IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 14, 2019, accepted September 3, 2019, date of publication October 7, 2019, date of current version November 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945338

Understanding Edge Computing: Engineering Evolution With Artificial Intelligence

JUN-HO HUH[1] AND YEONG-SEOK SEO[2], (Member, IEEE)
[1]Department of Data Informatics, Korea Maritime and Ocean University, Busan 49112, South Korea
[2]Department of Computer Engineering, Yeungnam University, Gyeongsan 38541, South Korea

Corresponding author: Yeong-Seok Seo (ysseo@yu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2017R1C1B5018295).

ABSTRACT The key to the explosion of the Internet of Things and the ability to collect, analyze, and provide big data in the cloud is edge computing, which is a new computing paradigm in which data is processed from edges. Edge Computing has been attracting attention as one of the top 10 strategic technology trends in the past two years and has innovative potential. It provides shorter response times, lower bandwidth costs, and more robust data safety and privacy protection than cloud computing. In particular, artificial intelligence technologies are rapidly incorporating edge computing. In this paper, we introduce the concepts, backgrounds, and pros and cons of edge computing, explain how it operates and its structure hierarchically with artificial intelligence concepts, list examples of its applications in various fields, and finally suggest some improvements and discuss the challenges of its application in three representative technological fields. We intend to clarify various analyses and opinions regarding edge computing and artificial intelligence.

INDEX TERMS Edge computing, cloud computing, Internet of Things, artificial intelligence, engineering.

I. INTRODUCTION

For over a decade, centralized cloud computing has been considered as a standard information technology (IT) platform and a new paradigm. Cloud computing is a service that enables users who access the Internet to access computing resources centered on storage [1]–[7]. Cloud computing can be used to efficiently manage the resources stored on a centralized cloud server and utilize those resources without the limitations of time and space. In addition to processing engines such as the Google File System [8], MapReduce [9], Apache Hadoop [10], and Apache Spark [11], which support cloud services, scalable infrastructures also have a major impact on how businesses operate. However, as the technology of future innovation has been developed recently, the disadvantages of cloud computing are revealed, and also new requirements for the high technology are needed.

With the recent increase in Internet of Things usage [12], the number of devices connected to the Internet is increasing daily. In 1992, the number of connected devices reached 1 million, and in 2003, notebook usage increased to more than 500 million. Based on the inclusion of wearable devices in 2012, this figure became 8.7 billion. In 2013, this number became 11.2 billion owing to connected home appliances, and in 2014, 14.4 billion owing to the usage of smart devices. By 2020, the number of devices connected to the Internet is expected to reach approximately 50 billion [13], [14].

This will make it impossible to process the data from this large number of Internet of Things devices, irrespective of how vast the cloud server is [15]. As the cloud usage increases, the time required to process the data also increases, thus resulting in longer latency for users, which in turn increases the load on the server and network. In addition, cloud computing is vulnerable to security and network environments. As mentioned above, when a cloud data center that stores important information of a growing number of devices is attacked, extensive information leakage occurs [16]–[20].

As a solution to these problems, a new computing technology called edge computing is attracting attention. Edge computing is known for its use of a variety of concepts such as fog computing [21]–[23], cloudlet [24]–[26], and mobile edge computing [27]–[29]. Edge computing is a technology that instantly analyzes and processes data at the edge of the network where the data is collected. It is not a physically remote data center that processes and computes data. It is a

The associate editor coordinating the review of this manuscript and approving it for publication was Joel Rodrigues.

technology wherein devices are located or data is analyzed and processed within a terminal device. In this manner, data is processed in real-time at physically close edges in order to support data flow acceleration. This greatly reduces the data latency and allows users to provide a fast service [30]–[38].

Fig. 1 shows the difference between cloud computing and edge computing. Cloud computing is responsible for all the tasks that IoT devices perform through the cloud. Edge computing, on the other hand, works by distributing the work of the device to the edge node near the physical distance as well as the cloud. Edge computing, in contrast to cloud computing, does not have centralized computing resources, and thus, the data overload is reduced, which can in turn reduce the backhaul traffic to the central repository. In addition, important data such as personal data can be processed without having to send it to the cloud server, which enhances data security as sensitive information is processed at the edge, and only the remaining data is encrypted and transmitted to the cloud server [15], [30], [31], [39], [40].

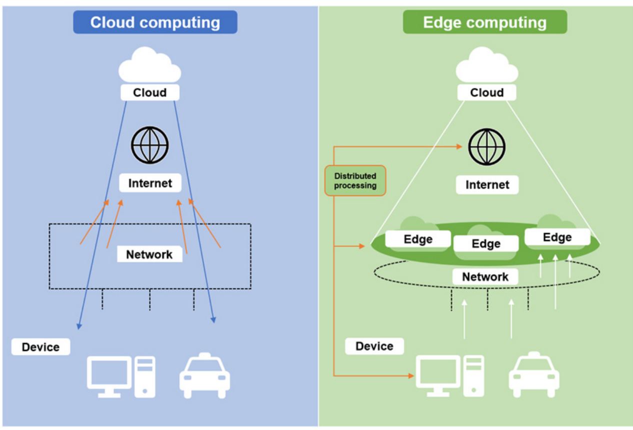

FIGURE 1. **Paradigm of cloud computing and edge computing.**

With these advantages, edge computing is suitable for application in technologies that make use of a vast amount of data such as autonomous vehicles, smart factories, and mobile applications, which are currently being developed, and which are important for real-time processing. Recently, there has been a trend of companies focusing on edge computing technology rather than cloud computing, and Gartner and Trendforce have selected edge computing as a promising future technology.

The contributions of this research work are:

- The concept of currently popular edge computing and AI-based edge computing have been described more clearly while considering the architectures, technologies, and challenges related to the edge computing technology.
- The direction of future study has been pointed out after understanding the trends of existing studies through most recent research surveys conducted for edge computing.
- The concept and operating mode of edge computing have been described in an easier and clearer way, while the comparative analysis between cloud computing and edge computing.
- The concepts of typical technological fields (i.e., Cloud offloading, Video content analysis, Autonomous Vehicle, Smart Home, Smart Factory, Smart City, and MEC) which can be utilized for edge computing were introduced along with the application forms and the possible methods of utilizing the edge computing technology by analyzing the relevant contents in detail.
- Corresponding to the core keyword 'Convergence' in the era of the 4th Industrial Revolution, the technological fields (i.e., MEC, IoT, AI-based autonomous vehicles) which are being considered as some of the typical challenges associated with technological convergence of edge computing have been explained together with their current issues, existing solutions, and the future research direction.

This paper is organized as follows: Section 2 introduces the concept and architecture of edge computing. Section 3 introduces the techniques comprising the use of edge computing, and Section 4 discusses the challenges faced with the implementation of edge computing. Section 5 concludes this paper. Please note that the abbreviation and acronyms used in the paper are presented in Table 1.

TABLE 1. **Abbreviations and acronyms.**

Abbreviations and acronyms	Description
5G	Fifth-Generation wireless communication
ACO	Ant Colony Optimization algorithm
AI	Artificial Intelligence
AVE	Autonomous Vehicle Edge
CCTV	Closed-circuit television
CIoT	Cloud-centric Internet of Things
CMU	Central Monitoring Unit
CPU	Central Processing Unit
edgeOS	Edge operating system
ESS	Energy Storage System
FEMS	Factory Energy Management System
GPS	Global Positioning System
ICT	Information and Communications Technology
IMU	Inertial Measurement Unit
IoT	Internet of Things
LAN	Local area network
LIDAR	Light Detection And Ranging
MDP	Markov Decision Process
MEC	Mobile Edge Computing
QoE	Quality of Experience
RAN	Radio Access Network
VCA	Video Content Analysis
WAN	Wide area network
XaaS	X as a Service (Anything as a Service)

II. CONCEPTS AND ARCHITECTURE

Edge computing is a concept similar to distributed cloud computing, but as shown in Fig. 1, there are differences between these two methods, in that, the physical distance to which the data is transmitted. If cloud computing is a method used for communicating directly with a central data center wherein computing resources are stored, edge computing comprises communication primarily with a so-called edge data center (edge server [15]) located close to the endpoint device and processes secondary work (storage functions, etc.) on cloud server.

Many people are familiar with edge computing concepts such as fog computing and cloudlets. "Fog computing" means mist-distributed computing technology (as can be interpreted from the meaning of the word "fog"), and the term "cloudlet" is used to indicate platforms [21]–[26].

Figs. 2 and 3 show that there is little difference between the fog computing and cloudlet paradigms. Another concept, MEC has a paradigm similar to that shown in Fig. 2. Eventually, the primary work is executed on the edge server close to the terminal device having the same structure as edge computing, and the secondary work is processed in the central cloud.

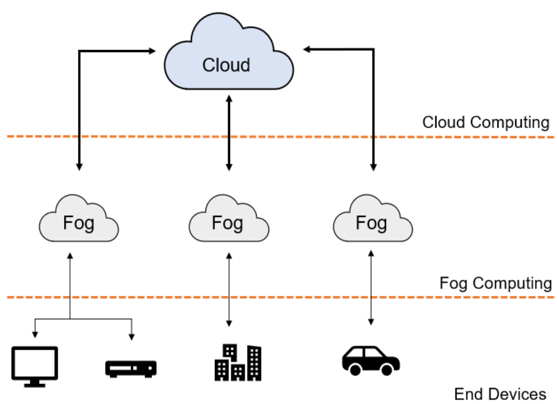

FIGURE 2. Paradigm of fog computing.

As shown in Figs. 2 and 3, edge computing divides cloud computing into two categories. We can consider the first process to be a complementary technology to the centralized cloud computing by processing the secondary process in the subordinate distributed cloud in the upper central cloud. Often two computing technologies seem to compete in a completely different manner, but in fact, they are closer to symbiosis than competition.

Table 2 is a comparison between cloud computing and edge computing. The major difference is that the former is a centralized computing system whereas the latter is a distributed computing system. Such a structural difference makes the differences described in this table.

To explain the principle of service provision, cloud computing allows the user to perform his/her task by using the data center through internet and edge computing lets the user use the nearby local network device (edge) [16]. Although

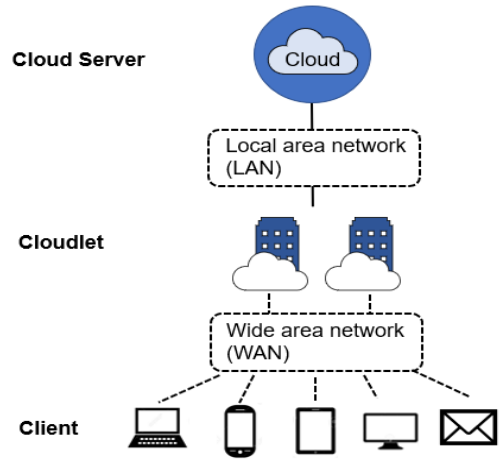

FIGURE 3. Paradigm of cloudlet.

TABLE 2. Comparison of cloud computing and edge computing.

	Cloud computing	Edge Computing
Geo-distribution	Centralized	Distributed
Location of Service	Within the Internet	Edge of the local network
Scalability	Easy to scale	Less scalable than fog computing
Location-aware	No	Yes
Data load	High	Low
Security	Undefined	Can be defined
Latency / Response time	Long	Short
Power consumption	High	Low
No. of server nodes	Few	Very large
Mobility	Limited	Supported
Distance between client and server	Multiple hops	One hop

the cloud computing system consumes more power and its data load can be heavy as all the data has to be processed at the data center, it is relatively easier to expand the system. In contrast, the edge computing system has to provide an edge to each area but requires rather a small amount of power and its data load is lighter since it just needs to process the local user's data only. In addition, location-aware function is possible by providing a service from the edge located close to the users. Because of its short physical distance between edges and users, a shorter latency and response time can also be expected.

In terms of security, it is logical to assume that the edge computing system having a distributed form is more secure than the cloud computing system where all the data is being gathered in one place. Thus, it will be possible to increase system security by processing some of the sensitive data at the edge and send the rest to the data center after encrypting them. Additionally, the stability of the infrastructure adopting the edge computing system will be increased more since data can be processed at the edge without any interruption, maintaining some of the services. The cloud computing system cannot achieve this as all the IoT equipment connected to the central data center will cease to operate when it shuts down for some reason.

Fig. 4 is a three-tiered description of the architecture of edge computing. As shown in Fig. 4, edge computing consists of the cloud, edge, and device tiers. The device tier is a user-used device, and the green block in the edge tier indicates the edge server. The edge server collects, aggregates, analyzes, and processes data before overloading the data to the cloud tier. When a user requests a service through the device, the data is collected, aggregated, analyzed, and processed by the edge server before it is distributed to the cloud server. If the edge server alone can complete the work, it will process the data in real time and provide the required services; however, if this is not the case, the same tier will be offloaded to another server or offloaded to the cloud [15].

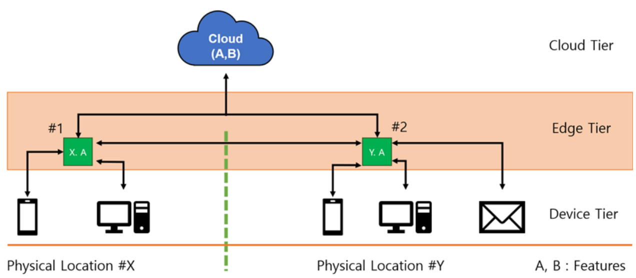

FIGURE 4. Three-tiered architecture of edge computing.

Let us assume that the cloud in Fig. 4 provides functionality A and B; if one of the devices in physical location X is attempting to perform Task B, activity A is already performed at the edge server in # 1, and thus, the data regarding Work B must be offloaded to another edge server or cloud. However, in # 2's edge server, the data is offloaded to the cloud because the work A for the physical location Y is being performed. The offloaded data is processed in the cloud and sent back to the end device.

In this manner, the edge has a two-way computing stream that can perform off-roading, data storage, caching, and processing, but there are aspects to consider before off-loading the data. First, the edge itself must meet the requirements of services, such as reliability, security, and privacy, while considering the priority of the task, the resource utilization of the server, the cost of computing, and the physical distance or distance cost of the server used [15].

III. APPLICATION FIELDS AND TECHNOLOGIES

With the advent of edge computing, existing technologies may develop due to edge computing, and new technologies based on edge computing have been developed. This section introduces technologies related to edge computing and describes how the technology can be changed by using edge computing.

A. CLOUD OFFLOADING

Fig. 5 presents the architecture of the mobile cloud computation offloading system with three main components: resource monitoring, cost, and partitioning. Resource monitoring models are used to collect data regarding CPU usage, battery level, and speed. The partition model divides the application's configuration classes into remote partitions and local partitions, which are offloaded to the cloud and the latter runs locally on the mobile device. The cost model is the most important, and the off-road decision is determined based on six criteria—energy, price, storage cost, performance, robustness, and security—as shown in Fig. 5 [41].

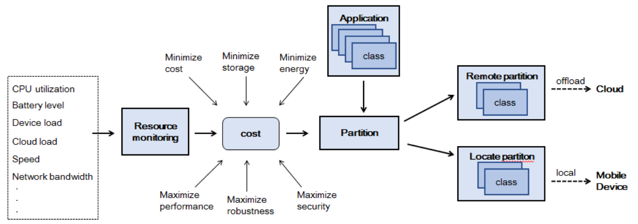

FIGURE 5. Architecture of cloud offloading system.

In this paradigm of cloud computing, most of the calculations are performed in the cloud, which requires more time (e.g., long tail waiting time) than edge computing. To date, we have dealt with cloud off-roading as a side of energy performance compromise in a mobile cloud environment. Edge computing has a specific computational resource in edges and provides an opportunity to offload some of the work in the cloud [42]–[46]. The evolution of distributed computing environments has resulted in the development of many technologies that facilitate the division of tasks that can be performed in multiple geographical locations (for example, by dividing workflows into different locations).

However, now that data is not only cached on the edge server, the task of processing the data must also be cached in the edge. When the task is cached in the edge, there exists a problem that the edge node must be performed in an automated manner. Therefore, it is necessary for developers who can use edge nodes to develop a scheduler that defines the computational pipeline and distributes the divided work to the edge node for flexibility of the work.

The program that can benefit the most from the development of edge computing is online shopping services. By default, all changes to online shopping are made in the cloud, and new shopping items are updated on the customer's device. This process takes time depending on the speed of the network and the load level of the server. However, as the bandwidth of the mobile network is relatively low, the delay time may be longer if online shopping is used on a mobile device. Thus, the waiting time will be greatly reduced if the online shopping cart updates are offloaded from the cloud servers to the edge nodes.

When edge computing is applied to online shopping services, all the shopping cart data and related tasks (such as adding, updating, deleting, and modifying items) can be performed on the edge node. However, the data of the edge node must be synchronized with the cloud server, but this can be performed in the background [30].

Other applications of edge computing are as follows.

- The navigation program allows navigation or searching services to be performed at the edge of the local area, where only some map blocks are involved.
- In content filtering/counting, you can perform tasks such as filtering/counting on edge nodes to reduce the amount of data being sent.
- Vision-aid entertainment games and mobile applications that require real-time requests can also be responded to promptly using the edge node.

Another problem in edge computing is that when a user moves from one edge node to another, it involves the collaboration of several edges. A simple solution to this is to cache data on all edges that you reach, but this creates a synchronization problem between the next edge nodes. If these issues are resolved, it will become possible to provide users with a higher service quality owing to the reduction in the waiting time.

B. VIDEO CONTENT ANALYSIS

VCA, also known as intelligent video analysis, is based on the analysis of visual scenes (such as CCTV images). As described above, the objective of the VCA is to understand visual scenes in order to learn, interpret, and extract meaningful information from the video sequence. The applications of VCA include video search, human detection and tracking, anomalous behavior detection, real-time vehicle monitoring and traffic control, surveillance, face recognition, and fire detection.

A typical VCA system consists of a video collection device, a video storage and display device, and a processing device network, as shown in Fig. 6. The camera collection works within the application for visual data collection. The visual data of the camera can be partially analyzed within the camera before being transmitted to the CMU and can be displayed on the display device without processing it. Video streams are stored separately in storage devices for subsequent playback. Processing units facilitate the analysis of video streams using VCA software. The results of the analysis are stored in storage, displayed on the monitor, or used to take action on the results [47].

The following are the details for the performance of the VCA.

- *Ease of deployment*: The system's users can easily configure the VCA application; they can fine-tune the parameters of the performance measurements and applications. The interface should be easy for users to use, especially when multiple video streams are being monitored simultaneously, such as when alarms and triggers are required.
- *Computational efficiency*: This indicates the ability to achieve a high accuracy without increasing the load of the video stream analysis work to a certain extent, and the system should have a high reliability and availability. This affects the processor and scalability with an increase in the frame size, speed, and number of channels in the video streams.
- *Real-time processing*: It is a function that matches the real-time responses to scenarios of other applications. For example, in order to analyze the video stream and take necessary measures to prevent crime, the necessary work must be performed in a short time and a separate alarm must be set up to inform the person concerned about the incident.
- Cost-*efficiency*: Cost-effective VCA systems include the parameters of the video stream analysis algorithm, and the scaling of the number of processors and system components.

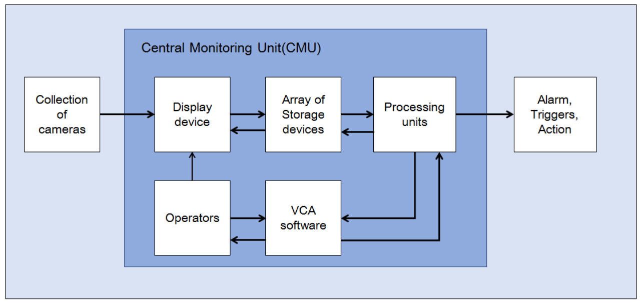

FIGURE 6. Components of video content analysis system.

The recent increase in the use of mobile phones and network cameras has popularized video analysis as a new and innovative technology. Cloud computing is unsuitable for applications that require video analysis owing to data transmission latency and privacy issues—a typical example of such an application is finding a person. When a person is missing, the camera captures the missing scene. However, camera data is not generally uploaded to the cloud owing to privacy issues or costs, which makes it difficult to use visual data obtained from a lot of cams. Even if the visual data is uploaded to the cloud, a large amount of time is required to upload, search, and analyze vast amounts of visual data [48].

On using edge computing, visual data is received from the edge servers that are physically closer to the cloud, analyzed, and responded to in real time, and the above disappearance or other crime cases are quickly processed. In addition, as bidirectional streams are possible, the analyzed results can be reported to the cloud again.

C. AUTONOMOUS VEHICLE (SELF-DRIVING CAR)

Automatic vehicles are cars that recognize the driving environment without human intervention, judge the risk, plan the driving route, and operate themselves. They comprise obstacle detection systems, central control devices, and actuators and require the use of advanced technologies such as robotics and computer engineering, global positioning system, geographic information system, precision sensors, and electronic control.

The simple system structure of the autonomous vehicle system is presented in Fig. 7 [49]. As shown in Fig. 7, the required functions in autonomous vehicles are divided into two main functions: the function of detecting danger and that of controlling the vehicle. The ability to detect danger includes the functions of lane departure warning, speed limit indication, forward collision warning, automatic emergency braking, adaptive cruise control, and traffic jam assist.

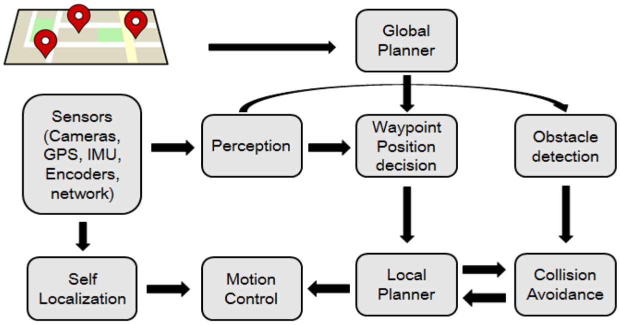

FIGURE 7. System architecture of autonomous vehicle.

Recently, the development of such deep-learning technology has been progressing rapidly, which is accelerating the realization of autonomous driving cars. However, the current reliability and accuracy of this technology are insufficient for autonomous driving cars to perform with the same reliability and accuracy as that exhibited by people. This is because the sensor used cannot collect and process data at the same level as humans in the process of collecting information regarding the driving environment as well as owing to the limit of the extant AI technology [50].

Edge computing technology is a key strategy of 5G, a network technology that is currently attracting attention as a technology that is optimized for data collection and processing. Edge computing provides an optimal service environment for real-time access to wireless network information and location recognition as well as a short waiting time and high bandwidth. If 5G is realized by the advantage of edge computing, it can provide high service by analyzing data obtained from autonomous vehicle sensor in a short time. If edge computing technology and AI technology are developed together in the future, autonomous driving vehicles will be realized [51], [52].

D. SMART HOME

Smart home technology can be used to monitor and control various devices such as home appliances, energy consumption devices, and security devices by connecting them to a communication network. A smartphone or AI speaker is used to recognize the user's voice, connects all the IoT devices in the house, and operates them automatically or remotely.

In order to realize a smart home, we use machine learning, connectivity, voice recognition, and big data technology in terms of software, which is a vital to providing personalized services to users. In the past, if you thought about the optimized method and movement line for the user, you now learn individual behavior patterns and present the best way to consider the user's habits and usage space. Connectivity is an essential element of connecting multiple products in the house in order to facilitate control at a single contact point. This function is performed by smart speakers or certain home appliances that can recognize mobile apps or voice and by special edgeOS. If the edge gateway runs a special edge operating system at home, it can easily connect and manage home appliances and process data from the edge to reduce the burden on the network bandwidth [30]. An edgeOS uses multiple communication networks to collect and fuse data obtained from various sources in order to deliver information to users.

Gartner, a global ICT research firm, expects global IoT equipment usage to increase to 8.4 billion units in 2017 and reach 20.4 billion units by 2020. Big data, one of the key driving forces of the AI-based Internet, is also an essential element affecting this increase. Furthermore, computing technology is required to manage big data in the IoT environment and to achieve business innovation through machine learning. Therefore, there is an operation processor in IoT devices, and edge computing technology that can respond to data input/output in real time is attracting attention. Edge computing technology responds in real time and ensures stability because it analyzes and processes data on the network edge.

Amazon Echo accounted for 67% of the global smart speaker market—as shown in Table. 3 [53]—because it was an excellent scalable and open platform for developers, and it was also able to effectively utilize edge computing to meet the varied and diverse needs of users. It is a tool that can be connected with and controlled using electronic products in the home, and it is advantageous to add only new products to control new products using Amazon Eco.

Each technology that involves the implementation of the smart home ecosystem may not be very new. However, it is also true that technologies such as machine learning, AI, and voice recognition are based on the smart home market. While global companies are intent of developing smart homes, it is necessary to note that edge computing is still problematic, and there is plenty of room for improvement.

E. SMART FACTORY

A smart factory is a factory wherein sensors (used for IoT) are implemented in the factory facilities and machines, and data can be collected and analyzed in real time, such that all situations in the factory can be observed at a glance and analyzed and controlled according to the requirement.

As the recession of the manufacturing industry has had a great impact on the economy owing to the global

TABLE 3. Global market share of smart speakers by manufacturer.

(Shipments in Millions of Units)

Vendor	Q3 '17 Shipments	Q3 '17 Market Share	Q3 '16 Shipments	Q3 '16 Market Share	Growth Y/Y
Amazon	5.0	66.9%	0.9	93.5%	478%
Google	1.9	25.3%	0.0	0.0%	-
JD.com	0.1	1.6%	0.0	0.0%	-
Xiaomi	0.1	1.3%	0.0	0.0%	-
Alibaba	0.1	0.9%	0.0	0.0%	-
Others	0.3	3.9%	0.1	6.5%	383%
Totals	**7.4**	**100.0%**	**0.9**	**100%**	**708%**

financial crisis, the world is paying attention to the importance of manufacturing and making alternatives such as large-scale investment and policy establishment for manufacturing innovation. The fourth industrial revolution is accelerating and creating corporate competitiveness by combining manufacturing and ICT, and thus, building a smart factory, which would be a comprehensive solution for creating new value beyond manufacturing, is essential.

The objective of the smart factory is to connect industrial devices and the entire production process to the network and to build a flexible response system for addressing customer needs. We thus need to systemize comprehensive, real-time interactions between humans and automated machines and integrate forecast simulations and pre-stage engineering to improve the production efficiency as well as market and customer-customized production systems. Technical overview is presented in Fig. 8.

When the smart factory is implemented, it has a data-based factory operating system that analyzes data and makes decisions based on the data collected from each factory. Thus, you can derive a correlation between the worksite situation and the problem, identify the causes of sudden failure or quality defects, and respond immediately. Edge computing technology can be used to respond in real time. It is suitable for a manufacturing environment wherein real-time data processing is important, and it does not require broadband to collect and process data directly from each edge. It is a new alternative to the construction of a smart factory because of its high response speed. In addition, the overload of data is greatly reduced, and problems that have occurred in the existing bandwidth can be alleviated.

The introduction of the smart factory would be a shortcut to securing and improving corporate competitiveness by enabling customized production processes, multi-product production, procurement and logistics innovation, and machine and human collaboration.

F. SMART CITY

The smart city is an urban area in which various types of electronic data collection sensors are used to provide the information necessary for the efficient management of assets and resources [15]. This includes data collected from citizens, devices, and assets, which is then processed or analyzed to monitor and manage transportation and transportation systems, power plants, water supply networks, waste management, law enforcement, information systems, schools, libraries, hospitals, and other community services. The concept of the smart city involves the integration of the Internet and information and communication technologies,

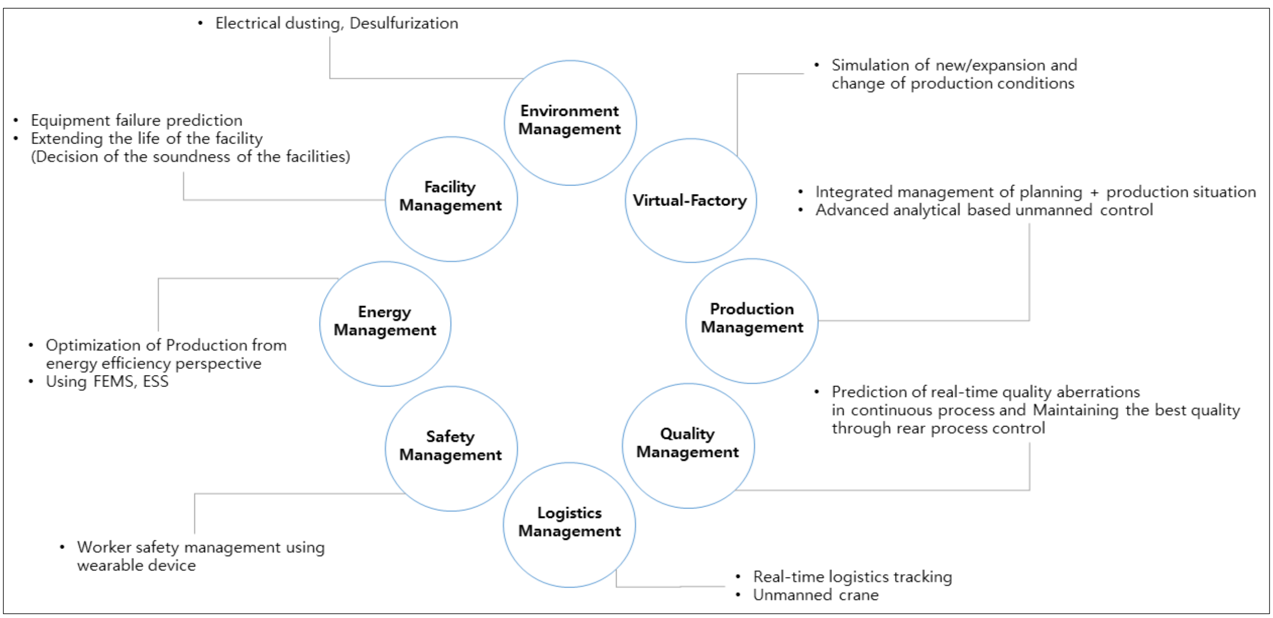

FIGURE 8. Technical range of smart factory.

which comprise various physical devices connected to the network to optimize the efficiency of the city operations and services and to connect with citizens. An example of the smart city model is presented in Fig. 9.

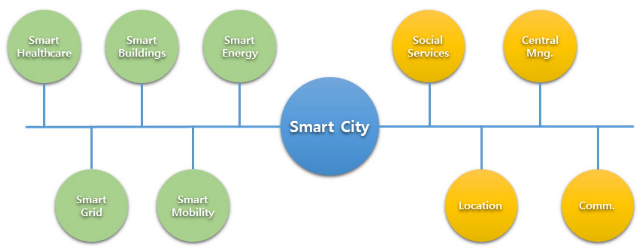

FIGURE 9. View of the smart city model.

As previously stated, the smart city functions provide urban mobility and urban resilience as well as aid in solving local community issues such as lighting, parking, waste, and public safety. Past and real-time data can be used for enabling urban mobility by facilitating the observation of traffic and population movements from a global perspective and for providing important information regarding future economic and infrastructure plans. Urban elasticity can be facilitated through the collection and analysis of data obtained from many suppliers to realize complete visibility and to manage artificial and natural risk factors based on this information. This facilitates super-localized monitoring that can be used to measure toxic-gas, fire, flood, and noise levels and prevent using simulations and real-time forecasting [54].

In the case of a smart city, it is necessary to collect, analyze, and process large amounts of data to perform all the aforementioned functions, which causes data flooding. Data floods are a challenge that is required to be steadily alleviated in order to realize sustainable smart cities, which are difficult and costly to manage. Edge analysis automatically computes and processes the data collected from the sensors, network switches, and other devices using the edge analysis algorithm in order to determine the information value and send or discard information to the cloud. This reduces the decision-making time of the connected device. Edge analysis requires hardware and software platforms to store, prepare, educate, and process algorithms and to process and store data.

However, edge analysis alone is insufficient for realizing the smart city of cutting-edge intelligence. Edge intelligence comprises state-of-the-art computing through machine learning that can save time and money by reducing data floods and communication delays by providing closer data sources using data preprocessing and decision-making capabilities.

If we can analyze all the data collected on the edge and all the actions performed to give citizens a new and improved quality of life, the value of edge computing used in the smart city will be enormous.

G. MEC

MEC is a technology which can create new local services by relieving the communication network congestion by providing various types of services and caching contents from the position close to the user, which can be achieved by applying the distributed cloud computing technology to the radio base stations. This technology offers a cloud computing capability as well as some type of an IT service environment to the application developers or contents providers at the mobile network edge while providing ultra-low latency, high-capacity bandwidth, and real-time network information access capability to the application programs, in addition to situation awareness, rapid inter-RAT handoff, and low power consumption capabilities. The network providers leave their RAN edges open to the qualified external institutions as a third party to provide this technology to introduce some innovative applications or services to mobile subscribers, firms, and industries rapidly. The introduction of MEC technology along with its industry standard and platform will create a new profit model for the firms while becoming a differentiator between them.

MEC can create some innovative services/applications and provide a better QoE for the users. The possible service scenarios with the MEC technology are as follows [55]–[58]:

- Intelligent Video Acceleration (Fig. 10): This is a technology that accelerates mobile contents delivery by guiding the video source according to the real-time network capacity. The MEC applications linked to the radio analytics provide information about the processible volume by the wireless downlink interface to the remote video server. This information is used to make an adjustment according to the estimated capacity of the wireless downlink. The application of MEC improves user QoE through the provision of decreased 'time-to-start' effect for the contents and/or the video-stalling effect, as well as maximization of wireless network resources.

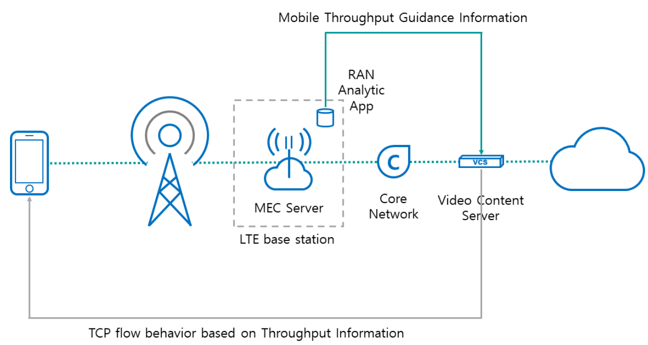

FIGURE 10. Overview of intelligent video acceleration in MEC.

- Video Stream Analysis (Fig. 11): This is a technology that analyzes video streams and uses its results for the monitoring system. The collected information is sent to a remote cloud-based monitoring system and such technology is being used for security purpose at various facilities. A large-volume transmission on the network can be prevented by delivering only small-size information to the MEC server after processing it locally.

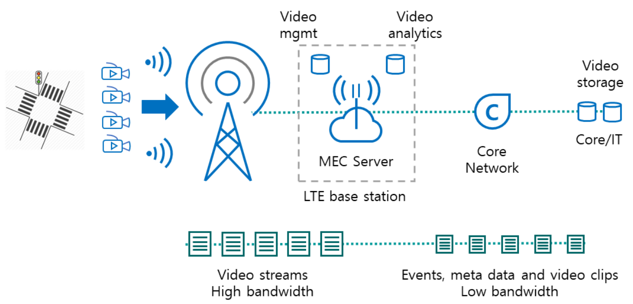

FIGURE 11. Overview of video stream analysis in MEC.

- Augmented Reality (Fig. 12): As part of the virtual reality technology, this is a computer graphics technology which synthesizes (compose) a virtual object or information in a real environment to make it appear as if it is actually existing in that environment. One of the typical methods being used is to let a person face the object he/she is interested through the mobile terminal and while doing so, some additional information will be added to the view. Since a rapid information update is required to show the augmented image on a real-time basis, a MEC server which can perform low-latency processing for a large amount of data is used commonly.

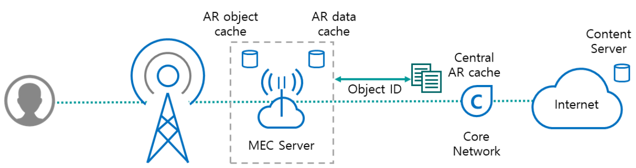

FIGURE 12. Overview of augmented reality in MEC.

5G is a next-generation mobile communication system which satisfies various technological requirements in AI, big data, IoT, and cloud systems. As shown in Fig. 13, MEC is the core technology for 5G and plays a vital role in reallocating the network functions or applications. Especially, as an important structural technology that achieves low latency, it plays a complementary role with the NFV technology which is the core technology for the network softwarization. And yet, MEC requires much more research for improvement and needs to surmount various technological challenges in order to be completely adapted to the 5G environment.

IV. CURRENT RESEARCH STUDIES FOR EDGE COMPUTING

The current research trends are compared and analyzed in Table 4.

V. DESIGNING AI TO THE EDGE: ENGINEERING APPROACH

Following the rapid development of AI technology playing a central role in almost every advanced system, the approaches for analyzing various types of data generated on the 'edge' to derive a meaningful insight are being developed in the field of edge computing. During the course of transformation from the era of cloud computing to the era of edge computing, intelligence has migrated to the edge systems such as mobile devices, smart speakers, sensors, wearable sensors, sensor networks, etc. enabling a much speedier data processing tasks including inference, pattern matching, and others. Ultimately, smart IoT devices or sensors instantly process data and provide the most optimal information promptly as the need for transferring data to the central server has been reduced due to the fact the hardware in the edge systems has become powerful enough to support AI-based algorithms, especially in terms of power and function. At the same time, system sensors have become smaller in size but their memory or processing capacity is continuously increasing.

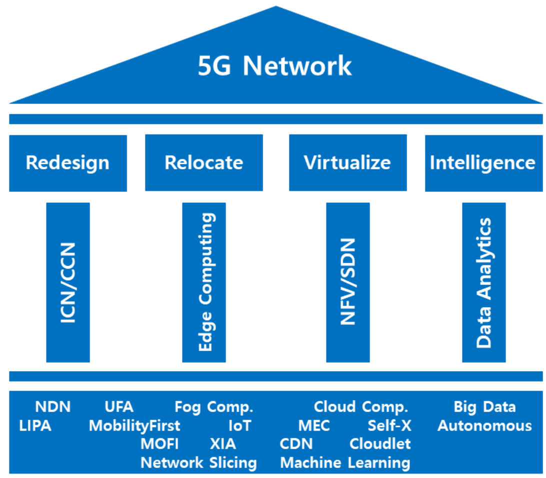

FIGURE 13. 5G network technology.

Meanwhile, since the IoT-based smart homes, factories, and cities are now holding some of the central position in the technology strategy in the era of the 4th industrial revolution, analysis of data streaming through AI technology-based edge computing has become possible largely increasing visibility and cognitive ability. Also, the edge systems (e.g., IoT devices, etc.) in edge computing are gradually becoming smarter with the assistance of AI so that they are able to read a pattern or detect abnormality in the data generated a smart sensor or device automatically through machine learning. The information could be about temperature, pressure, or vibration. Such an intelligent performance allows a timely provision of new service or product, increased operating efficiency, and proper risk management.

The main reason for AI and edge computing should be combined is that cloud computing is being centralized too much following the continuous increase in its computing scale so that increase in the number of small-size edge systems along with automated service could solve some of the problems inherent in it. Many more players who can deal with various types of demands associated with data, infrastructure, or application should appear in the future to understand the data generating from a variety of context.

TABLE 4. Feature and proposed techniques of related studies.

Authors	Title of Paper	Summary	Evaluation	Core technology	Published year	Domain
H. Wu K. Wolter Q. Wang	Tradeoff between Performance Improvement and Energy Saving in Mobile Cloud Offloading Systems [41]	To extend the battery life of a mobile device and shorten the execution time, authors analyze the offloading method based on a tradeoff analysis to explore the compromise point for the purpose.	A new model was proposed and validated through the formula which had considered the performance improvement and energy saving aspects of offloading.	Mobile Cloud offloading	2013	Cloud offloading
L. Tang S. He	Multi-User Computation Offloading in Mobile Edge Computing: A Behavioral Perspective [59]	Authors model the computation offloading decision making problem among multiple mobile device users.	Authors illustrate the impact of subjective parameters on the users' offloading decisions, for the expected utility theory (EUT) and prospect theory (PT)	Distributed computation offloading	2018	Cloud offloading
S. Yi Z. Hao Q. Zhang Q. Zhang W. shi Q. Li	LAVEA: Latency-Aware Video Analytics on Edge Computing Platform [60]	Authors propose LAVEA that offloads the computation between a customer and an edge node, and collaborates with nearby edge nodes to provide low latency video analytics.	A comparative verification has been performed between our proposed 'client-edge configuration-centric LAVEA' and local configuration and client-cloud configuration-centric LAVEAs by using different testbeds and pixel datasets.	ALPR, Edge computing, VCA	2017	VCA
C. Long Y. Cao T. Jiang Q. Zhang	Edge Computing Framework for Cooperative Video Processing in Multimedia IoT Systems [61]	Authors investigate video processing phenomenon of multimedia IoT system in three scenarios (smart city, satellite networks, Internet-of-Vehicle) and propose a method to implement collaborative video processing in edge computing platform.	The performance of the proposed edge computing framework was verified through the simulated studies involving lead time and processing time of the edge node.	Offloading, Edge computing, VCA	2018	VCA
J. Feng Z. Liu C. Wu Y. Ji	AVE: Autonomous Vehicular Edge Computing Framework with ACO-Based Scheduling [62]	Authors propose a framework called autonomous vehicular edge (AVE) to improve the computational capabilities of vehicles, and introduce a workflow supporting autonomous driving in an edge environment.	An evaluation was made with traffic simulator for the performance of the proposed scheme. The results showed that the proposed scheme was superior to the existing schemes in terms of total utility and job delay.	Edge computing, ACO algorithm	2017	autonomous vehicle

TABLE 4. ***(Continued.)*** **Feature and proposed techniques of related studies.**

N. Hassan S. Gillani E. Ahmed I. Yaqoob M. Imran	The role of edge computing in internet of things [63]	Authors investigate latest advances in edge computing from the IoT perspective. IoT-based edge computing is classified according to wireless network, service level objectives, data type, etc.	Ideal roles and research challenges were presented after categorizing the major roles of edge computing in an IoT environment.	MEC, C-RAN, Edge computing	2018	IoT
C. Chang S. Srirama R. Buyya	Internet of Things (IoT) and New Computing Paradigms [64]	Authors provide an overview in fog and edge computing (FEC). They suggest Storage, Compute, Acceleration, Networking and Control (SCANC) mechanisms and XaaS as technologies for implementing FEC's business models in smart services.	A method of solving the cloud-related issues in a cloud-centric IoT environment with fog and edge computing was proposed along with some of the new challenges.	CIoT, MEC, FEC, XaaS	2018	IoT
M. Sapienza E. Guardo M. Cavallo G. LaTorre G. Leombruno O. Tomarchio	Solving Critical Events through Mobile Edge Computing an Approach for Smart Cities [65]	Authors propose an architectural model based on the MEC concept to respond rapidly time- and mission-critical situations.	To detect abnormal and critical events such as terrorists, threats, natural and man-made disasters in smart city, authors present a scenario and evaluate the usefulness of the proposed architecture.	MEC, MCC, Fog computing, RAN	2016	Smart city
M. Santos de Brito S. Hoque R. Steinke A. Willner T. Magedanz	Application of the Fog computing paradigm to Smart Factories and cyber-physical systems [66]	Authors propose a Fog Node architecture for Fog computing paradigms in smart factories, as well a discussion on the orchestration system and programmable characteristics of the Fog Node. In addition, they introduce OpenFog Reference Architecture (OpenFog RA).	To demonstrate and evaluate the proposed concept, a simulation model was designed and two approaches were used to show how the fog computing paradigm impacts the production at the shop floor.	CPS, AAS, Fog computing, M2M	2017	Smart factory Cyber physical system
J. Cao L. Xu R. Abdallah W. Shi	EdgeOS: A Home Operating System for Internet of Everything [67]	Authors introduce EdgeOS for home operating system in Internet of Everything environment. Moreover, they provide how EdgeOS should be designed, how the data should be handled, how the security and privacy should be examined.	To help designing and developing smart home systems using Edge OS, authors discuss open non-functional issues such as open testbed, user experience, system cost and delay.	EdgeOS, Internet of Everything, Cloud computing	2017	Smart home

The merits which can be obtained by grafting AI into edge computing are as follows:

- The edge-based AI responds quite prompt. When compared to a cloud computing-based model, it performs almost on a real-time basis leading to higher customer confidence or friendliness.
- Improved security can be provided by minimizing the risk by reducing the number of data transferring/receiving processes or steps.

The design of an AI-based edge computing environment can be described as below.

As presented in Fig. 14, basically, in a cloud used in cloud computing, many functions such as broad network access, on-demand self-service, availability, certifiability, and resource pooling are involved to provide a single service. However, it is possible to extend the design of data analytics part within the cloud to let it play the role of a core platform which is able to perform a smooth AI-based analysis in an edge system. It is also possible to deploy a machine-learning model in an edge system which will then carry out inference locally using the pre-learned learning process and exchange information among the various types of sensors while reducing response delays and increasing flexibility to provide a better service. Thus, a foothold for AI technology to gradually pervade into our daily lives through edge computing by reducing the learning time or providing a quick response has been prepared.

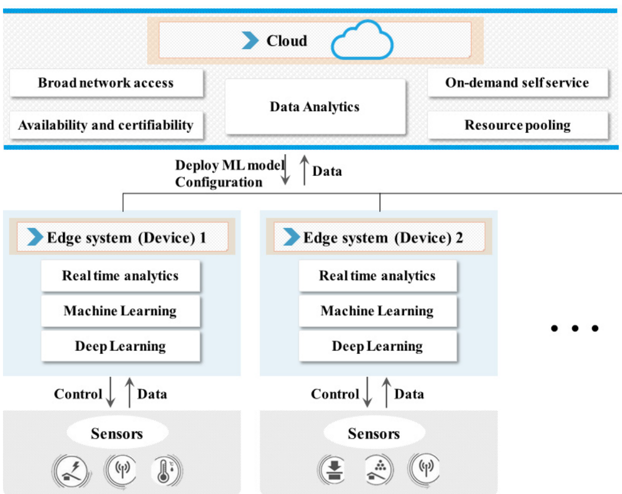

FIGURE 14. A design of edge computing environment supported by AI.

VI. CHALLENGES

A. MEC

Nowadays, as the mobile communication industry establishes 5G communication network, the services like autonomous driving which requires a low-latency and a stable connection are increasing gradually. 5G offers a network which satisfies the technological characteristics such as low latency, high bandwidth, mass connectivity and has a service-oriented architecture that can provide such network characteristics individually or on a per-service basis. As a service-oriented model, edge network requires multi-accessing resources and a service API and currently, MEC is being regarded as the technology that can satisfy such a network system, along with SDN and NFV technologies which are the key components in a 5G network.

The standardization of MEC is being carried out by ETSI (European Telecommunications Standards Institute) and many attempts are being made in various industries to apply it to an LTE or 5G network. Even though a variety of MEC-applied applications, services, or platforms have been developed or launched by the companies in a variety of areas, any tangible results are yet to be seen in the market. MEC still has a number of issues such as how to distribute contents or applications between the edge and central data center, or how to achieve an ultra-low latency in a 5G network.

MEC is a technology that distributes the load and traffic of the enormous amount of data flowing into the central server due to the technological advancement in AI, IoT, or big data technologies, to which edge computing technology has been applied to allow the applications requiring a low-latency to access a network in real time. In a RAN playing the role of an edge, the task that can be processed by MEC itself is handled by itself and if not, it will be assisted by peripheral MEC server or the data center of a cloud (Fig. 15). The workload and traffic can be distributed in this way so that low-latency services such as autonomous driving or traffic control which are sensitive to the latency can be implemented.

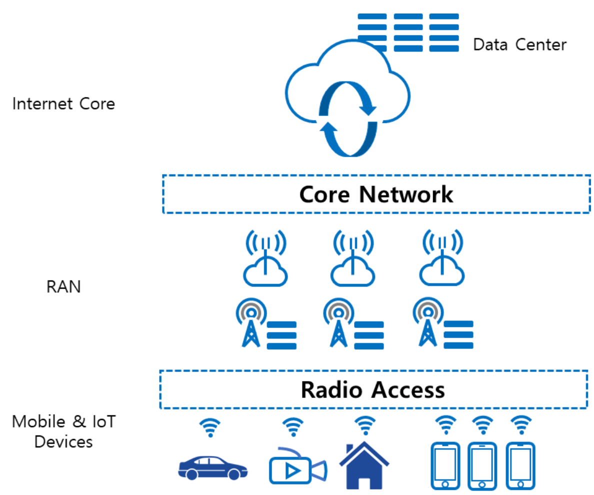

FIGURE 15. MEC system.

Such a technology can be utilized for the computation offloading through MEC's computer resources or the computation migration using an MDP. Basically, computation offloading is often performed to enhance computing power by utilizing MEC's computer resources and there are two ways to achieve it: CloneCloud [42] processes the codes by separating them as the code that should be handled by the terminal or the MEC by the thread unit, whereas MAUI [68], [69] separates applications by the method unit and proceed with the code that should be handled by MEC only. Meanwhile, computation migration is used when a service has been performed by the terminal but has to be moved to a nearby MEC server after moving to another area. The migration is carried out in a way that minimizes the cost after measuring the distance between two MEC servers by using an MDP.

As such, MEC reduces latency by allocating the computing load from the existing cloud computing to edge computing and create a new service by reducing the congestions in the central network. However, MEC is still in its early stages and require more research to solve the problems involved with resource management, partitioning of online jobs, or optimization:

First, the offloading process should be carried out through several channel blocks. This is the task that requires has to be researched even for a single-user MEC system having a deterministic job arrival [70]–[75].

Second, it is necessary to prevent the occurrences of inefficient and unfeasible offloading and degradation of computing performance by establishing an online job partitioning strategy. Currently, the job partitioning algorithms have been developed for the applications having a serial job call graph only and the solutions for general job models are yet to be utilized [76].

Third, it is necessary to co-manage resources to provide several mobile devices simultaneously through collaboration between multiple MEC servers to achieve full-scale optimization. However, as the size of a network increases, a series of problems in optimizing resource management could occur in relation to allocating wireless/computing resources or many offloading decisions [77]. A full-scale distributed and low-complex optimization algorithm should be developed to manage resources effectively to solve such problems [33].

There are other necessary improvements to be made as well for the problems involving layer scheduling for computing resources, or service instance handover. Although it may not be easy to standardize the 5G networking technology which satisfies all these requirements, the problems have to be solved as early as possible for the diversity of service migrations in mobile networks.

B. INTELLIGENT IOT

Since the introduction of the IoT technology and its smooth pervasion into people's daily lives, various types of technological demands are rising so that major global IT companies are developing their own IoT-driven systems to strengthen their business competitiveness. There various products available on the market to keep up with the demands and at the same time, the research on the method of processing huge amount of data collected through various types of IoT equipment is performed in priority, regarding Edge computing as one of the effective technologies for that purpose.

As the IoT becomes increasingly practical, the amount of data to be processed in real time is increasing. As it is difficult to process a large amount of data at the central data center in the cloud, it is expected that edge computing technology will be used to process, analyze, and utilize the data collected from the terminal. Edge acts as a gateway to optimize the industrial data and can immediately deal with task-related matters. Edge and cloud computing are attracting attention in all business areas because they can respond more efficiently, quickly, and safely in order to analyze and apply a large amount of data.

One such typical example is traffic signal facilities such as traffic lights and streetlights, which use the edge-to-cloud platform to process information from edge-computing-based devices and transmit data to the cloud. Traffic-related applications collect data from their respective locations, respond to changes, and perform meaningful analyses.

In the IoT field, edge computing is a new technology, and there are many challenges yet to be tackled: heterogeneity, standard protocols and interfaces, availability, security, and privacy [63].

Firstly, it is necessary to solve the problem of the heterogeneity of computing and communication technologies. Communication technology can be heterogeneous in relation to the data rate, transmission range, and bandwidth; thus, it is necessary to develop a software solution that is portable across different environments. However, as this software solution may vary depending on the hardware, it is necessary to develop a programming model that can execute workloads simultaneously at multiple hardware levels [63].

Secondly, communication protocols and interfaces must be standardized. Standard protocols and interfaces should be developed because the interface for communicating with different devices in a heterogeneous environment and the communication protocol required differ [63], [78].

Thirdly, the availability of resources and services for IoT devices must be guaranteed. It is difficult to guarantee the availability of resources and services for the increasing number of IoT devices. So, we have to maximize the average time between errors, and optimize the availability by minimizing the failure probability and average recovery time [79].

Fourthly, robust security and privacy are required. In the IoT field, the network is not protected by the very dynamic environment at the edge of a network. Data obtained from various devices contains sensitive personal information [79], [80], and thus, it is necessary to prevent intruders from using data for illegal purposes. Therefore, through the use of a hierarchical fog computing framework, the personal information of user content is protected, and the data traffic on the core network is reduced.

C. AUTONOMOUS VEHICLES WITH AI

One of the keywords in the 4th Industrial Revolution is 'convergence' between two areas and the most typical example can be found in autonomous driving technology where auto industry meets with AI. The autonomous driving applied with a deep learning technology allows the vehicle to drive safely based on a more accurate assessment of the current situation without any intervention from the driver most of the time. It is expected that the technologies similar to this one will bring about innovations in a wide range of infortainment services including 5G network-based HD map, video phone, media streaming, etc., generating some additional value-added services like the in-vehicle electronic payment systems. The business models being pursued in some of the IT-advanced countries include robot taxi, car sharing, and unmanned delivery systems.

It is quite clear that these business models will cause high-level latency and traffic overload due to the exponential increase in data exchange volume. Edge computing is a model that can analyze and process big data. For example, as a latency-sensitive technology, autonomous driving requires a rapid computing capability to accurately control the vehicle and edge computing is the right one for this purpose.

As shown in Fig. 16, each sensor installed in an autonomous vehicle collects data by checking the surrounding terrain and road conditions/traffics and this data is used to rapidly deal with the various types of incidents that can occur while driving. The tasks which can be processed in an intermediate stage are handled in real time by the edge computing system without having to reach the cloud server.

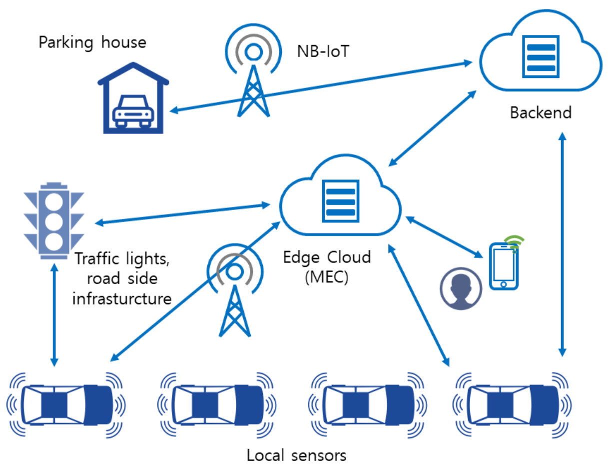

FIGURE 16. Architecture of autonomous vehicles with edge computing.

The current autonomous vehicles are not completely autonomous and stay at SAE (Society of Automotive Engineers) Level-2, meaning partially automated vehicles [81]. The industries involved in designing this type of vehicles are trying to develop a perfect autonomous vehicle but there are many technological obstacles to be cleared.

The issues pertaining to the future research subjects associated with edge computing are as follows:

First, an edge computing system that can rapidly process a huge volume of the vehicle's internal sensor data and driving data is required. The most important sensor currently used for autonomous vehicles is a LIDAR sensor (image sensor) which generates the data size from hundreds of megabytes to gigabytes per second and them in a cloud form. If the data is to be collected in raw form, it could amount up to terabytes per hour, not to mention the hundreds of gigabytes of data that can be generated from the image cameras installed in the vehicle. Since more sensors are expected to be used to accurately assess the surrounding conditions, the data volume will expand much more so that the researches on the technologies that can enhance computing power should be consistently carried out along with the data compression/conversion techniques.

Second, the big data collected continuously from the traffic infrastructure needs to be processed in addition to the sensor data generated by the autonomous vehicles or the traffic infrastructure, all of which should be computed by the edge computing device to return responses immediately. It is also essential that the issues involved in the architecture of network infrastructure and intelligent network operating systems have to be dealt with to guarantee uninterrupted data stream for collection.

In order for autonomous vehicles to establish themselves as a safe traffic infrastructure in today's society, the technologies involved in designing/manufacturing have to be developed along with the communication technology that can increase their safety. The importance of edge computing is emphasized in this aspect and in the near future, this technology will evolve into a model which supplements the cloud service by assuming the role of a central service on edge device as well as on the central server. It is expected that some special AI chips will be mounted on a variety of edge devices within five years to establish a perfect traffic infrastructure for safe autonomous driving.

VII. CONCLUSION

Edge computing is a new technology that has recently attracted attention, and there is much room for improvement in order to obtain efficient distributed computing. By processing data from edges, reducing the response times, and improving the reliability, a greater number of services will move from the cloud to the edge. The three problems of the cloud are service delay, increased processing load, and information infringement; edge computing technology can be used to solve these problems.

Edge computing is a good solution to perform stable operations in real time with specific functions in a field, and it is distinct from the cloud. It is expected to become an important success factor in the market if we use various technologies such as AI and Big data as well as IoT in a balanced manner to meet our objectives.

Recently, one of the keyword for edge computing is empowered edge. Gartner, an information technology consulting firm, stated that empowered edge is one of the top 10 strategic technology trends in 2019. Empowered edge will soon evolve into a model that complements cloud services managed as central services in the central servers as well as distributed on-premises and edge devices themselves. Over the next 5 years, special AI chips can be trained with huge volumes of data, such as images, text, and language will be installed in various edge devices, and it will be difficult to manage the long-life cycle of assets such as heterogeneity and industrial systems in these embedded IoT fields. When these problems are solved, the expanded edge computing environment will be used to build central services and strong communications with 5G.

Moreover, the hybrid computing method that combines cloud and edge computing is also becoming widely used. This does not mean that edge computing is replacing the cloud. The cloud can provide the same performance as that provided by edge computing with its advanced (AI) processor. Thus, we will use the cloud computing for high-level operations and edge computing for low-level operations. The challenge in this is to complement and effectively leverage the cloud and edge computing with the advanced (AI) processor.

The authors hope that the significance and potential of edge computing will be understood better by both academia and industry through this research work and more works or projects will be performed in the future.

ACKNOWLEDGMENT

This paper is the result of invaluable contributions from many individuals. The authors would especially like to thank Hyunsoo Jang and Jeongwon Kim for their work on the initial design of the research.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, Apr. 2010.

[2] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud computing: Distributed Internet computing for IT and scientific research," *IEEE Internet Comput.*, vol. 13, no. 5, pp. 10–13, Sep./Oct. 2009.

[3] D. C. Marinescu, *Cloud Computing: Theory and Practice*. 2nd ed. Cambridge, MA, USA: Morgan Kaufmann, 2017.

[4] Y. Jadeja and K. Modi, "Cloud computing—Concepts, architecture and challenges," in *Proc. ICCEET*, Kumaracoil, India, Mar. 2012, pp. 877–880.

[5] L. M. Kaufman, "Data security in the world of cloud computing," *IEEE Security Privacy*, vol. 7, no. 4, pp. 61–64, Jul./Aug. 2009.

[6] J.-H. Huh, "Server operation and virtualization to save energy and cost in future sustainable computing," *Sustainability*, vol. 10, no. 6, p. 1919, 2018.

[7] C. Mukundha and K. Vidyamadhuri, "Cloud computing models: A survey," *Adv. Comput. Sci. Technol.*, vol. 10, no. 5, pp. 747–761, 2017.

[8] S. Ghemawat, H. Gobioff, and S. Leung, "The Google file system," *ACM SIGOPS Operating Syst. Rev.*, vol. 37, no. 5, pp. 29–43, 2003.

[9] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, Jan. 2008.

[10] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. MSST*, Incline Village, NV, USA, 2010, pp. 1–10.

[11] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *Proc. HotCloud*, Boston, MA, USA, 2010, pp. 1–7.

[12] K. Ashton, "That 'Internet of Things' thing," *RFID J. Adv. Internet Things*, vol. 22, no. 7, pp. 97–114, Jun. 2009.

[13] NCTA. (Aug. 2015). *Behind The Numbers: Growth in the Internet of Things*. [Online]. Available: https://www.ncta.com/whats-new/behind-the-numbers-growth-in-the-internet-of-things-2

[14] D. Evans. (Apr. 2011). *The Internet of Things: How the Next Evolution of the Internet is Changing Everything*. [Online]. Available: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

[15] V. Gezer, J. Um, and M. Ruskowski, "An extensible edge computing architecture: Definition, requirements and enablers," in *Proc. UBICOMM*, Barcelona, Spain, 2017, pp. 148–152.

[16] G. Volkan, U. Jumyung, and R. Martin, "An introduction to edge computing and a real-time capable server architecture," *Int. J. Adv. Intell. Syst.*, vol. 11, nos. 1–2, pp. 105–114, 2018.

[17] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelffle. (Mar. 2010). *Vision and Challenges for Realising the Internet of Things*. [Online]. Available: http://www.internet-of-things-research.eu/pdf/IoT_Clusterbook_March_2010.pdf

[18] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generat. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[19] CISCO. (Nov. 2018). *Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf

[20] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: Research problems in data center networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, pp. 68–73, Jan. 2009.

[21] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. MCC*, Helsinki, Finland, 2012, pp. 13–16.

[22] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog computing: Platform and applications," in *Proc. HotWeb*, Washington, DC, USA, 2015, pp. 73–78.

[23] S. Yi, Z. Qin, and Q. Li, "Security and privacy issues of fog computing: A survey," in *Proc. WASA*, LNCS, vol. 9204, 2015, pp. 685–695.

[24] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Comput.*, vol. 8, no. 4, pp. 14–23, Oct. 2009.

[25] U. Shaukat, E. Ahmed, Z. Anwar, and F. Xia, "Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges," *J. Netw. Comput. Appl.*, vol. 62, pp. 18–40, Feb. 2016.

[26] M. Satyanarayanan, "The role of cloudlets in hostile environments," in *Proc. MCS*, New York, NY, USA, 2013, pp. 1–2.

[27] E. Ahmed and M. H. Rehmani, "Mobile edge computing: Opportunities, solutions, and challenges," *Future Generat. Comput. Syst.*, vol. 70, pp. 59–63, May 2017.

[28] L. Gupta, R. Jain, and H. A. Chan, "Mobile edge computing—An important ingredient of 5G networks," *IEEE Softw. Defined Netw. Newslett.*, Mar. 2016. [Online]. Available: https://sdn.ieee.org/newsletter/march-2016/mobile-edge-computing-an-important-ingredient-of-5g-networks

[29] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[30] 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, Oct. 2016.

[31] G. Premsankar, M. Di Francesco, and T. Taleb, "Edge computing for the Internet of Things: A case study," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1275–1284, Apr. 2018.

[32] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, 3rd Quart., 2017.

[33] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[34] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.

[35] B. Cohen, G. Kathirvel, G. Katsaros, A. Lebre, A. Mitry, C. Price, P.-A. Raymond, A. Reznik, P. Vaananen, I. Vansca, and I. Wells. *Cloud Edge Computing: Beyond the Data Center*. Accessed: Oct. 23, 2019. [Online]. Available: https://www.openstack.org/edge-computing/cloud-edge-computing-beyond-the-data-center?lang=en_US

[36] B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, and D. S. Nikolopoilos, "Challenges and opportunities in edge computing," in *Proc. SmartCloud*, 2016, pp. 20–26.

[37] P. G. Lopez, A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber, and E. Riviere, "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 37–42, 2015.

[38] M. Satyanarayanan, P. Simoens, Y. Xiao, P. Pillai, Z. Chen, K. Ha, W. Hu, and B. Amos, "Edge analytics in the Internet of Things," *IEEE Pervasive Comput.*, vol. 14, no. 2, pp. 24–31, Feb. 2015.

[39] B. Iyer. (Jun. 2016). *VMware Introduces Liota—The IoT Developer's Dream*. [Online]. Available: https://www.vmware.com/radius/vmware-introduces-liota-iot-developers-dream/

[40] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Commun. Mobile Comput.*, vol. 13, no. 18, pp. 1587–1611, Dec. 2013.
[41] H. Wu, Q. Wang, and K. Wolter, "Tradeoff between performance improvement and energy saving in mobile cloud offloading systems," in *Proc. IEEE ICC*, Budapest, Hungary, 2013, pp. 728–732.
[42] B.-G. Chun, S. Ihm, P. Maniatis, M. Naik, and A. Patti, "CloneCloud: Elastic execution between mobile device and cloud," in *Proc. EuroSys*, Salzburg, Austria, 2011, pp. 301–314.
[43] A. Rudenko, P. Reiher, G. J. Popek, and G. H. Kuenning, "Saving portable computer battery power through remote process execution," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 2, no. 1, pp. 19–26, 1998.
[44] G. C. Hunt and M. L. Scott, "The coign automatic distributed partitioning system," in *Proc. OSDI*, vol. 99, New Orleans, LA, USA, 1999, pp. 187–200.
[45] K. Kumar and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *Computer*, vol. 43, no. 4, pp. 51–56, Apr. 2010.
[46] 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," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 945–953.
[47] T. A.-Q. Tawiah, "Video content analysis for automated detection and tracking of humans in CCTV surveillance applications," Ph.D. dissertation, Dept. School Eng. Des., Brunel Univ., London, U.K., Aug. 2010
[48] *An Introduction to Video Content Analysis—Industry Guide*, BSIA, Worcester, U.K., Aug. 2016.
[49] E. Alcala, L. Sellart, V. Puig, J. Quevedo, J. Saludes, D. Vázquez, and A. Lopez, "Comparison of two non-linear model-based control strategies for autonomous vehicles," in *Proc. MED*, Athens, Greece, 2016, pp. 846–851.
[50] L. Eliot. (Jan. 2018). *Edge Computing for AI Self-Driving Cars*. [Online]. Available: https://www.aitrends.com/ai-insider/edge-computing-ai-self-driving-cars/
[51] M. Finnegan. (Mar. 2013). *Boeing 787s to Create Half a Terabyte of Data Per Flight, Says Virgin Atlantic*. [Online]. Available: https://www.computerworlduk.com/data/boeing-787s-create-half-terabyte-of-data-per-flight-says-virgin-atlantic-3433595/
[52] M. Rijmenam. (Jul. 2013). *Self-Driving Cars Will Create 2 Petabytes of Data, What are the Big Data Opportunities for the Car Industry?* [Online]. Available: https://datafloq.com/read/self-driving-carscreate-2-petabytes-data-annually/172
[53] D. Watkins. (Nov. 2018). *Global Smart Speaker Vendor & OS Shipment and Installed Base Market Share by Region: Q3 2018*. [Online]. Available: https://www.strategyanalytics.com/access-services/devices/connected-home/smart-speakers-and-screens/market-data/report-detail/global-smart-speaker-vendor-os-shipment-and-installed-base-market-share-by-region-q3-2018
[54] C. Perera, Y. Qin, J. C. Estrella, S. Reiff-marganiec, and A. V. Vasilakos, "Fog computing for sustainable smart cities: A survey," *ACM Comput. Surv.*, vol. 50, p. 32, Mar. 2017.
[55] N. Antonopoulos and L. Gillam, *Cloud Computing: Principles, Systems and Applications*, 2nd ed. Cham, Switzerland: Springer, 2017, pp. 3–26.
[56] S.-R. Yang, Y.-J. Tseng, C.-C. Huang, and W.-C. Lin, "Multi-access edge computing enhanced video streaming: Proof-of-concept implementation and prediction/QoE models," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1888–1902, Feb. 2019.
[57] Y.-H. Hung, C.-Y. Wang, and R.-H. Hwang, "Optimizing social welfare of live video streaming services in mobile edge computing," *IEEE Trans. Mobile Comput.*, to be published.
[58] A. Zhu, S. Guo, B. Liu, M. Ma, H. Feng, and X. Su, "Adaptive multi-service heterogeneous network selection scheme in mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6862–6875, Aug. 2019.
[59] L. Tang and S. He, "Multi-user computation offloading in mobile edge computing: A behavioral perspective," *IEEE Netw.*, vol. 32, no. 1, pp. 48–53, Jan./Feb. 2018.
[60] S. Yi, Z. Hao, Q. Zhang, Q. Zhang, W. Shi, and Q. Li, "LAVEA: Latency-aware video analytics on edge computing platform," in *Proc. ICDCS*, Atlanta, GA, USA, 2017, pp. 2573–2574.
[61] C. Long, Y. Cao, T. Jiang, and Q. Zhang, "Edge computing framework for cooperative video processing in multimedia IoT systems," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1126–1139, May 2018.
[62] J. Feng, Z. Liu, C. Wu, and Y. Ji, "AVE: Autonomous vehicular edge computing framework with ACO-based scheduling," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10660–10675, Dec. 2017.
[63] N. Hassan, S. Gillani, E. Ahmed, I. Ibrar, and M. Imran, "The role of edge computing in Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 110–115, Nov. 2018.
[64] C. Chang, S. N. Srirama, and R. Buyya, "Internet of Things (IoT) and new computing paradigms," in *Fog and Edge Computing: Principles and Paradigms*. Hoboken, NJ, USA: Wiley, 2018, pp. 1–23.
[65] M. Sapienza, E. Guardo, M. Cavallo, G. La Torre, G. Leombruno, and O. Tomarchio, "Solving critical events through mobile edge computing: An approach for smart cities," in *Proc. SMARTCOMP*, St. Louis, MO, USA, 2016, pp. 1–5.
[66] M. S. de Brito, S. Hoque, R. Steinke, A. Willner, and T. Magedanz, "Application of the fog computing paradigm to smart factories and cyber-physical systems," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 4, May 2017, Art. no. e3184.
[67] J. Cao, L. Xu, R. Abdallah, and W. Shi, "EdgeOS_H: A home operating system for Internet of everything," in *Proc. ICDCS*, Atlanta, GA, USA, 2017, pp. 1756–1764.
[68] E. Cuervo, A. Balasubramanian, D.-K. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "MAUI: Making smartphones last longer with code offload," in *Proc. MobiSys*, San Francisco, CA, USA, 2010, pp. 49–62.
[69] K. Peng, V. C. M. Leung, X. Xu, L. Zheng, J. Wang, and Q. Huang, "A survey on mobile edge computing: Focusing on service adoption and provision," *Wireless Commun. Mobile Comput.*, vol. 2018, Oct. 2018, Art. no. 8267838.
[70] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013.
[71] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.
[72] X. Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Trans. Veh. Techn.*, vol. 66, no. 4, pp. 3435–3447, Apr. 2017.
[73] O. Muñoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Trans. Veh. Techn.*, vol. 64, no. 10, pp. 497–508, Oct. 2015.
[74] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.
[75] Y. Yu, J. Zhang, and K. B. Letaief, "Joint subcarrier and CPU time allocation for mobile edge computing," in *Proc. GLOBECOM*, Washington, DC, USA, 2016, pp. 1–6.
[76] W. Zhang, Y. Wen, and D. O. Wu, "Collaborative task execution in mobile cloud computing under a stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 81–93, Jan. 2015.
[77] Z. Sheng, C. Mahapatra, V. C. M. Leung, M. Chen, and P. K. Sahu, "Energy efficient cooperative computing in mobile wireless sensor networks," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 114–126, Jan./Mar. 2018.
[78] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.
[79] F. A. Kraemer, A. E. Braten, N. Tamkittikhun, and D. Palma, "Fog computing in healthcare—A review and discussion," *IEEE Access*, vol. 5, pp. 9206–9222, 2017.
[80] X. Sun and N. Ansari, "EdgeIoT: Mobile edge computing for the Internet of Things," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 22–29, Dec. 2016.
[81] T. Pollard. (Jan. 2019). *What are Autonomous Car Levels? Levels 1 to 5 of Driverless Vehicle Tech Explained*. [Online]. Available: https://www.carmagazine.co.uk/car-news/tech/autonomous-car-levels-different-driverless-technology-levels-explained/

JUN-HO HUH was born in Kyoto, Japan. He received the B.S. degree in science from the Department of Major of Applied Marine Sciences (Currently Faculty of Marine Biomedical Sciences) and the B.S. degree in engineering (Double Major) from the Department of Major of Computer Engineering, Jeju National University, Ara, Jeju, South Korea, in August 2007, and also, completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act, in August 2007; the M.A. degree in education from the Department of Major of Computer Science Education, Graduate School of Education, Pukyoug National University, Daeyeon, Busan, South Korea, in August 2012, and also completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act, in August 2012; and the Ph.D. degree in engineering from the Department of Major of Computer Engineering, Graduate School, Pukyoug National University, in February 2016.

He was finished the Cooperative Marine Science and Engineering Program, Texas A&M University at Galveston, USA, in August 2006. He was a Research Professor with Dankook University, Jukjeon, Yongin, South Korea, from July 2016 to September 2016. He was an Assistant Professor with the Department of Software, Catholic University of Pusan, South Korea, from December 2016 to August 2019. Since September 2019, he has been an Assistant Professor (Tenure Track) with the Department of Data Informatics, (National) Korea Maritime and Ocean University, South Korea. His research interests include green IT, smart grid, edge computing, network security, the IoT, big data, and system architecture.

He received the Best Paper Award from Korea Multimedia Society eleven times (November 2014, May 2015, November 2015, May 2016, October 2016, May, 2017; three times, October 2017, October 2018), the Undergraduate Student Paper Bronze Medal (Corresponding Author) Awarded, Korea Information Processing Society, in April 2017, the Best Paper Award the 10th 2016 International Interdisciplinary Workshop Series from HSST, in August 2016, the Best Paper Award the 16th International Conference on Control, Automation and Systems October 2016, ICROS with IEEE Xplore, and the Springer Nature Journal Award, Human-centric Computing and Information Sciences most cited paper award 2019 (Research published in the journal, from 2016 to 2018; SCIE IF=3.212). In addition, he received the Best Paper Award (Two times) from the 15th International Conference on Multimedia Information Technology and Applications (MITA 2019), and the Best Poster Award (First Place) from the 7th International Conference on Big Data Applications and Services (BIGDAS 2019). He has been serving as chairs in some conferences and workshops; MUE 2017, BIC 2017, CUTE 2017, WITC 2018, MUE 2018, PDCAT 2018, BIC 2018, CUTE 2018, WITC 2019, MUE 2019, MIMI 2019, MIMA 2019, MITA 2019, BIC 2019, and CUTE/CSA 2019, He is Technical Committee at IFAC (International Federation of Automatic Control), CC 1 (Systems and Signals), TC 1.5. (Networked Systems). And, Technical Committee at IFAC, CC 3 (Computers, Cognition and Communication), TC 3.2 (Computational Intelligence in Control). Technical Committee at IFAC, CC 7 (Transportation and Vehicle Systems), TC 7.2. (Marine Systems). He is an Associate Editor with Human-centric Computing and Information Sciences (HCIS), Springer Berlin Heidelberg (SCIE/SCOPUS indexed). He is an Associate Editor with the *Journal of Information Processing Systems* (JIPS), Korea Information Processing Society (SCOPUS/ESCI indexed). He is a Managing Editor of the *Journal of Multimedia Information System*.

YEONG-SEOK SEO (M'19) received the B.S. degree in computer science from Soongsil University, Seoul, South Korea, in 2006, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008 and 2012, respectively.

From September 2012 to December 2013, he was a Postdoctoral Researcher with the KAIST Institute for Information and Electronics. From January 2014 to August 2016, he was a Senior Researcher with the Korea Testing Laboratory (KTL), Seoul. Since 2016, he has been an Assistant Professor with the Department of Computer Engineering, Yeungnam University, Gyeongsan, Gyeongbuk, South Korea. His research interests include software engineering, the Internet of Things, artificial intelligence, data mining, and big data analysis.

Prof. Seo is a member of board of directors of software engineering society in South Korea. He has served as the Proceedings Co-Chair for APSEC 2018, the Publicity chair for WITC 2019, and the Program Chair for HCIS 2019. He was a recipient of The 2nd JIPS Survey Paper Awards, in 2019. He is an Associate Editor of the *Journal of Information Processing Systems* (JIPS), Korea Information Processing Society (SCOPUS/ESCI indexed), and a Guest Editor of the *Journal of Systems and Software* (JSS), Elsevier (SCIE indexed). Also, he is involved in international standardization activities and is a member of the Korean National Body mirror committee to ISO on IT Service Management and IT Governance (ISO/IEC JTC1/SC40).