DOI: 10.3969/j. issn. 1673-5188. 2017. 02. 004
http://kns.cnki.net/kcms/detail/34.1294.TN.20170419.1031.002.html, published online April 19, 2017

Key Technologies and Application of Edge Computing

TU Yaofeng, DONG Zhenjiang, and YANG Hongzhang
(Nanjing R&D Center, ZTE Corporation, Nanjing 210021, China)

Abstract

Cloud computing faces a series of challenges, such as insufficient bandwidth, unsatisfactory real-time, privacy protection, and energy consumption. To overcome the challenges, edge computing emerges. Edge computing refers to a process where the open platform that converges the core capabilities of networks, computing, storage, and applications provides intelligent services at the network edge near the source of the objects or data to meet the critical requirements for agile connection, real-time services, data optimization, application intelligence, security and privacy protection of industry digitization. Edge computing consists of three elements: edge, computing, and intelligence. Edge computing and the Internet of Things (IoT) mutually create, and edge computing and cloud computing complement each other. In the architecture of edge computing, resources are distributed to the edge nodes, and therefore the storage system is near users while the computation function is near data. In this way, the stress on the backbone network can be lessened. With this architecture, the existing key technologies for computation, networks, and storage will change significantly. ZTE's edge computing solutions can ensure the service quality of operators and greatly enhance the experience of mobile users.

Keywords

edge computing; cloud computing; IoT

The era of the Internet of Things and big data is coming and many industries are moving toward digital transformation. Storage and computing are increasingly dependent on the cloud, which makes the cloud become a hub of constructing the information society. However, the cloud cannot solve all the problems. The network edge side of intelligent Internet faces the following challenges:

1) Insufficient bandwidth for the connection and analysis of massive data: The explosive growth of IoT terminals and data densely occupies the bandwidth at the network edge, and the continuous development of the communication technologies and the increasingly rapid growth of network construction still cannot afford the massive transmission caused by the connection between man and things, man and man, or things and things.
2) Steady increase of the demand of real-time experience: Many data streams are generated by edge equipment. However, it is difficult to make real-time decisions through "remote" cloud computing and analysis. For example, the response time of the visual service of a wearable camera must be within 50 ms, which cannot be met if cloud computing is used because the delay is quite long.
3) Security and privacy protection: Security is a basic requirement for both cloud computing and edge computing and therefore end-to-end protection is required. It is difficult to greatly and extensively improve the access control and threat protection of the network edge because the network edge is closer to the IoT devices.
4) Constant energy consumption: As more and more applications are transferred to the cloud, energy demands are growing. Therefore, the computing strategy for maximizing the energy efficiency becomes a particularly urgent demand. It is not necessary to transmit each piece of original data to the cloud during the collection and processing of basic information about some embedded small devices, saving a lot of energy costs.

To overcome the network-edge challenges to the cloud, the industry and the academia attempt to build a new architecture called edge computing.

1 Definition of Edge Computing

Edge computing [1] refers to a process where the open platform that converges the core capabilities of networks, computing, storage, and applications provides intelligent services at the network edge near the source of the objects or data to meet the critical requirements for agile connection, real-time services, data optimization, application intelligence, security and privacy protection of industry digitization.

To implement edge computing, the following three elements are required: edge, computing, and intelligence [2].

1) Edge: For cloud computing, all data is gathered to the back-end data center for processing, which is a cloud-focused process. For edge computing, however, the focus is the physical area of the "edge" (or "end"). It is obvious that the timeliness requirement of the real-time services is more likely to be satisfied if the network, computation, and storage resources can be provided nearby for the "edge".
2) Computation: Computation is of great importance during edge computing. Because the computation is implemented on the edge, the bottleneck problem concentrating the computing power to the cloud can be avoided.
3) Intelligence: Edge computing covers the operation technology (OT), information technology (IT), and communication technology (CT) fields, and involves network connection, data collection, sensing, chips, and industry applications. The autonomous management and orderly collaboration of various heterogeneous platforms, and the collection, optimization, and unified presentation of heterogeneous data can be implemented only through edge intelligence. This is another difference from cloud computing.

In summary, compared with the large and comprehensive functions implemented by cloud computing, those implemented by edge computing are said to be smaller and more beautiful. Based on the data source, edge computing is complementary with cloud computing applications in a real-time and quick way.

2 Value of Edge Computing

2.1 Mutual Creation of Edge Computing and IoT

Edge computing makes the IoT smarter. The core of the IoT is to implement smart connection and operation [3] of each object while that of edge computing is to implement the sensing, interaction, and control between objects through data collection and analysis. Edge computing is the key to the application of the IoT and can implement the comprehensive perception, seamless interconnection, and high intelligence of the smart terminals in the autonomous areas.

Edge computing is one of the key technologies used in the IoT and big data. The essential changes occurring during the evolution from the Internet, mobile Internet, to IoT are the amount of data: the rapid increase of the number of devices connected to the Internet and the data generated per unit. With the mass deployment of IoT terminals, the data generated by the IoT will experience an exponential growth. Data filtering and intelligent analysis on the network edge through scattered terminals and IoT is an important development direction of IoT applications.

Edge computing can enhance the efficiency and security of the IoT. Edge computing is already successfully used in efficient energy and intelligent manufacturing. After the intelligent transformation based on the IoT technologies, the data analysis, processing, and monitoring on the network edge can improve both the efficiency and the response speed. In addition, data gradually becomes an important asset of enterprises. In the case of network or data center failures, the intelligent network edge can ensure data security to help enterprises avoid risks.

2.2 Edge Computing and Cloud Computing Complement Each Other

Not all computations need the cloud. Cloud computing is a basic technology, and almost all data must be connected to the cloud for storage, computation, and analysis. However, cloud computing is not the only solution. The types of computation typically not suitable for the cloud are as follows:

1) Time-delay sensitive computation: With traditional cloud computing, after you click a button, the back-end system makes a computation and then responds with the result. However, some services with real-time responses (such as online videos, augmented reality, and virtual reality) have high requirements for the delay, caching, and security. If the service processing completely depends on the heavyweight cloud computing far away from customers, bottlenecks are inevitable.
2) Low value density: In a traditional cloud computing model, terminals can only collect data. All big data is analyzed and processed after being transmitted to the cloud through the network, which puts a huge interaction pressure on the network. If there is only a small amount of valuable data, the analysis and processing of data are a waste of bandwidth. Therefore, it is reasonable to transmit only valuable data.
3) Emergency power outage: In the case of power outage, the advantages of the autonomous system of edge computing can be developed. In the traditional cloud computing architecture, a terminal can hardly cope with a disaster because it totally depends on the cloud for data processing.

Moreover, the advantages of edge computing and cloud computing can be combined, and the cloud and fog can collaborate with each other.

The processing and storage of massive data must rely on a solid cloud platform. Focusing on the big data analysis of non real-time and long-period data, cloud computing can take advantage of "logic concentration of resources" in periodic maintenance and service decision-making support. Compared with cloud computing, edge computing is secure, fast, easy to manage, and more applicable to the intelligent processing and execution of local services in real time.

The two solutions can be highly collaborative based on the complementary advantages. On the one hand, close to execution units, edge computing can act as the collector of the high-value data required by the cloud to better support the big data analysis of cloud applications. On the other hand, the service rules output through big-data analysis and optimization of cloud computing can be delivered to the edge for the optimization and processing of service execution by edge computing based on new service rules. The deployment of lightweight edge computing on the edge of networks will undoubtedly reduce the burden of the upper cloud computing center.

In most cases, edge computing needs to interact with cloud computing (**Fig. 1**) [4]. Analysis on the edge combines in-depth analysis on the cloud.

2.3 Edge Computing and Cloud Computing Jointly Promote the Collaboration of Cloud and Fog

The concept of fog computing [5] was put forward by Cisco in 2011. The fog mainly uses the equipment on the network edge. The equipment can be any traditional network device, such as routers, switches, and gateways, or the local servers that are specially deployed. In general, the deployment of special equipment needs more resources, while the sharing of abundant traditional network equipment can greatly reduce the cost. The resource capacity of a single specially-deployed device and a single traditional network device is far less than that of a data center; however, a huge number of the two devices can produce enormous resource capacity. A fog platform consists of a vast number of fog nodes. In contrast to the data center where resources are gathered, the fog nodes can be geographically distributed in different locations.

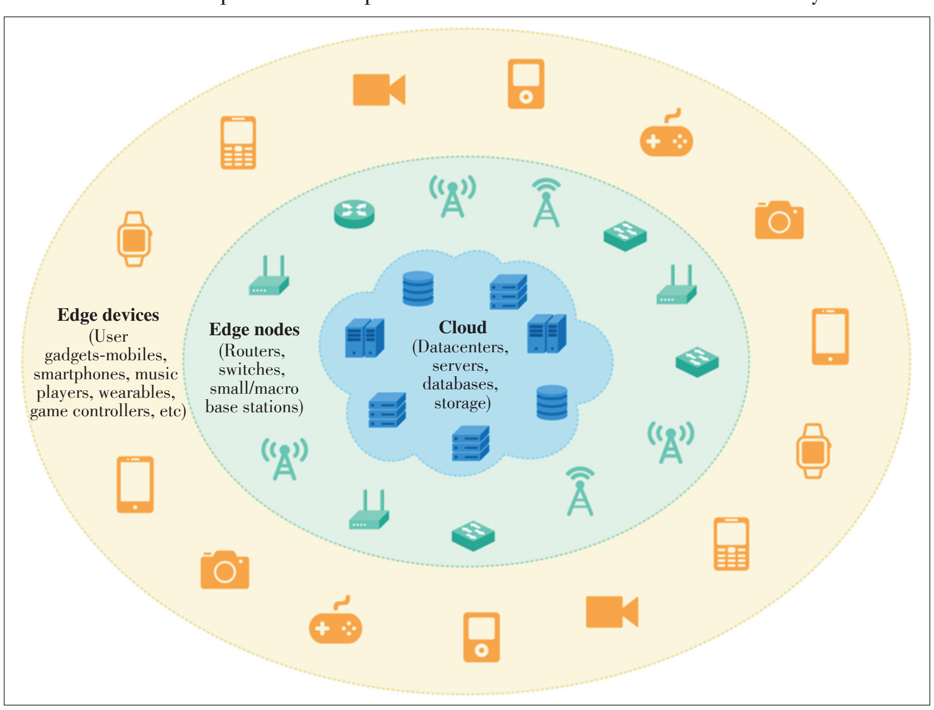
Ginny Nichols, who proposed the concept of fog computing, has an interesting view: the fog is the cloud close to the ground. In other words, the common point of fog computing and cloud computing is that they provide the resources from the shared resource pool for multiple users based on the virtualization technology. The difference between fog computing and cloud computing is their positions in the network topology.

However, edge computing is essentially different from fog computing and cloud computing because it does not target resource virtualization. Edge computing is a solution to processing real-time big data and aims to reduce the great pressure on the data center that processes massive data alone in the Internet of Everything (IoE) era, the advanced stage of the IoT.

Sharing similar ideas, both edge computing and fog computing provide computation near the on-site applications. In terms of the nature, both edge computing and fog computing are the counterpart of cloud computing and will be definitely integrated into the architecture of "collaborative cloud-fog development."

3 Key Technologies of Edge Computing

In the architecture of cloud computing, the data center is the controller of resources and users apply for resources on demand. User data is uploaded to the data center for computation and processing and the data center finally sends a feedback of the processing result to the user through the network. However, in the architecture of edge computing, users do not need to entirely rely on the data center because resources are distributed to the edge nodes. Therefore, the storage system is near users while the computation function is near data, meeting the real-time requirements of the network and effectively using the computing resources. With this architecture, the existing key technologies for computation, networks, and storage will change significantly.



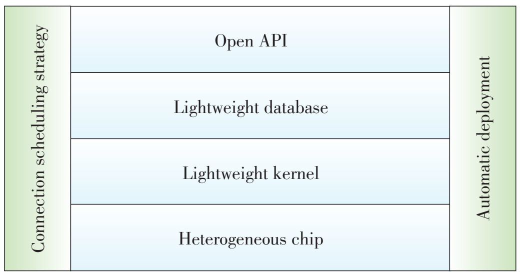
▲Figure 1. Edge computing architecture.

3.1 Key Computing Technologies

In the resource-gathered cloud computing architecture, the computing model is relatively mature, and the computing resources are abundant. In the edge computing architecture, however, the edge nodes and the technology capabilities vary significantly, and some key technical challenges (**Fig. 2**) need to be overcome, such as intelligent connection scheduling, parallel processing, and automatic deployment.

1) Intelligent connection scheduling policy

For edge computing, the biggest difficulty lies in how to dynamically deploy the computation and storage capacity on a large scale, and how to implement efficient coordination and seamless connection between the cloud and equipment [6]. The continuous development of distributed computation leads to many tech-



▲**Figure 2. Key computing technologies.**

nologies to drive split task execution in many geographic locations. Task splitting is usually clearly specified in the programming languages or management tools. However, the edge node-and splitting-based computing poses some challenges, such as how to effectively split computing tasks, and how to automatically calculate the capabilities and locations of edge nodes without clearly specifying. Therefore, a new scheduling method is required to deploy the split tasks to every edge node. An ideal strategy is that the edge side can fully intelligently determine when to use edge nodes and which data is processed with edge nodes.

Another challenge is how to ensure that the edge nodes can still reliably operate when making additional computation. For example, if a base station is overloaded, the edge devices connected to the base station may be affected. Therefore, the operation of the edge nodes needs to be intelligently sensed and controlled and the tasks need to be flexibly split and scheduled. How to integrate and separate a complex algorithm between the cloud and edge devices can be realized only with the technologies that can simultaneously control the cloud, channels, and devices.

2) Computation of heterogeneous nodes

"Heterogeneity" is the most distinctive characteristic of massive edge devices and the edge devices manufactured by most of the vendors support general-purpose computation through software solutions. With the increase of the number of edge nodes supporting general-purpose computation, the demand for the development frameworks and toolkits will also rise. The programming model needs to support the parallel processing of tasks and data by the edge nodes, and simultaneously makes computations on the hardware at multiple levels. The programming language needs to consider the heterogeneity of hardware and the computing power of various resources in the workflow. This is more complicated than the existing cloud computing model. The container technology is becoming mature. The mobile containers that reuse hardware across multiple virtual devices can provide the performance equal to that of local hardware and can quickly deploy applications on heterogeneous platforms. The container- or virtualization-based solutions can overcome the challenges to the deployment of applications on the heterogeneous nodes.

3) Lightweight database and kernel

Unlike large servers, edge nodes may not support large software due to hardware limitations. For example, the small Intel T3K base station with the concurrent dual-mode SoC is designed with a four-core ARM CPU and limited memory, and cannot afford complex data processing. Apache Quarks, a lightweight database, can be used on small edge devices (for example, smart phones) for real-time data analysis. However, a single Quarks cannot perform advanced analysis tasks. The lightweight database that consumes less computation and storage resources is more suitable for the application edge computing.

4) Automatic deployment and service discovery

By 2020, about 50 billion terminals and devices will be connected to the Internet. The shorter product life cycle, higher individual demand, and more obvious trend of whole-life-cycle management and service need an automatic deployment mechanism to provide a powerful technical support to deal with routine operation and maintenance, such as rapid deployment in batches, intelligent configuration, automatic troubleshooting, and service recovery. In addition, how to find resources and services in an edge computing environment is also an area for further development.

3.2 Key Storage Technologies

Focusing on storage, the existing distributed storage architecture is suitable for centralized computing systems [7]. With the in-depth development of edge computing, the computation capability is shared by the edge. Storage in the future, especially the local file storage system on the edge, will focus on computation. The transformation from the storage-centric mechanism to the computation-centric mechanism is a reverse of the design idea of the existing storage systems.

1) Data distribution

This is a basic issue that must be considered for a distributed storage system at the beginning of the design. The purpose of typical data distribution algorithms, such as Controlled Replication under Scalable Hashing (CRUSH), Distributed Hash Table (DHT), and consistent Hash, is to evenly distribute data to every node of the system after splitting and fully breaking up the data. Such data distribution can equalize the computing power of the central computing node. However, for the architecture where edge nodes are involved in computation, data should be stored where it is needed rather than in a random place. In other words, data should be stored where it is computed. The focus should be the reduction of the computation delay instead of data equalization.

2) Data consistency in a distributed environment

In a distributed scenario, data typically has multiple copies and the copies may be read and written at the same time. Therefore, the resulting data consistency becomes a long standing problem for the distributed storage system. In the edge computing architecture, data is accessed on the edge instead of by other clients, so the traditional consistency guarantee mechanism is not required. A new architecture will be used in the future.

3) New requirements for storage software from new storage hardware

Edge computing is a delay-sensitive application. Especially for the Internet and embedded applications, it is a general trend in storage devices to replace the mechanical disks with flash drives. However, the design of the existing storage systems depends too much on the characteristics of disks instead of flash drives. With the continuous development of edge computing, the high-speed, energy-saving, and small flash drives will be heavily deployed on the edge nodes. Whether a single disk or a full-flash server need matched storage software, flash-oriented software storage systems will become a key technology of edge computing in the future.

3.3 Key Network Technologies

With the shifting of the storage and computing resources from the cloud data center down to the edge nodes, the network bottlenecks migrate from the backbone networks to the edge nodes. In this case, the internal and external interaction on the servers greatly increases, which imposes a demanding requirement that cannot be met by traditional TCP/IP technologies. To respond to this challenge, InfiniBand (IB), Remote Direct Memory Access (RDMA), and Data Plane Development Kit (DPDK) become key network acceleration technologies of edge computing.

1) With RDMA, data is directly transferred to the storage area of a computer through the network, which means that data is quickly moved from one system to the memory of a remote system without affecting the operating system. Therefore, this technology does not have a high requirement for the processing capability of computers. It removes the text copy and exchange operations on external memories and therefore can improve the system performance because it can reduce the occupation of the memory bandwidth and CPU cycle time [8].
2) The InfiniBand is a cable conversion technology that supports concurrent links. With the high bandwidth, low delay, and high expandability, InfiniBand is applicable to the communications between a server and server, a server and storage device, and a server and the network.
3) DPDK is an application development kit provided by Intel to increase the data processing speed of data-plate packets. Supported by Userspace Input/Output (UIO), DPDK supports the drivers in the application space and the network adapter driver operates in the user space, reducing the number of duplications of the packets between the user space and application space. With the Linux affinity, DPDK binds the control plane threads and data plane threads to different CPU cores, reducing the scheduling of CPU cores among the threads. It provides memory pool and lock-free ring buffer management to improve the memory access efficiency.

3.4 Key Security Technologies

Security is a basic requirement for both cloud computing and edge computing and therefore end-to-end protection is required. It is difficult to greatly and extensively improve the access control and threat protection of the network edge because the network edge is closer to the IoT devices. Edge security mainly involves the security of devices, networks, data, and applications. In addition, the integrity and confidentiality of key data are the focus of security.

If terminals act as shared edge nodes, the associated risks for the users and owners of the edge devices need to be defined. If the terminals act as edge computing nodes, the original functions of the devices cannot be damaged. For the multiple users on the edge nodes, security is the first priority. Therefore, the minimum service level should be guaranteed for the users on the edge nodes.

4 ZTE Edge Computing Solution

With the rapid evolution of mobile data networks from 2G to 4G and the wide spread of mobile smart terminals, man enters a booming mobile Internet era (**Fig. 3**). Traffic rise and abundant new services place an increasing pressure on the networks and services of the operators. From the perspective of both the revenues and user development, the mobile operators are more concerned about user experience while content providers want the operators to provide a more intelligent channel, which creates a good opportunity for the development of edge computing. ZTE's 5G Mobile Edge Computing (MEC) solution and MEC content delivery network (CDN) solution can bring superior service experience for users and safeguard the operators' networks. The two solutions were put into commercial use in some offices at home and abroad.

4.1 5G MEC Solution

The application of edge computing in mobile communications is called mobile edge computing [9]. MEC can provide IT services and cloud computing functions for telecom users through the nearby wireless access networks, creating a carrier-class service environment with a high bandwidth, low delay, and high performance, accelerating the download of the contents, services, and applications in the networks, and let users enjoy uninterrupted and quality network access.

As the leader and initiator of MEC, ZTE actively participates in and leads the establishment of the standards and specifications for MEC in ETSI and 3GPP. Based on the cloud architecture, the 5G MEC solution integrates computation, storage, networks, and hardware resources (for example, special functional units) into a natural, flexibly-scheduled, and easily-extended resource pool, implements the componentization of software units and Application Programming Interface (API)

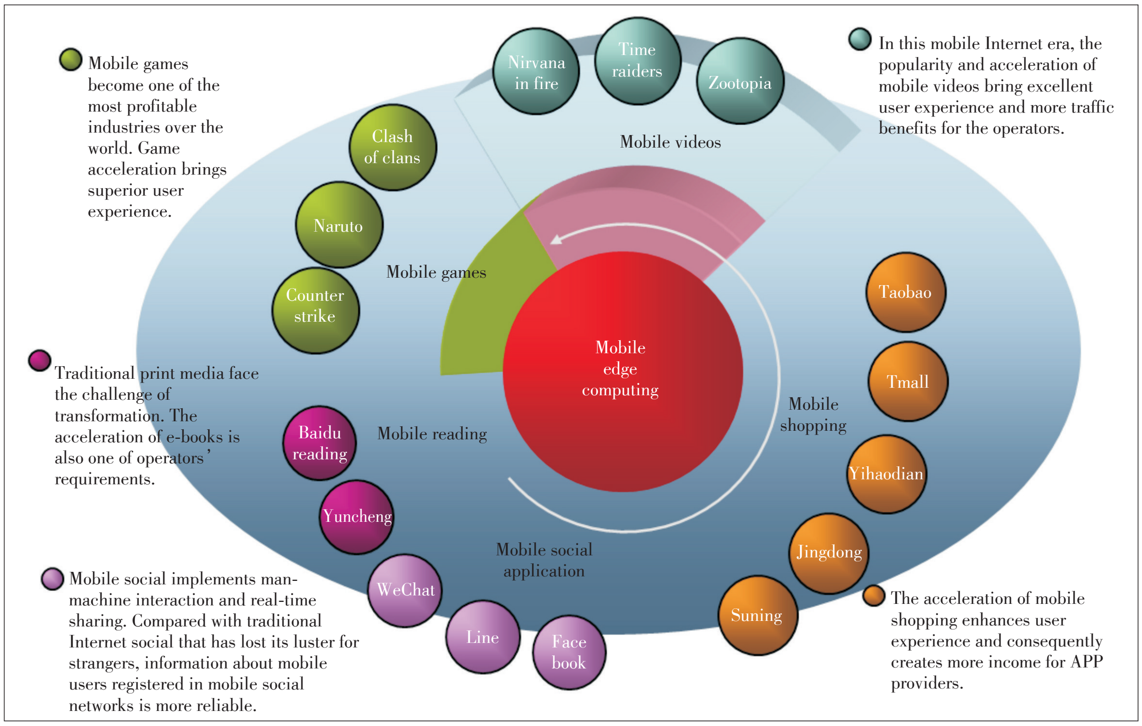


Figure 3. Service application requirements on mobile networks.

based capability opening units on the central Information And Communications Technology-Platform as a Service (ICT-PaaS) platform [10], and realizes integrated deployment, monitoring, management, and operation. Based on the ICT-PaaS platform architecture, the edge data center, region data center, and core data center can be deployed at multiple levels through component collocation, implementing multi-level distributed cloud and fog deployment. Each component of the Virtualized Network Function (VNF) can be flexibly deployed in different data centers based on their service processing features, resource requirements, and user experience. Through component sharing, the common function modules of different VNFs are extracted as common components that are shared among various service flows, greatly improving the processing efficiency of the whole network. After the componentization, operators can flexibly combine independent virtual networks according to the capacity and function requirements of individual users, home users, and enterprise users, and automatically slice networks based on traffic volume (elastic scaling) [11]. In general, the 5G MEC solution realizes the convergence and development of network slicing, software-defined networking (SDN)/network function virtualization (NFV), cloud computing, and virtualization. It can meet the requirements for on-demand network slicing, quick service chains, dynamic orchestration, fast introduction of new features, maximum level of software and hardware decoupling, and strict separation between forwarding from control, finally implementing end-to-end service delivery and full-range network-wide service delivery.

The 5G MEC solution deeply integrates wireless networks and the Internet. The MEC nodes are deployed near the radio access network (RAN) side and provide resident services, such as the LBS service and RAN cache service, meeting the strict requirements of the online video, augmented reality (AR), and virtual reality (VR) services in the era of big data for caching, delay, policy control, and security, saving bandwidth, and improving user experience [12].

Independent of the radio link technologies, the 5G MEC solution can also be implemented and deployed in 4G and 4.5G networks. MEC devices are loosely coupled with the existing 4G network devices in terms of the standards. In addition to customized applications, the standard use cases do not affect the network quality and versions of the existing 4G networks.

The 5G MEC solution (**Fig. 4**) provides an open platform that allows third parties to rapidly develop and deploy innovative applications and services, promoting the sound development of the MEC application ecosystem, facilitating the innovation of mobile edge service interfaces, and providing abundant applications.

The 5G MEC solution provides third-party applications with the capability engines for the integrated positioning, video, IoT services to accelerate the development and deployment of the third-party applications and improve user experience. It has the following characteristics:

1) Positioning and tracing: The base stations can provide the positioning function accurate to within five meters. This is a specific function of mobile networks. This function can help businesses to push advertisements for nearby business areas and stores to mobile users, bring operators and service providers backward advertising revenues. It can also be used for football match analysis and campus management.

D:\EMAG\2017-04-56/VOL15\F3.VFT——9PPS/P6

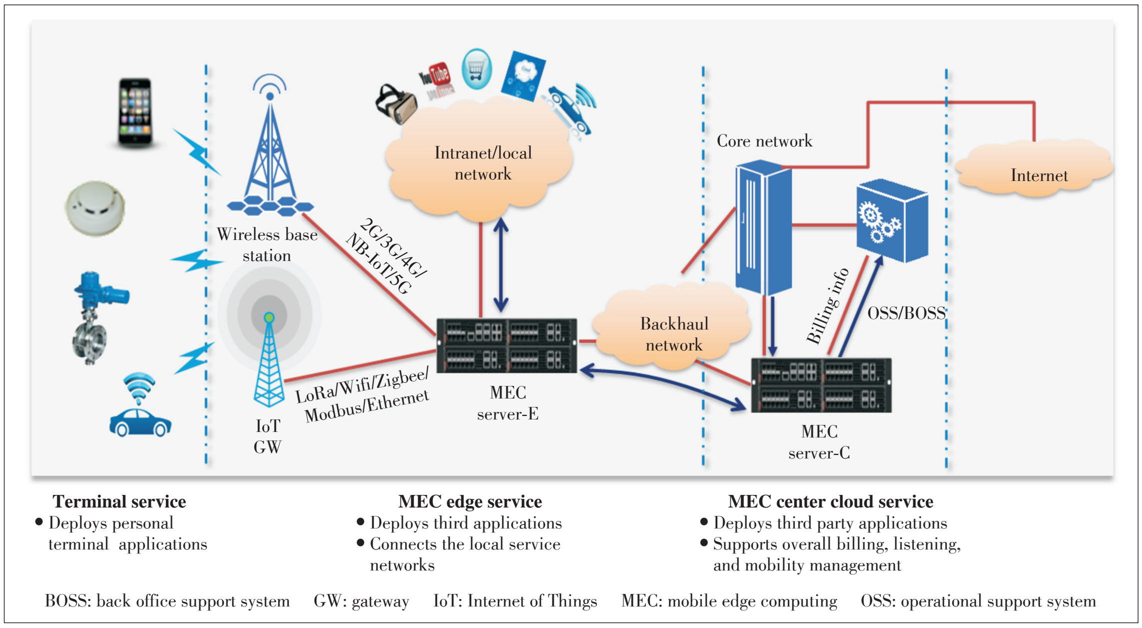


Figure 4.
5G MEC solution.

2) Video analysis: The MEC server side implements image management (camera management and image capture) and image analysis. The users of the information are independent of MEC and can be a smart city, retailer, advertisement, and related APP or server. They can be public places such as airports and railway stations for managing lost luggage. They can also be the security personnel of public places for locating missing children and old men. Video analysis can also be used in other scenarios, such as high-speed monitoring and building security, to help the construction of smart cities.

3) Optimization of wireless network perception: Apps are deployed on the MEC server side for wireless network perception and content optimization servers are deployed in the core network. The apps transmit the information about user IDs, cell load, and link quality to the content optimization servers to help the servers to dynamically optimize contents based the cell load and link quality, improve the QoE and network efficiency, shorten the time delay, and provide optimal user experience.

4.2 MEC CDN Solutions

The CDN is used to clear the Internet bandwidth bottleneck. The basic idea of the CDN is to store the hot-spot content on the server near the access side of the network. When users access the hot-spot content, they no longer access the servers on backbone side, reducing the demand for the traffic from the backbone network and improving the QoS. Through the intelligent virtual network on the existing Internet formed by server nodes, the CDN system can re-direct the user requests to the nearest server node in real time based on the network traffic, node connections, load on the nodes, the distance to the users, and the response time to relieve network congestion and improve the response to users [13].

Both the CDN and edge computing focus on the edge side. The difference lies in that the CDN focuses on storage while edge computing on computation based on storage. Therefore, it is extremely advantageous to introduce edge computing to the existing CDN service.

With the development of mobile Internet and HD videos, the content, users, and service scope of the CDN are greatly expanded. How to quickly and accurately delivering the content to the users who really need it becomes a bottleneck to the development of new services. Internet CDN operators (for example, ChinaCache) and traditional telecom equipment manufacturers (such as NSN and Ericsson) have started the technology and algorithm researches on mobile network content delivery, and piloted their research results in the positioning and advertising services. China Net Center puts forward the concept of community cloud that integrates edge computing with the existing CDN platform [14]. The MEC CDN solution of ZTE can overcome the traditional CDN difficulties and satisfy the needs for rapid development of new services.

Based on the hyper-converged infrastructure (HCI) and the introduction of edge computing to CDN nodes, the ZTE MEC CDN solution (**Fig. 5**) can change the storage nodes to edge computing nodes without massive node creation, implementing all-around flexible deployment and interconnection, centralized resource cloudification, nearby service deployment, and collaboration between edge computing and cloud computing. Through standard interfaces and service procedures, the solution focuses on the improvement in content introduction, storage, and delivery.

The MEC CDN solution supports the intelligent scheduling of application perception. For a traditional CDN edge server, video storage and deletion are passive and depend on the push

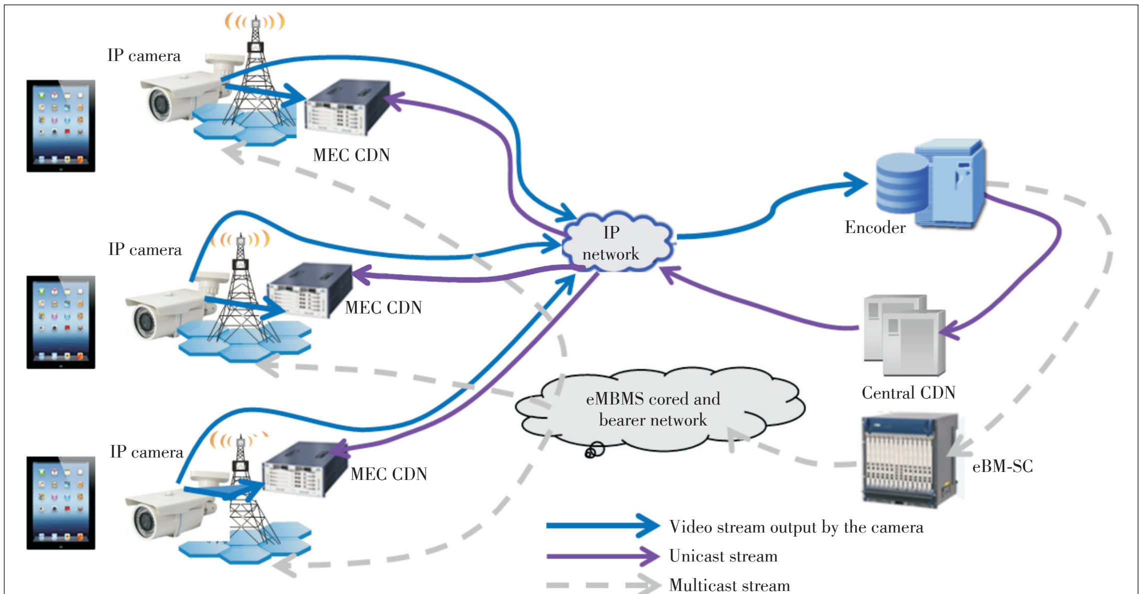


Figure 5. MEC CDN Solution.

of the core data center. An MEC CDN edge node can monitor traffic in real time. When finding an access preference or hot spot in the local area, it actively obtains data from the core data center. It can also schedule services based on applications and requirements to provide users with optimal services. An MEC CDN edge node has some network management functions, such as node self-healing and intelligent load balancing, which can effectively prevent hot-spot bottlenecks, enhance system reliability, and ensure to provide users with quality and stable services.

Users access the MEC CDN through intelligent terminals, resulting in the requirements for the mobility of the content. Due to the limitation of network resources, the movement speed of the content does not match the mobility of users in real time; namely, the content cannot be moved in real time after the location of a user changes. The MEC CDN can set up a behavior track model of users, forecast and analyze the behavior patterns and tracks of users to improve the accuracy of the prediction about hot-spot locations and the hot content at the hot-spot locations. The delivery of content in advance to the next location or the hot-spot location of users can solve the real-time matching of movement between users and the content. The prediction about content placement can reduce the delay in user's access to the content and prevent too much resource occupation during the content delivery request to implement network load balancing and avoid network congestion.

The MEC CDN solution realizes integrated multi-service bearing and supports the unified access of various terminals. The wireless resources allocated to each user over the radio link vary in real time with the changes in the type and number of terminals. In addition, the bandwidth for mobile terminals is directly affected by the signal strength for radio network access. Signal instability can lead to imperfect video playing and slow response when users watch videos. With the stream auto-adaptation technology, the MEC CDN solution dynamically detects the changes in user resources over the radio link, encapsulates transcoding in real time, and adjusts the format and code rate of the transmitted content in a timely manner to fully ensure smooth video playing and enhance user experience. For the same content, the MEC CDN system keeps only one copy of the bit rate file. The system converts the format and bit rate in real time based on the bandwidth and the type of terminals to ensure effective and continuous playing of videos, lowering the requirements for the storage space of the MEC CDN system and the bandwidth requirements for content transmission.

The MEC CDN solution realizes the combination of cloud technologies and mobile network technologies and supports distributed content delivery and DNS caching. With a top-down construction mode, the caches are gradually deployed to the base station side from the core network side, and further to the terminals based on apps. In this way, the backhaul network bandwidth can be reduced by about 35%. The system can provide better QoS and service experience and supports the content delivery of some big video services, for example, AR. When a mobile user who is downloading files or watching videos is handed over between base stations, the user identification session storage technology can ensure continuous file download and video playing without an interruption even though a handover occurs for base station access. In this way, the effect of the location change in the mobile network can be completely avoided.

The traditional TCP mechanism is used in the scenarios with a low error rate, such as wired networks and data packet networks. Assume that packet loss is caused by network congestion and data packets are retransmitted. If the network environment is adverse, the retransmission further aggravates the condition. However, a wireless network is not suitable for the traditional TCP transmission mechanism due to its high bit error rate, low transmission bandwidth, and mobility. The MEC CDN solution optimizes TCP transmission for wireless networks. With the Westwood algorithm, the MEC CDN solution determines the threshold for slow start and the size of the congestion window based on the effective bandwidth in the case of congestion (estimated through the receiving rate of ACK packets), avoiding the effect of random packet loss on the bandwidth utilization. The Westwood optimization algorithm is more suitable for the characteristics of wireless networks and can make full use of the transmission bandwidth.

5 Conclusions

Edge computing is essentially "lightweight cloud computing." The edge computing solutions of ZTE implements a unified architecture through component collocation, multi-level distributed cloud and fog deployment, on-demand network slicing, quick service chains, dynamic orchestration, fast introduction of new features, and end-to-end fast service delivery.

Edge computing is a new concept and the Edge Computing Consortium [15] has been just formed. The rapid development of edge computing in the future depends on the participation of the academia, overseas enterprises, and Internet enterprises. As more members join the Edge Computing Consortium, unified industry norms and standards must be developed in the future to jointly and orderly promote the development of edge computing.

References

[1] P. G. Lopez, A. Montresor, D. Epema, et al., "Edge-centric computing: vision and challenges," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 5, pp. 37–42, Oct. 2015. doi: 10.1145/2831347.2831354.

[2] T. Guo, "Solution to edge computing for digital transformation," *China Information World*, no. 45, p. 3, Nov. 28, 2016.

[3] Y. Ye, "Future applications and solutions to the IoT," *ZTE Technology Journal*, vol. 18, no. 2, pp. 15–17, Apr. 2012.

[4] Iot101. (2017, Jan. 12). *Edge of computing will be popular in 2017, to count the value, the opportunities and challenges* [Online]. Available: http://www.iot101.com/kpwl/2017-01-12/12755.html

[5] Y. Shi, "Software architecture analysis and application research on the fog computing platform," M.S. thesis, Zhejiang Normal University, Hangzhou, China, 2016.

[6] Iotworld. (2017, Jan. 19). *Value, opportunities, and challenges of edge computing that will be popular in 2017* [Online]. Available: http://www.iotworld.com.cn/html/News/201701/8122ed7f43e02c1f.shtml

[7] L. Shen, Q. Wu, and S. Yang, "Research on distributed storage energy saving technology on ceph," *Computer Engineering*, vol. 41, no. 8, pp. 13–17, Aug. 2015.

[8] S.-G. Wang, W.-X. Xu, Wu Dan, Z.-B. Pang, and J. Xia, "A fast RDMA offload method for unreliable interconnection networks," *Journal of Hunan University (Natural Sciences)*, vol. 42, no. 8, pp. 100–107, Aug. 2015.

[9] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief. (2017, Jan. 11). *Mobile edge computing: survey and research outlook* [Online]. Available: https://arxiv.org/pdf/1701.01090.pdf

[10] Y. Wu and L. Yang, "Introduction and value exploration on NFV PaaS," *ZTE Technologies (Chinese Version)*, no. 2, pp. 20–22, Feb. 2016.

[11] D. Liu. (2016, Apr. 20). *ZTE's NFV path: "three trends" based network transition from IaaS to PaaS* [Online], Available: http://www.c114.net/news/127/a950179.html

[12] Y. Lu. (2016, May 12). *ZTE 5G MEC test on the way* [Online], Available: http://www.cww.net.cn/zhuanti/html/2016/5/12/20165121547158041.htm

[13] J. Liang, G. Chen, and Y. Zhuang, *Content Delivery Network (CDN) Key Technologies, Architecture, and Applications*, Beijing, China: Posts & Telecom Press, 2013.

[14] W. Shu, "ChinaNetCenter: community cloud based on the convergence of edge computing and CDN," *Communications World*, no. 33, pp. 37–38, Dec. 2016.

[15] H. Huang, "Influence of the establishment of the edge computing consortium," *Communications World*, no. 33, pp. 52–53, Dec. 2016.

Manuscript received: 2017-02-10

Biographies

TU Yaofeng (tu.yaofeng@zte.com.cn) received his master's degree from Nanjing University of Aeronautics and Astronautics, China. He is a deputy director of the Cloud Computing & IT Institute of ZTE Corporation. His research interests include telecommunication services, cloud computing products, big data, and artificial intelligence.

Dong Zhenjiang (dong.zhenjiang@zte.com.cn) received his master's degree from Harbin Institute of Technology, China. He is a deputy director of the Cloud Computing & IT Institute of ZTE Corporation. His research interests include cloud computing, artificial intelligence and media processing, and mobile Internet and security.

YANG Hongzhang (yang.hongzhang@zte.com.cn) received his master's degree in computing technology from Chinese Academy of Sciences, China. He is a technology pre-research engineer with the Cloud Computing & IT Institute of ZTE Corporation. His research interests include cloud computing, big data, and distributed storage system.