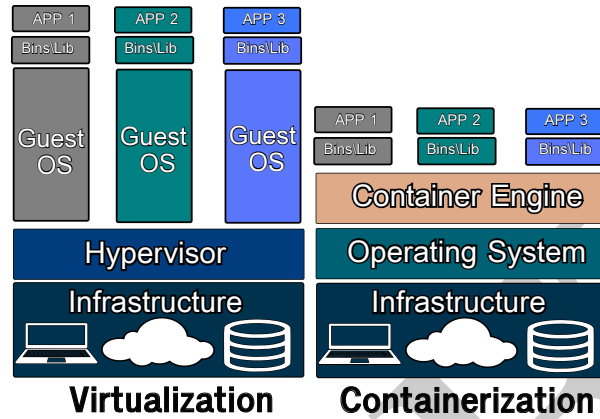


Fig. 5 Virtualization vs Containerization

7 Edge Computing in IoT-Based Intelligent Systems: Case Studies

With the rise of intelligent systems, edge computing offers the most efficient computing and storage solutions for devices with limited computational capabilities. This section delves into the applications of edge computing in IoT-based intelligent systems, including healthcare, manufacturing, agriculture, and transportation. We chose these four case studies because they have a substantial impact on improving human life.

7.1 Edge Computing in IoT-Based Healthcare

The term "geriatric care" refers to a branch of healthcare that emphasizes meeting the special mental, physical, and social needs of the aged. Geriatric care, which is specifically designed to meet the unique demands of elderly individuals, seeks to enhance their general well-being and health while successfully treating age-related illnesses and diseases. Its ultimate goal is to give them the means to maintain their independence, preserve their well-being, and enjoy the greatest degree of comfort as they age [7]. In the area of "Geriatric Care," the danger of falling is regarded as a crucial concern. Unfortunately, many older people fall and unfortunately pass away

because they lack good balance. While the fall itself may be the primary cause, the severity of the outcome stems from the inability to recover, leading to deteriorating physical and cognitive health. Numerous studies back up the idea that elderly people can potentially avoid physical consequences like brain injuries if immediate aid is given within seven minutes of a fall. As a result, the death and disease rates among the aging population would both dramatically decline. In [5] the authors suggest an intelligent edge-monitoring system that utilizes cameras to detect instances of falling. Real-time video analysis is essential for continuously capturing photos, classifying them as either normal (sleeping) or abnormal (falling) circumstances, and instantly sounding an alarm in emergency scenarios. Due to the significant amount of data involved, relying solely on cloud processing would be impractical because of the resulting delays. As a result, cameras serve as IoT devices, gathering data and sending it to nearby edge computing equipment for processing locally. In this approach, the Edge computing server is equipped with Deep Learning models that have already undergone pre-training and are specifically created to detect falls with high accuracy and low latency. They successfully reduce transmission delays by maintaining the computing process' independence from the cloud. This approach not only allows them to generate valuable insights on-site, reducing response time and latency but also addresses privacy concerns by handling sensitive and personal data at the edge.

7.2 Edge Computing in IoT-Based Manufacturing

Regarding intelligent manufacturing, the growing number of terminal network devices has presented new issues in terms of data center operation and maintenance, scalability, and dependability. To tackle these challenges, edge computing has advanced, which moves computation from centralized data centers to the network's edge. This approach enables intelligent services to be deployed near the manufacturing units, meeting essential demands such as highly responsive cloud services, data analytics via edge nodes, and a privacy-policy plan [22]. In [25], the authors introduced a cutting-edge visual sorting method created especially for adaptable manufacturing systems. The system utilizes a visual sorting approach based on CNNs. To support this system, they developed a cloud-edge computing platform, which makes it easier to quickly compute and continuously maintain and improve services.

7.3 Edge Computing in IoT-Based Agricultural

The IoT is typically applied in agricultural development through a monitoring network comprised of a considerable number of sensor nodes. As a result, agriculture increasingly moves away from a production model that is centered on humans and toward one that is centered on information and software [14]. Concerning the notion

of edge computing in the context of Agricultural IoT, numerous researchers have made contributions from diverse perspectives. The authors of [15] developed a system that meets the demanding needs of Precision Agriculture (PA) by integrating automation, IoT technologies, edge and cloud computing through virtualization. Three essential layers make up a multi-tier platform that has been developed: (1) a local layer of Cyber-Physical Systems (CPS) that is connected to agricultural greenhouses. (2) The authors suggest a new edge computing architecture in which control modules are placed on virtualized nodes near the access network. (3) A cloud section outfitted with powerful computing and data analytics tools to help farmers make smart crop management decisions. The entire system was successfully tested in a real greenhouse located in southeast Spain. Through specialized software that was accessible to the end farmers via the platform, this innovation made it possible to control a closed hydroponic system in real-time. To validate the effectiveness of the architecture, two tomato crop cycles were conducted. The results showed remarkable benefits compared to a traditional open crop approach. Significant water savings of over 30% were achieved, which is particularly crucial in their semi-arid region. Additionally, certain nutrients saw improvements of up to 80% thanks to the system's efficient management.

7.4 Edge Computing in IoT-Based Transportation

The Internet of Vehicles (IoV), a new paradigm introduced by the IoT, employs edge computing to offer ground-breaking applications for transportation systems. Using sensors and geofencing technologies, IoV connects various cars with Roadside Units (RSUs) and other vehicles in an Intelligent Transportation System (ITS). Edge cloudlets are used by IoV for service provisioning and orchestration. Currently, substantial research on smart vehicles is being undertaken in both academic and industrial domains [23]. Traffic flow detection plays a vital role in ITS. By obtaining real-time urban road traffic flow data, ITS can intelligently guide measures to alleviate traffic congestion and reduce environmental pollution. In [24], the YOLOv3 (You Only Look Once) model was used by the authors to create a vehicle detecting method. The YOLOv3 model was trained on an extensive dataset of traffic data and subsequently pruned to achieve optimal performance on edge devices. Additionally, by retraining the feature extractor, they improved the DeepSORT (Deep Simple Online and Realtime Tracking) algorithm, enabling multi-object vehicle tracking. Through the integration of vehicle detection and tracking algorithms, they developed a counter for real-time vehicle tracking capable of accurately detecting traffic flow. Finally, the Jetson TX2 edge device platform received and implemented the vehicle detection network and multi-object tracking network.

8 Challenges and Future research Directions

In the preceding section, we outlined four possible uses of edge computing in IoT-based systems. To achieve the full potential of IoT, we emphasize the need for seamless collaboration between IoT devices and edge computing. Now, in this section, we will summarize some of the challenges faced in implementing edge computing in IoT-based systems and propose potential solutions and research opportunities. These include resource allocation, heterogeneity, privacy and security, and microservices.

- **Resource Allocation** : Edge devices play an essential role in enabling latency-critical services. The majority of end IoT devices often experience resource limitations; for example, local CPU computation capabilities and battery capacity are frequently constrained. Some workloads can be offloaded to more powerful edge devices to bypass these constraints and meet the performance requirements of applications. Edge computing improves IoT device capabilities, allowing them to handle more resource-intensive applications. However, Practically speaking, edge computing devices have a finite amount of processing power. As a result, it is unable to handle the massive computing tasks generated by all of the end devices in its service region. As a result, the allocation of resources becomes highly crucial in such environments. Traditional optimization approaches like convex optimization and Lyapunov optimization have been used to tackle the computation offloading problem and find the best scheme. However, these methods are limited when it comes to making optimal decisions in dynamic environments. In contrast, modern resource allocation algorithms, powered by artificial intelligence and deep learning, such as deep reinforcement learning, offer more effective solutions for achieving optimal allocation.
- **Heterogeneity** : The existence of numerous computing technologies in edge computing, such as distinct hardware architectures and operating systems, has made it difficult to develop a viable approach that can be employed with ease in diverse scenarios. This problem can be solved by developing a programming model for edge nodes using software-based techniques, which enables workloads to be executed effectively on numerous hardware configurations simultaneously.
- **Privacy and Security** :At the network's edge, the primary services that require assurance are the protection of usage privacy and data security. When IoT is used in a home, the usage data collected can be used to infer important private information. For instance, examining the patterns of energy or water use can show whether the home is vacant or occupied. This presents a significant challenge in providing services while safeguarding privacy. Having a reliable architecture is crucial before users can feel confident in embracing new technologies. "Privacy by design" can be considered as a reliable approach to enhance security in edge computing. It involves incorporating privacy features directly into the design, taking preventive measures instead of just reacting after privacy breaches, and ensuring data privacy throughout its entire lifecycle.
- **Microservices** : Recently, both edge and cloud services have been changing from monolithic, standalone systems to loosely coupled, independent microservices.

When running complex computations such as deep learning, there are several software requirements, and it's important to find a way to separate different deep learning services when using shared resources. Currently, the microservice framework, which can be used to host complex services on the edge, is still developing and in its early stages. However, it shows great potential for efficiently introducing services in the future.

9 Conclusion

As IoT continues to grow, edge computing is increasingly regarded as a promising and viable solution to address the complexities of managing numerous sensors and devices, along with the demands for resources they require. Edge computing, in contrast to standard cloud computing, involves placing data processing and storage to the edge of the network, bringing them closer to the end users. Thus, by dispersing compute nodes across the network, it is possible to reduce message exchange latency and relieve the computational load on the centralized data center. In conclusion, our chapter has explored the computing paradigms for IoT, edge computing paradigms like cloudlet and MEC, the architecture of edge computing-based IoT, the benefits it offers, and the enabling technologies such as artificial intelligence and lightweight virtualization. Additionally, it presented case studies showcasing how edge computing is applied in intelligent systems based on IoT and highlighted the issues with current research and suggested future directions for further exploration in this field.

References

1. Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., Kong, J., & Jue, J. P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. *Journal of Systems Architecture*. <https://doi.org/10.1016/j.sysarc.2019.02.009>
2. Elazhary, H. (2019). Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions. *Journal of network and computer applications*, 128, 105-140.
3. Satyanarayanan, M., Bahl, P., Caceres, R., & Davies, N. (2009). The case for vm-based cloudlets in mobile computing. *IEEE pervasive Computing*, 8(4), 14-23.
4. Alessio, B., De Donato, W., Persico, V., & Pescapé, A. (2014). On the integration of cloud computing and internet of things. *Proc. Future internet of things and cloud (FiCloud)*, 23-30.
5. Naveen, S., Kounte, M. R., & Ahmed, M. R. (2021). Low latency deep learning inference model for distributed intelligent IoT edge clusters. *IEEE Access*, 9, 160607-160621.
6. Edge Computing Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services, Edge-managed Platforms), By Application, By Industry Vertical, By Region, And Segment Forecasts, 2023 - 2030. Accessed: July 15, 2023. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/edge-computing-market>
7. Paulauskaite-Taraseviciene, A., Siaulyis, J., Sutiene, K., Petravicius, T., Navickas, S., Oliandra, M., ... & Balciunas, J. (2023, April). Geriatric Care Management System Powered by the IoT and Computer Vision Techniques. In *Healthcare* (Vol. 11, No. 8, p. 1152). MDPI.

8. Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. *Future Generation Computer Systems*, 97, 219-235.
9. Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2017). A survey on the edge computing for the Internet of Things. *IEEE access*, 6, 6900-6919.
10. Lorenzo, B., Garcia-Rois, J., Li, X., Gonzalez-Castano, J., & Fang, Y. (2018). A robust dynamic edge network architecture for the internet of things. *IEEE network*, 32(1), 8-15.
11. Delfin, S., Sivasanker, N. P., Raj, N., & Anand, A. (2019, March). Fog computing: A new era of cloud computing. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1106-1111). IEEE.
12. Atlam, H. F., Walters, R. J., & Wills, G. B. (2018). Fog computing and the Internet of Things: A review. *Big Data and Cognitive Computing*, 2(2), 10.
13. Alwakeel, A. M. (2021). An overview of fog computing and edge computing security and privacy issues. *Sensors*, 21(24), 8226.
14. Zhang, X., Cao, Z., & Dong, W. (2020). Overview of edge computing in the agricultural internet of things: key technologies, applications, challenges. *Ieee Access*, 8, 141748-141761.
15. Zamora-Izquierdo, M. A., Santa, J., Martínez, J. A., Martínez, V., & Skarmeta, A. F. (2019). Smart farming IoT platform based on edge and cloud computing. *Biosystems engineering*, 177, 4-17.
16. Mach, P., & Becvar, Z. (2017). Mobile edge computing: A survey on architecture and computation offloading. *IEEE communications surveys & tutorials*, 19(3), 1628-1656.
17. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). Mobile edge computing: Survey and research outlook. *arXiv preprint arXiv:1701.01090*, 1-37.
18. Haibeh, L. A., Yagoub, M. C., & Jarray, A. (2022). A survey on mobile edge computing infrastructure: Design, resource management, and optimization approaches. *IEEE Access*, 10, 27591-27610.
19. Abbas, N., Zhang, Y., Taherkordi, A., & Skeie, T. (2017). Mobile edge computing: A survey. *IEEE Internet of Things Journal*, 5(1), 450-465.
20. Chen, S., & Zhou, M. (2021). Evolving container to unikernel for edge computing and applications in process industry. *Processes*, 9(2), 351.
21. Morabito, R., & Beijar, N. (2016, June). Enabling data processing at the network edge through lightweight virtualization technologies. In *2016 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)* (pp. 1-6). IEEE.
22. Chen, B., Wan, J., Celesti, A., Li, D., Abbas, H., & Zhang, Q. (2018). Edge computing in IoT-based manufacturing. *IEEE Communications Magazine*, 56(9), 103-109.
23. Rafique, W., Qi, L., Yaqoob, I., Imran, M., Rasool, R. U., & Dou, W. (2020). Complementing IoT services through software defined networking and edge computing: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3), 1761-1804.
24. Chen, C., Liu, B., Wan, S., Qiao, P., & Pei, Q. (2020). An edge traffic flow detection scheme based on deeplearning in an intelligent transportation system. *IEEE Transactions on Intelligent Transportation Systems*, 22(3), 1840-1852.
25. Wang, Y., Hong, K., Zou, J., Peng, T., & Yang, H. (2019). A CNN-based visual sorting system with cloud-edge computing for flexible manufacturing systems. *IEEE Transactions on Industrial Informatics*, 16(7), 4726-4735.
26. Hassan, N., Yau, K. L. A., & Wu, C. (2019). Edge computing in 5G: A review. *IEEE Access*, 7, 127276-127289.
27. Fazeldehkordi, E., & Grønli, T. M. (2022). A survey of security architectures for edge computing-based IoT. *IoT*, 3(3), 332-365.
28. Sha, K., Yang, T. A., Wei, W., & Davari, S. (2020). A survey of edge computing-based designs for IoT security. *Digital Communications and Networks*, 6(2), 195-202.
29. Qiu, T., Chi, J., Zhou, X., Ning, Z., Atiquzzaman, M., & Wu, D. O. (2020). Edge computing in industrial internet of things: Architecture, advances and challenges. *IEEE Communications Surveys & Tutorials*, 22(4), 2462-2488.
30. Wang, F., Zhang, M., Wang, X., Ma, X., & Liu, J. (2020). Deep learning for edge computing applications: A state-of-the-art survey. *IEEE Access*, 8, 58322-58336.

31. Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, 107(8), 1655-1674.
32. Marchisio, A., Hanif, M. A., Khalid, F., Plastiras, G., Kyrkou, C., Theodorides, T., & Shafique, M. (2019, July). Deep learning for edge computing: Current trends, cross-layer optimizations, and open research challenges. In *2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)* (pp. 553-559). IEEE.
33. Idrees, A. K., Alhussaini, R., & Salman, M. A. (2020). Energy-efficient two-layer data transmission reduction protocol in periodic sensor networks of IoTs. *Personal and Ubiquitous Computing* 27 (2), 139-158.
34. Shawqi Jaber, A., & Kadhum Idrees, A. (2020). Adaptive rate energy-saving data collecting technique for health monitoring in wireless body sensor networks. *International Journal of Communication Systems*, 33(17), e4589.
35. Alhussaini, R., Idrees, A. K., & Salman, M. A. (2018, September). Data transmission protocol for reducing the energy consumption in wireless sensor networks. In *International conference on new trends in information and communications technology applications* (pp. 35-49). Cham: Springer International Publishing.
36. Idrees, A. K., Abou Jaoude, C., & Al-Qurabat, A. K. M. (2020, October). Data reduction and cleaning approach for energy-saving in wireless sensors networks of IoT. In *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (pp. 1-6). IEEE.
37. Idrees, A. K., & Khelif, M. S. (2023) Lossless EEG Data Compression using Clustering and Encoding for Fog Computing based IoMT Networks, *International Journal of Computer Applications in Technology* 72 (1), 77-83.
38. Idrees, A. K., & Khelif, M. S. (2023). A new lossless electroencephalogram compression technique for fog computing-based IoHT networks. *International Journal of Communication Systems*, e5572, <https://doi.org/10.1002/dac.5572>.
39. Idrees, A. K., & Khelif, M. S. (2023). Efficient compression technique for reducing transmitted EEG data without loss in IoMT networks based on fog computing. *The Journal of Supercomputing*, 79(8), 9047-9072.
40. Idrees, A. K., Ali-Yahiya, T., Idrees, S., & Couturier, R. (2022, September). DaTOS: Data Transmission Optimization Scheme in Tactile Internet-based Fog Computing Applications. In *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)* (pp. 01-06). IEEE.
41. Idrees, A. K., Idrees, S. K., Couturier, R., & Ali-Yahiya, T. (2022). An edge-fog computing-enabled lossless EEG data compression with epileptic seizure detection in IoMT networks. *IEEE Internet of Things Journal*, 9(15), 13327-13337.
42. Idrees, A. K., & Jawad, L. W. (2023). Energy-efficient Data Processing Protocol in edge-based IoT networks. *Annals of Telecommunications*, 1-16, <https://doi.org/10.1007/s12243-023-00957-8>.