

Edge Computing Technology: Development and Measures

Hong Xuehai^{1,2}, Wang Yang²

1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

2. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Edge computing is an emerging technology that reduces transmission delays and bandwidth consumption by placing computing, storage, bandwidth, application, and other resources on the edge of a network. Moreover, edge computing allows application developers and content providers to provide perceptible services based on real-time network information. Mobile terminals, the Internet of Things, and other devices provide the necessary front-end support for computing-sensitive applications such as image recognition and online games to share the cloud workload with the processing capability of edge computing. This paper discusses the concept of edge computing, key problems that require solutions, main advances in edge computing, influences of edge computing development, and opportunities and measures of edge computing development.

Keywords: cloud computing; edge computing; fog computing; mobile edge computing; Internet of Things; front-end intelligence

1 Concept of edge computing

The basic, initially conceived functions of the cloud, network, and terminals were as an extremely powerful data center responsible for processing data, a network to facilitate data transmission, and nodes of the Internet of Things (IoT) responsible for collecting data and transmitting it to the cloud, respectively. In IoT cloud networks, the decisions are made in the cloud based on data analysis before returning the results to the terminals. In this model, the cloud handles intelligent computing, while the terminal nodes handle data collection and decision implementation.

However, initially conceived functions of the cloud, network, and terminals encounter many difficulties in actual implementation. The first is the data transmission overhead. IoT nodes typically transmit all unprocessed, raw data over wireless networks to the cloud, resulting in bandwidth demands that exceed the capabilities of the network infrastructure. The overhead also comes from the power consumption necessary for wireless transmis-

sion. If all unprocessed data are to be transmitted to the cloud, the wireless transmission module of the terminal node must support high-speed wireless transmission, leading to high power consumption by the wireless module. This is inconsistent with the initial idea about the low power consumption of IoT nodes. The second difficulty is the network delay. The implementation of tasks of a large number of nodes is extremely sensitive to delays. For example, in driverless car, augmented reality, or virtual reality applications, tasks often cannot be accepted if network transmission delays exceed tens of milliseconds. Occasionally, weak network signals will cause delays of several seconds or worse, disconnection. In response to these difficulties, edge computing was conceived.

Presently, there is no strict definition of edge computing. The physical boundary of edge computing currently believed is shown in Fig. 1.

Shi et al. [1] have defined edge computing as a new computing model in which calculations are performed at the edge of the network, for example in downstream cloud services and

Received date: April 18, 2018; **Revised date:** April 20, 2018

Corresponding author: Hong Xuehai, Chinese Academy of Sciences, Researcher. Major research fields include high performance computing, cloud computing, and research on information technology and information strategy, etc. E-mail: hxh@ict.ac.cn

Funding program: CAE Advisory Project "Development Strategy of 'Internet Plus' Action Plan" (2016-ZD-03); Project of National Natural Science Foundation of China (91646127); Special Project for Document Intelligence Capability Construction of Chinese Academy of Sciences: Academy and Subordinate Institute Coordination Project

Chinese version: Strategic Study of CAE 2018, 20 (2): 020–026

Cited item: Hong Xuehai et al. Edge Computing Technology: Development and Measures. *Strategic Study of CAE*, <https://doi.org/10.15302/J-SSCAE-2018.02.004>

upstream Internet of Everything services. According to the definition of China’s Edge Computing Consortium, edge computing refers to an open platform that converges the core capabilities of network, computing, storage, and applications on the edge of the network—close to the “thing” or data source—and provides edge intelligent services nearby to meet the industry’s key digital demands for agile connection, on-time operation, data optimization, application intelligence, and security and privacy protection. For edge computing, the core concept for upstream interconnection or downstream cloud service and converged service computing is that “computing should be closer to the data source as well as the user,” and the task of intelligent computing is moved to the physical end, not just the cloud’s end. Before the emergence of edge computing, concepts such as sea-cloud computing, fog computing [2], and the recent mobile edge computing (MEC), were explored, all of which attempt to overcome the heavy load, extensive delays, security risks, and other issues associated with cloud computing and cloud services. The goal is to get cloud computing closer to users and data sources with a new computing model of prepositive intelligent computing, complemented by cloud computing and cloud services. Therefore, edge computing is a new model following distributed, grid, and cloud computing. The authors believe edge computing to be a new computing model with the advantage of deeply collaborative resource-scheduling capabilities. The model focuses on cloud computing, takes use of modern communication networks as the approach, makes mass terminals as the perceived front-end, and implements the optimization of resource allocation to enable more intelligent computing, storage, transmission, applications,

and services. It is a new type of computing model that integrates the cloud, network, terminals, and intelligence.

2 Key issues to be solved during development of edge computing

In edge computing, the terminal node is no longer completely responsible for computing, and instead passes partially processed data to the cloud. In this way, issues arising from network delays and bandwidth limits can be avoided, as transmissions are much smaller. Furthermore, the specific computational complexity at the edge depends on the trade-off between the power consumption of computing and that of wireless transmission. A higher computational complexity at the terminal will generally result in greater power consumption for computing and less power consumption for wireless transmission, though performance may differ according to systems. Additionally, because the data at the terminal node are not all transmitted to the cloud through the network, data security can be guaranteed under certain data-sensitive application scenarios. The following major problems need to be solved for edge computing.

2.1 Architecture of edge computing

Fog, sea-cloud, ternary, and MEC can all essentially be classified under edge computing, which covers the IoT, cellular networks, vehicle networking, the Internet, and mobile Internet. It is the optimal solution for resources and facilities including computing, storage, network, software, and input/output. It

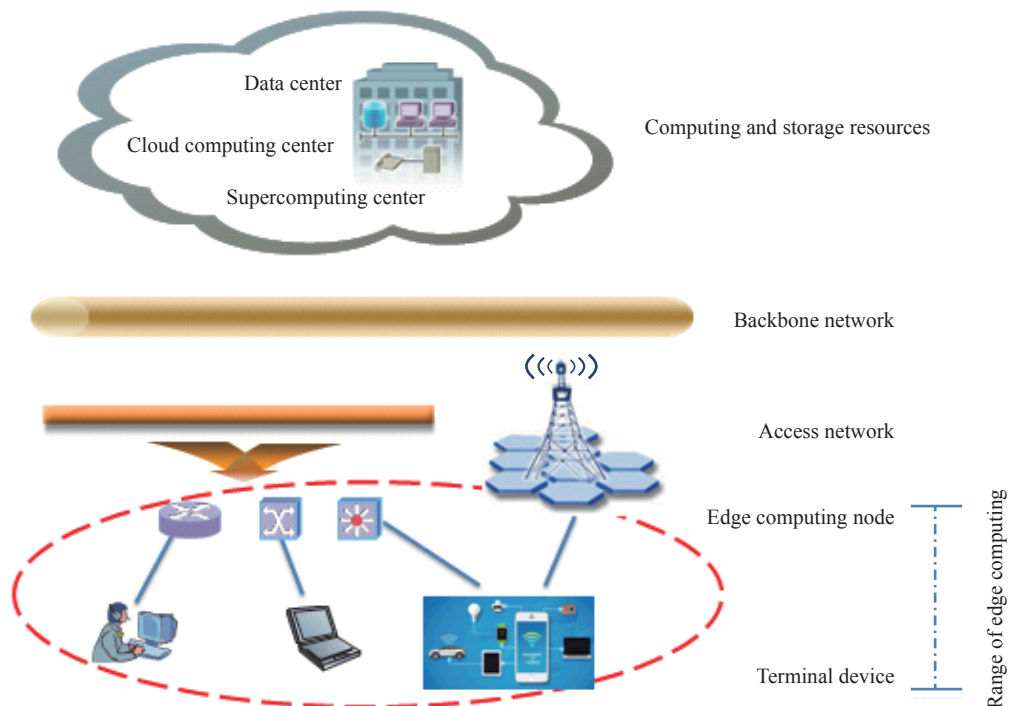


Fig. 1. Physical boundary of edge computing.

facilitates algorithm improvement, computing optimization, and data rearrangement. However, while edge computing directly addresses numerous application scenarios and modes at the edge of the network, not only various general architectures are required to face these application scenarios and modes, but also a general architecture such as specific architectures—including fog computing and MEC architectures—should be designed for different application scenarios and must deal with complex application scenarios and modes. Based on the above, the allocation of resources such as computing, storage, network, and software is planned to optimize application performance, power consumption, and security.

2.2 Processor power of edge computing devices

Different application scenarios set different requirements for processor power, power consumption, and other demands of the edge computing device. Edge computing applications based on object-side control have relatively weak demands for computing power, while in other applications, data are processed in a place near to the object to reduce transmission volume and network delay. In applications that require processing of real-time information (including high-definition video monitoring), data should be processed according to the principle of proximity, which imposes high requirements on computing power. Therefore, the processor power of the edge computing device should be provided with an object-side device processor that corresponds to its capability for the application scenario.

2.3 Edge computing algorithms and memory design

An important application scenario of edge computing is to extend the intelligent processing capabilities in the cloud to the object end, and this extension is the so-called intelligent front-loading of computing. However, the size of mainstream deep neural network processing algorithms spans several megabytes to several hundred megabytes, posing a challenge for the processor and memory configuration of the object-side computing nodes. Because it is generally impossible to configure a large processor and memory for the terminal device, compression of the algorithm model should be carried out, and the hard-coded IP of the intelligent processing algorithm should be configured for the object-side processor. Additionally, there is a need to further consider in-memory computing if the very small model is not achieved.

2.4 Capabilities and resource coordination of edge computing and cloud computing

Cloud computing is a centralized computing model, whereas edge computing is essentially a decentralized model. However, cloud computing and edge computing are not completely oppo-

site. In some application scenarios, edge computing is used to perform data preprocessing at the edge of the network before the data are transmitted to the cloud where further deep-scale analysis is performed, thereby using the strengths of each model. Edge computing will realize its full capabilities by effectively coordinating with cloud computing to truly achieve its purpose. For a collaborative process between cloud and edge computing, there is currently no good and unified model for appropriately distributing the computing load and scheduling tasks, or for defining the coordination protocol between the cloud-side and object-side devices.

2.5 System security of edge computing

Because of the natural connection between edge computing and the IoT, mobile Internet, and Internet, there are inherent system security issues. However, because the terminal edge computing system is generally of relatively low capacity, it cannot adopt the multiple layers and multiple security protection measures of conventional computing systems. Some edge computing systems are directly applied to system control, and their security becomes particularly important. Therefore, the security of edge computing systems remains a new research topic.

3 Main progress of edge computing research

At present, research on edge computing mainly focuses on mobile edge networks, fog computing, cloudlets, edge clouds, and architecture. Professor Shi Weisong from Wayne State University is a leader in edge computing research at home and abroad. He has described the research on edge computing technology from various aspects, such as edge computing foundation, edge computing system platform, and typical applications. In [3], mobile edge networks are introduced from the perspectives of fusion computing, caching, and communication. In [4], the cloudlets computing model advanced from cloud computing to the user side is explained. Additional studies on fog computing concepts have been conducted. Reference [5] mainly presents the importance of fog computing in the IoT. Reference [6] describes how to reduce the energy consumption of cloud computing through fog computing, whereas Reference [7] discusses whether fog computing is the future of cloud computing. Reference [8] primarily introduces the concept of, application for, and problems faced by fog computing. In [9,10], the security, privacy, and control methods of fog computing are explored. Reference [11] introduces the fog computing platform and its upper application. Reference [12] introduces the role of edge computing in extending enterprise applications to the edge of the network, whether the delay and performance analysis are introduced in Reference [13]. Reference [14] introduces the rise, definition, architecture, research field, and technical challenges of edge computing. In addition, the Edge Computing Consor-

tium discusses the definition, architecture, and industrial ecology of edge computing. MEC is one of the key technologies for the fifth generation of mobile communications (5G) era [15], as well as a new programming model for large-scale IoT applications [16]. Reference [17] describes the related technologies in greater detail. In summary, these research advances focus on the following aspects.

3.1 General edge computing architecture is clearly defined

Although both fog computing and edge computing have their own areas of focus, they essentially describe the same type of computing model. The current research aims to define the general edge computing architecture horizontally and longitudinally: at the longitudinal dimension, the general edge computing architecture consists of an infrastructure layer, a unified interface layer, and an application service layer; at the horizontal dimension, it represents the dynamic intelligence, security assurance, and operation and maintenance control.

(1) Infrastructure layer: This layer primarily consists of the hundreds of millions of heterogeneous smart terminal devices distributed at the edge of the network, close to users. This layer serves as a pool for the allocation of resources such as computing, storage, bandwidth, and cache, etc., providing the infrastructure for edge users.

(2) Unified interface layer: There are heterogeneous issues in many smart devices that must be integrated so that they can be managed in an effective and unified manner. The goal of this layer is to eliminate these heterogeneous issues so that the infrastructure devices can be interconnected according to a unified and standardized interface, enabling collaborative management of resources.

(3) Application service layer: Close to the application service provider, this is the upper application and system program that can satisfy user needs. This layer distributes, uploads, and delegates tasks—including load, application, and services—to other resource providers, to meet the actual needs of users.

(4) Dynamic intelligence layer: The layer horizontally covers the three longitudinal layers. This layer is designed to manage and control its devices, interfaces, and applications efficiently, intelligently, and automatically, and relieve the labor force of cumbersome configuration and supervision to realize intelligent computing.

3.2 Mobile edge computing has become the focus of research and a unified edge computing reference frame has been proposed

As of now, MEC has evolved into a key component of future mobile broadband networks. Together with software-defined network/network function virtualization, MEC will become one of the key technologies for the next generation of mobile com-

munications networks. Additionally, the consensus reached by the industry is: MEC is a key enabler for the IoT and vertical industry communication with low delay and high reliability. It has numerous application scenarios in many industries. MEC is also recognized by the industry as one of the key architectural concepts and technologies of 5G. Presently, the European Telecommunications Standards Institute (ETSI) has defined the computing framework for MEC (Fig. 2) and is now accelerating the standardization of MEC platforms and interfaces. In the MEC platform, functional modules such as mobile edge host, mobile edge platform management, mobile edge programming, and virtual infrastructure management have been defined [18,19].

(1) With the mobile edge host module, the capacity of the mobile edge platform and the data-forwarding capability of the virtual infrastructure are open to public use, and the deployment of the mobile edge application is realized.

(2) With the mobile edge platform management module, mobile edge platform network element management, mobile edge application life cycle management, mobile edge application rules, and demand management are realized.

(3) With the mobile edge programming module, deployment, instantiation, and standardization of the mobile edge application in the global scope are realized.

(4) With the virtual infrastructure management module, to achieve unified allocation, management, and configuration of virtual resources of the infrastructure, as well as collection and reporting of virtual resource performance and failure, are realized.

3.3 Multiple solutions for enabling technologies of edge computing

For enabling the technology of MEC, China Unicom has proposed a specific solution that divides the enabling technology into three aspects: cloud and virtualization, high-capacity servers, and enable applications and service ecosystems. Cloud and virtualization technology and large-capacity servers are targeted for edge computing hardware. Enable applications and service ecosystems bring software and application providers to the edge computing market and provide rich applications to create an ecosystem. In other words, a wealth of edge computing apps is provided, as well as a variety of programming models for app programming interfaces (APIs) based on open standards, related tool chains, and software development kits.

For the terminal computing capability of edge computing, the existing idea is to perform different resource configurations for different application scenarios. If it is a kind of control-purposed edge computing scenario, a weak performance microcontrol unit (MCU) for the terminal is enough. As for the applications with a high demand for terminal computing performance, one solution is to use a new instruction set to strengthen support for vector computing, or develop SIMD-like architecture with multiple

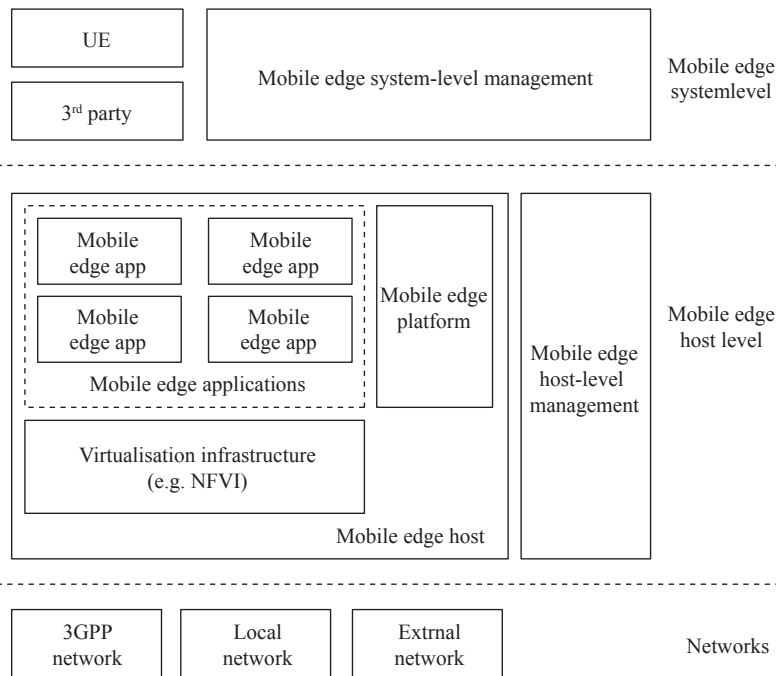


Fig. 2. MEC computing architecture defined by ETSI.

Note: VFVI refers to network function virtualization infrastructure solution

cores to strengthen the MCU; the second solution is to adopt heterogeneous computing, in which the MCU is simply responsible for control, and a dedicated accelerator IP is responsible for the computing task. In fact, most current artificial intelligence (AI) chips are subject to such dedicated AI algorithm accelerator IPs. The Cambrian IP of the Institute of Computing Technology, Chinese Academy of Sciences, which is embedded in Huawei mobile phones, is one such example. Additionally, the memory configuration scheme for terminal computing is important. Because terminal node computing cannot be added to dynamic random-access memory due to many factors, such as cost, volume, and power consumption, the flash storage (used for storing operating systems, etc.) is generally used as system memory. Since the cache must be completed on the processor chip and the cache is generally small, the algorithm must be able to compress the size of the model—that is, to perform “model compression.” It is also important to use in-memory computing (currently under study) and new types of memory including nonvolatile magnetic random-access memory and resistive random-access memory on high-density chips.

4 Influence of the development of edge computing technology

4.1 The edge cloud will contribute to decentralized construction of cloud computing center from a centralized manner and functional decoupling

The edge cloud is the expression of edge computing in the cloud architecture system. Instead of a unified concentration

of multiple devices into one region, the cloud computing center adopts a star topology with multisites and multicenters and connects numerous edge clouds with the center. Thus, the edge cloud distributed in other regions can provide nearest services for local users, avoiding excessive bandwidth consumption, data transmission, and access pressure. This model can balance the load, alleviate the excessive consumption of resources, and effectively improve the operating efficiency. Furthermore, the functions of the current cloud computing center are also decoupled. Put differently, edge computing can effectively divide the functions of the infrastructure. The resources in each region can be modularized, customized, and simplified for tasks, reducing the degree of coupling of applications, data, and services.

4.2 Edge computing makes a shift in computing capacity from centralization to decentralization

Edge computing is a kind of technical means to solve excessive pressure on cloud computing, under-utilization of resources, low reliability, poor usability, and insufficient bandwidth. It transforms the originally centralized advantages into distributed ones under the new form of the IoT. Because cloud computing cannot include all the massive intelligent terminals, and the computing tasks that should originally be solved at the cloud side can now be easily resolved locally on the terminal side (owing to the stronger edge device computing and storage capabilities), the overall computing model has undergone major changes from the originally centralized computing to distributed computing. Such a trend will have an influence on a number of problems arising from information construction. For example, the scale of equip-

ment investment at the cloud side will be greatly reduced along with bandwidth requirements and storage pressure.

4.3 Edge computing contributes to a shift of IT resources from isolation to coordination

Because of the massive construction scale of traditional cloud computing centers, big data centers, and supercomputing centers—and difficulty in nearby use due to isolation of users—significant amounts of money invested in infrastructure are wasted, and resources are not fully used. Alternatively, edge computing can break down the fragmented information-based construction model, enabling each branch of an organization to be interconnected and interoperable. This would eliminate isolated resource islands, contributing to complementary advantages and collaboration between the originally segregated resources.

4.4 Edge computing contributes to a shift in information system security from single centralization to burden sharing

The edge computing construction model will break down the “one size fits all” model for information construction, so that risks and hidden dangers can be distributed to other parts, ensuring information security. In the traditional centralized model, there is often a need for security-level indicators. Once a problem arises, cloud computing and data centers will be affected as a whole. The edge computing model can alleviate this burden and greatly reduce the risk of information leakage by allowing users to save their own data, instead of them being stored on the cloud side.

5 Opportunities and measures for development of edge computing

In February 2017, the Computer Community Coalition in America issued the Report on the Major Challenges of Edge Computing, which described the main challenges of edge computing in terms of applications, architecture, capabilities, and services. These challenges are summarized mainly as the following: multisource, heterogeneous and off-site management of edge devices, quality of information service guarantee of edge computing, data privacy and information security at the edge, distributed collaborative computing on the cloud and edge, and intelligent situational awareness and unified open platform. In the future, the development of the IoT, mobile Internet, and industrial Internet will enable the edge computing model to gradually break down the cloud computing model—simply centered on the Internet data center—and eventually form a complementary situation. With capabilities upgrading of mobile devices in storage, computing, security, the terminal node will do more works. Edge computing has developed toward the decentralization of computing models, and collaborative com-

puting will be the trend of future technology. Massive terminals will have an impact on AI, machine learning, and other technologies [20], promote the development of microkernel technology, and facilitate the embedding of algorithms and models into the firmware of massive devices. This will make front-end intelligence more promising; the openness, versatility, compatibility, interactivity, and security of the edge computing platform will be the problems to solve in the future [21]. If these challenging technical problems can be broken through, this will bring new opportunities for the development of the “Internet Plus,” as well as opportunities for industrial development. Iain Gillott, founder of iGR—a wireless and mobile industry research institute—stated at the Global Mobile Edge Computing Conference, held from September 25 to 27, 2017, that edge computing has changed the economic shape of mobile networks. It is estimated that during 2017–2026, the United States will spend 87 billion USD on edge computing, and Europe will spend 185 billion USD. Therefore, to respond to new development opportunities, it is suggested that the following measures be taken for the development of edge computing technology in China.

5.1 Strengthening the technical standards and norm construction of edge computing

Edge computing that involves a large number of terminal devices and edge nodes is the front-end of data collection, data aggregation, data integration, and data processing. These devices are often heterogeneous and come from different manufacturers with different data interfaces, data structures, transmission protocols, and underlying platforms. Thus, there is an urgent need for unified technical specifications and standards, which will greatly reduce the construction cost of the edge cloud.

5.2 Combining R&D and application of edge computing technology with R&D plans of “Internet plus,” cloud computing, big data, and next-generation communications technologies

Edge computing is a technology that is closely related to cloud computing and is highly associated with big data, 5G communications, and intelligent information processing technologies. Therefore, when formulating related research and development (R&D) plans, China should consider the R&D activities of edge computing technologies and applications to accelerate the R&D of related core technologies and enhance the sophistication of edge computing technologies.

5.3 Strengthening open-source ecological construction of edge computing

Edge computing is composed of massive terminal devices that can be adopted by a unified open-source operating system

from an open-source ecological environment. Under such a trend, all manufacturers will gain equal opportunities in maintaining the core code within the open-source ecology, to form the industry-recognized technology interfaces, key functions, and development paths.

References

- [1] Shi W S, Liu F. Edge computing [J]. Beijing: China Science Publishing & Media Ltd., 2018. Chinese.
- [2] Mung C, Bharath B. Fog computing: Technology, architecture and applications [M]. Translate by Yan S, Peng M G. Beijing: China Machine Press, 2018. Chinese.
- [3] Wang S, Zhang X, Zhang Y, et al. A survey on mobile edge networks: Convergence of computing, caching and communications [J]. *IEEE Access*, 2017, 5(99): 6757–6779.
- [4] Verbelen T, Simoons P, Turck F D, et al. Cloudlets: Bringing the cloud to the mobile user [C]. *ACM: ACM Workshop on Mobile Cloud Computing and Services*, 2012: 29–36.
- [5] Bonomi F, Milito R, Zhu J, et al. Fog computing and its role in the internet of things [C]. *ACM: MCC Workshop on Mobile Cloud Computing*, 2012: 13–16.
- [6] Fatemeh J, Kerry H, Robert A, et al. Tucker, “fog computing may help to save energy in cloud computing” [J]. *IEEE Journal of Selected Areas in Communications*, 2016, 34(5): 1728–1734.
- [7] Firdhous M, Ghazali O, Hassan S. Fog computing: Will it be the future of cloud computing? [C]. *The Third International Conference on Informatics & Applications (ICIA2014)*, 2014.
- [8] Yi S, Li C, Li Q. A survey of fog computing: Concepts, applications and issues [C]. *ACM: The 2015 Workshop on Mobile Big Data*, 2015: 37–42.
- [9] Dsouza C, Ahn G J, Taguinod M. Policy-driven security management for fog computing: Preliminary framework and a case study [C]. *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on IEEE*, 2014: 16–23.
- [10] Yi S, Qin Z, Li Q. Security and privacy issues of fog computing: A survey [C]. *WASA*, 2015: 685–695.
- [11] Yi S, Hao Z, Qin Z, et al. Fog computing: Platform and applications [C]. *Hot Topics in Web Systems and Technologies (HotWeb), 2015 Third IEEE Workshop on IEEE*, 2015: 73–78.
- [12] Davis A, Parikh J, Wehl W E. Edge computing: Extending enterprise applications to the edge of the internet [C]. *ACM: The 13th International World Wide Web Conference on Alternate Track Papers & Posters*, 2004: 180–187.
- [13] Malandrino F, Kirkpatrick S, Chiasserini C F. How close to the edge? delay/utilization trends in MEC [C]. *ACM: ACM Workshop on Cloud-Assisted Networking*, 2016:37–42.
- [14] Shi W S, Cao J, Zhang Q, et al. Edge computing: Vision and challenges [J]. *IEEE Internet of Things Journal*, 2016, 3(5): 637–646.
- [15] Hu Y C, Patel M, Sabella D, et al. Mobile edge computing—A key technology towards 5G [J]. *ETSI White Paper*, 2015, 11(11): 1–16.
- [16] Hong K, Lillethun D, Ramachandran U, et al. Mobile fog: A programming model for large-scale applications on the internet of things[C]. *ACM: The Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, 2013: 15–20.
- [17] Patel M, Naughton B, Chan C, et al. Mobile-edge computing introductory technical white paper [R]. *Mobile-Edge Computing (MEC) Industry Initiative*, 2014.
- [18] Edge Computing Consortium, Industrial Internet Industry Alliance. Edge Computing Reference Architecture 2.0 white paper [R]. *Edge Computing Consortium, Industrial Internet Industry Alliance*, 2017. Chinese.
- [19] Lv H Z, Chen D, Fan B, et al. Standardization progress and case analysis of edge computing [J]. *Journal of Computer Research and Development*, 2018, 55(3): 487–511. Chinese.
- [20] Zhou X R, Zhao L Y, Zhao D, et al. The promotion effects of artificial intelligence on e-science [J]. *E-Science Technology & Application*, 2016, 7(6): 14–26. Chinese.
- [21] Yang C, Yuan B, Tian Y, et al. A smart home architecture based on Internet of things identifier [J]. *E-Science Technology & Application*, 2015, 6(1): 80–88. Chinese.