6G Vision: An AI-Driven Decentralized Network and Service Architecture

Xiuquan Qiao
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications

Yakun Huang
Beijing University of Posts and Telecommunications

Schahram Dustdar
Vienna University of Technology

Junliang Chen
Beijing University of Posts and Telecommunications

***Abstract*—Recently, following the rapid commercial deployment of 5G networks, next-generation mobile communication technology (6G) has been attracting increasing attention from global researchers and engineers. 6G is envisioned as a distributed, decentralized, and intelligent innovative network. However, existing application provisioning is still based on a centralized service architecture, ubiquitous edge computing, and decentralized AI technologies have not been fully exploited. In this article, we analyze the problems faced by existing centralized service provisioning architecture, and propose design principles for a decentralized network and service architecture for a future 6G network. Finally, we discuss several open research problems to inspire readers to address these.**

DUE TO THE large number of commercial applications of 5G networks worldwide, potential 6G technologies are attracting attention from both academia and industry. Although 5G has achieved significant improvements in terms of communication performance, it remains difficult to meet demand for more intelligent communication in terms of information speed, multidomain coverage, artificial intelligence (AI), and security.[1] Recently, several governments have launched 6G projects to explore the requirements and key technologies of the next-generation mobile communication network. However, existing visions and discussions of 6G mainly focus on innovative wireless communication technologies, mobile

Digital Object Identifier 10.1109/MIC.2020.2987738
Date of current version 9 September 2020.

edge computing (MEC) and AI,[2] and there is a lack of deep and innovative insights into networking and service provisioning mechanisms. It is therefore necessary to create a blueprint for a disruptive service provisioning mechanism in future 6G networks.

Following the continuous evolution of 5G networks, 6G is envisioned as an ultrabroadband, ultra-low-delay, full-dimensional coverage (terrestrial, aerial, space, and maritime domains) ubiquitous intelligent network with native AI and security. 6G will seamlessly integrate communication, computing, control, caching, sensing, positioning, and imaging features to support various Internet of Everything (IoE) applications. In contrast to 5G, 6G will evolve from "human-machine-thing" interactions to "human-machine-thing-genie" interaction,[3] and will become a highly autonomous and intelligent ecosystem. Several innovative applications will become reality, such as holographic communications, brainwave-machine interaction applications, tactile mixed reality (MR) experiences (including vision, hearing, smell, taste, and touch), and high-precision manufacturing.[4] With the continuous maturity of AI technology and the rapid reduction in related hardware costs, increasing numbers of devices will have native AI functions, such as smartphones, AR/VR glasses, smartwatches, headsets, TVs, loudspeakers, and vehicle-mounted devices. Based on the user's movements, these ubiquitous devices will dynamically and autonomously collaborate with each other to achieve better user experience. 6G will be a highly dynamic, autonomous, decentralized, and intelligent network in which network nodes will collaborate autonomously and dynamically, user data will be stored in the network in a decentralized way, and services will migrate on demand. This new ubiquitous, decentralized, AI-driven flat 6G network needs a corresponding new decentralized service provisioning mechanism.

However, although MEC has pushed computing closer to the user, and the device-to-device (D2D) model has enabled nearby mobile devices to communicate with each other directly, the 5G network remains a centralized network and service provisioning architecture in terms of the data storage and access, service running mechanism, and application protocols used. The service provisioning mechanism for the existing 5G network has not changed a great deal compared with the 4G network. We therefore need to design a novel service provisioning mechanism to meet the tremendous shift toward decentralization in the future 6G network.

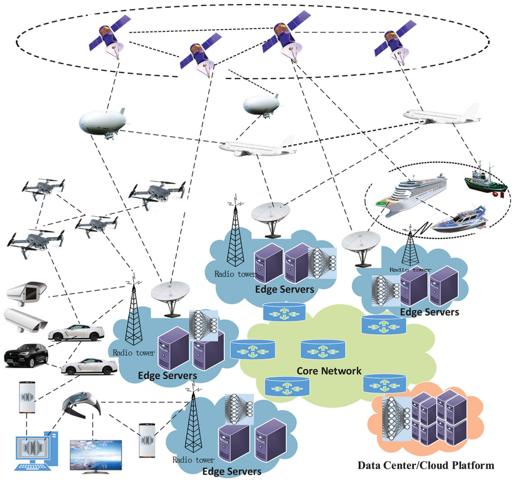

Figure 1. Vision for 6G: Integrated, ubiquitous, intelligent, and decentralized.

In this article, we first explain the decentralized trends of the 6G network and its new characteristics. We then analyze and discuss the problems faced by existing centralized service provisioning. Finally, we propose some design principles for a future decentralized 6G service provisioning mechanism, and discuss open research issues related to this new decentralized computing paradigm.

6G VISION AND ITS NEW CHARACTERISTICS

Although 6G is not yet the subject of a global consensus, some potential new characteristics and trends have been widely discussed. In this section, we present a comprehensive vision of the future 6G network from multiple perspectives, as shown in Figure 1.

1) *Network coverage view:* With the expansion of human activities, the existing closed and vertical dedicated networks and terminals cannot meet the demand for ubiquitous mobile communication anytime and anywhere. Unlike the previous 1G to 5G networks, 6G will extend mobile communication

coverage in an unprecedented way from terrestrial areas to the aerial, space, and maritime domains. A ubiquitous, integrated, multidimensional, and full-coverage mobile communication network will be available anywhere in the 6G era. Everything (including real-world objects and digital objects in virtual worlds) will be able to be connected with everything else, and a new IoE distributed ecosystem will be established based on this all-round connectivity.

2) *Capability convergence view:* With the enhancement of terminal capabilities, the large-scale deployment of MEC infrastructures[5] and the widespread applications of IoE, communication is no longer the only goal of the 6G network. The convergence of communication, computing, control, storage, and sensing capabilities will be the new trend in the 6G network. Based on these capabilities, increasing numbers of terminals and network nodes will become intelligent, autonomous information processing entities and act as both information producers and consumers.
3) *Interaction space view:* Based on the characteristics of eMBB, mMTC, and uRLLC, 5G has begun to support "human-machine-thing" interactions, which bridge the domains of cyber space, physical space, and society. 6G will further deepen and expand interaction spaces. Following the advances in wireless brain–computer interaction (BCI) technologies, the use of consciousness-based communication and control will create some new application scenarios. For example, brain–computer interfaces may be used to interact with ambient smart devices such as XR glasses, TV sets, or loudspeakers. 6G will also evolve from current "human-machine-thing" interactions to "human-machine-thing-consciousness" interactions. The real world and virtual world will be perfectly integrated, and the era of augmented reality (AR)/MR is imminent,[6] in which physical and digital objects coexist and interact in real time (i.e., dual worlds).
4) *AI view:* In the initial design phase of the 5G network, AI technologies were not sufficiently mature to act as an enabling technology. However, following rapid advances in big data, cloud computing, neural network, and dedicated chip technologies in recent years, AI technologies have begun to be employed in 5G network management, smart mobile phones, and various applications in a patchwork way. AI is considered the most innovative enabling technique for 6G, and will be an innate feature of the network from the application layer to the physical layer. In the 6G era, end devices with various AI capabilities will seamlessly collaborate with a variety of edge and cloud resources.[7] With the maturity of AI technology and the reduction in AI hardware costs, the number of smart end devices in use in daily life will constantly increase. Decentralized and collaborative AI services among distributed end devices and network nodes will also become a trend in 6G.
5) *Network architecture view:* As mobile communication networks have been updated from 1G to 5G over the decades, they have gradually evolved from a closed dedicated network to an open converged network based on general IT technologies. Network architectures are becoming increasingly flat, and the original customized hardware appliances used for each network function have been replaced by general IT devices and software platforms. This is particularly true for the 5G network, in which software-defined networks, network function virtualization, and network slicing technologies have been fully employed. To a certain extent, network carriers can utilize different software and processes in a flexible way on top of standard high-volume servers, switches, and storage devices, and can customize different virtual networks in order to meet demand arising from differentiated application scenarios such as high bandwidth, low latency, or massive numbers of connections. In addition, MEC and D2D communication technologies promote the migration of computing and service processing capability from the cloud platform to the network edges.[8] With enhancements to smart user and network equipment, edge or fog computing will become as important as cloud computing. Increasing numbers of local communication clusters will be formed dynamically and

autonomously, and applications will be processed both directly and locally. The network edge will be highly decentralized, and will take on some of the functions of the core network and cloud platform. In this model, the network edge is no longer simply an access network, but comprises a large number of ubiquitous, autonomous local networks that can integrate communication, computing, control, storage, and sensing capabilities. The network edges and core networks will have a more peer-to-peer structure, and in general, the network architecture will be much flatter and more flexible.

6) *Application architecture view:* Based on the above analysis, it can be seen that 6G will be a ubiquitous, distributed, decentralized, and intelligent innovative network. The existing application provisioning architecture mainly adopts B/S or C/S architectures, which were originally designed for centralized network. Clients often interact with centralized, specific application servers, and database servers in order to deal with user requests. In contrast, 6G will become more decentralized due to the long-term evolution of the 5G network. The application provisioning architecture of 6G will also therefore change significantly to cater to this shift. In the future 6G network, peer-to-peer and ad hoc networks will become more pervasive and popular, and current cloud-based serverless application provisioning architecture will gradually evolve toward a decentralized peer-to-peer application provisioning architecture. User data will be stored on a decentralized peer-to-peer network, and business processing logic will be divided into stateless and independent fine-grained services that can be migrated and run on any network node on demand.

EXISTING CENTRALIZED APPLICATION PROVISIONING MECHANISM AND RELATED ISSUES

Since the aforementioned vision for 6G and its new characteristics differ markedly from the existing 5G and 4G mobile networks, it is necessary to analyze and discuss the problems faced by current centralized service provisioning. After 40 years of development, the existing centralized application provisioning architecture has gradually become unsuitable for the application development needs of 6G.

1) *Limitations on B/S or C/S application architecture:* Most existing applications employ B/S or C/S application provisioning architecture, which was originally designed for the era of thin clients and powerful servers. An application is provided by a collaboration between user devices and edge/cloud servers. In a centralized architecture, applications are highly dependent on dedicated cloud servers, and information storage and business logic are all provided by servers. This architecture gives rise to the high computing, storage, and bandwidth costs on the server side. With the emergence of MEC in 5G, some of these application functions can now be offloaded to edge servers, and a "terminal+edge+cloud" collaboration computing architecture is being developed. However, 5G applications are only beginning to allow for distributed computing, let alone a decentralized computing model. Following significant advances in hardware and software, the capability of 6G terminals will be further improved, and some tasks will be processed by the local user terminal or by collaborating with ambient devices or edge/cloud servers. It is therefore necessary to explore a novel application architecture to support this ubiquitous decentralized computing paradigm.
2) *Disadvantages of the centralized data model:* In the existing centralized application architecture, data are generally stored in specific cloud servers or terminal devices, and are cached on edge servers or CDN networks. Data storage and access are all controlled by a centralized authority such as Yahoo, Facebook, or YouTube. This centralized data model results in some potential problems such as censorship, privacy, data leakage, and data control rights. For example, if the central point is hacked, the entire user database is at risk. In addition, the trust problem of the centralized authority is often challenged. In fact, some Internet service providers use the data for their own benefit, such

as selling it to advertising companies, meaning that the privacy and security of user data are not well protected.

3) *End-to-end application protocols:* Due to the use of a centralized data storage and service operation mechanism, most of the existing application protocols are based on an end-to-end communication model rather than a peer-to-peer model, and client requests need to be routed to dedicated application servers to be processed. Existing application protocols (such as HTTP) were originally designed for B/S or C/S application architectures, and are not suitable for this new dynamic and opportunistic form of connectivity and the ubiquitous edge and decentralized computing paradigm of the 6G network. In the era of 6G, application protocols will enable data access and service coordination on a peer-to-peer basis over a ubiquitous distributed network.
4) *Tight coupling of user data with specific applications:* Following the rapid development of the mobile Internet, increasing numbers of people rely on service provisioning by a few Internet giants such as Yahoo, Google, Facebook, Twitter, and WeChat, and the centralization of information has become more pronounced, with services and content being gradually aggregated by a few Internet oligarchs. This model of centralized information organization creates many information islands, and in this paradigm, users have no right to control their data. User data are tightly coupled with specific apps, and data utilization across different apps is often restricted due to commercial competition. These centralized information islands have gradually come to hinder the free dissemination of information.
5) *Shortcomings of centralized AI:* In recent years, due to the development of powerful cloud computing capability and big data, AI has become increasingly widespread. However, existing AI is mainly organized using a centralized application model. More specifically, massive training datasets are very valuable assets to enterprises. Training datasets and the creation and training of models are also controlled by a small number of large organizations, which increases the gap between large companies with access to large, labeled datasets, and smaller companies. At the same time, the centralization of model training requires the transmission of data from end devices to cloud servers, often resulting in high transmission and computing costs and giving rise to user privacy protection issues. Furthermore, current AI models are always deployed on either cloud/edge servers or end devices using a centralized operation model, without allowing for the efficient utilization of resources such as ubiquitous distributed network nodes.

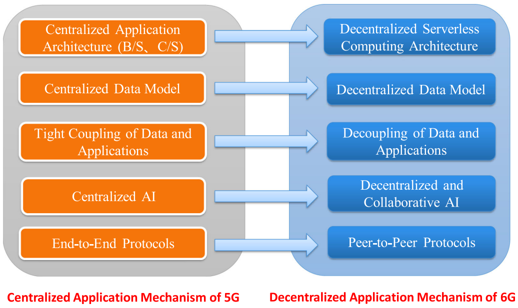

Figure 2. Evolution of 6G decentralized application provisioning.

DECENTRALIZED APPLICATION PROVISIONING ARCHITECTURE FOR 6G

Decentralized Application Provisioning Mechanism

Based on the above analysis, we expect that the application provisioning mechanism in 6G will change significantly from the existing centralized application mechanism, as shown in Figure 2. Some design principles for a future decentralized 6G application provisioning mechanism are presented below.

1) *Decentralized serverless computing architecture:* In the future 6G network, the communication, computing, and storage capabilities of network nodes will be greatly enhanced. The traditional client–server boundary will be eliminated, and each network node (including various terminals, base stations, gateways, routers, servers, etc.) will act not only as an information publisher but also an information consumer. In 6G, the decentralization of

the network infrastructure will be realized, and the whole network will become a service running environment. A microkernel-based distributed operating system will become popular, and this will be adaptively deployed on various types of hardware, including smartphones, AR/VR glasses, smart displays, wearable devices, in-car entertainment systems, and other IoT devices. The service environment will gradually expand from the existing cloud infrastructure to the network edge and ubiquitous end devices. The overall business logic will be composed of multiple fine-grained micro services; these will not be deployed on dedicated servers, and will be able to migrate to any network node on demand. Front-end client applications will resolve the application description file and invoke the related service components directly.

2) *Decentralized data model:* With the large-scale deployment of edge/fog computing, it is now possible to establish a ubiquitous decentralized storage infrastructure to address the problems faced by the existing centralized cloud storage model. Compared with the centralized data model, data will no longer be stored on specific servers, but will be distributed over a peer-to-peer network. This decentralized data model promises even greater advantages, such as efficient scalability, reliability, privacy, and data immutability. Since all the data are distributed among the different network nodes, decentralized data networks are better able to withstand massive user requests distributed between the nodes, since the pressure of these requests no longer falls on a few computers but on the network as a whole. This scheme can also deal with DDoS attacks more effectively. In addition, a decentralized data model can reduce dependence on the infrastructure of specific Internet giants, and will facilitate the disintermediation of the mobile Internet.

3) *Decoupling of data and applications:* In order to return data control rights to the users themselves, it is necessary to decouple user data from specific silo applications. In the future 6G network, user-generated data such as videos, social media posts, health data and tracking information, etc., will be completely controlled by the users themselves. These data will be stored in a decentralized P2P network in which users have the right to authorize certain applications to manipulate their data, and decide which users to share these data with. This new mechanism will facilitate information sharing and dissemination among different apps. For example, a user's profile information can be shared by different apps, thus avoiding the need for each app system to save a copy of the user's information. This scheme can also avoid data leakage by third-party application providers.

4) *Decentralized and collaborative AI:* In the 6G era, each network node can store and process data and can autonomously communicate and seamlessly interact with other ambient devices. Following the development of ubiquitous computing infrastructure, existing centralized AI will gradually evolve to a decentralized and collaborative model. In contrast to traditional centralized AI, in which all data samples are uploaded to dedicated cloud servers, the decentralized approach will train a model across multiple decentralized edge devices or servers with local data samples; this will be done without sharing data, and instead parameters will be simply exchanged between these local models at a certain frequency to generate a global model. This approach can efficiently avoid the need for transmission and centralized storage of training data, and can also address several critical issues such as data privacy, data security, data access rights, and heterogeneity of data. In addition, aided by advances of lightweight model technology, AI models can be deployed on any devices, from mobile phones to massive numbers of IoT devices. AI will be able to run, train, and even make decisions on local devices in this type of decentralized network. The autonomous collaboration of multiple network nodes is controlled by a distributed group of intelligent agents, which will be able to solve complex planning and decision making problems.

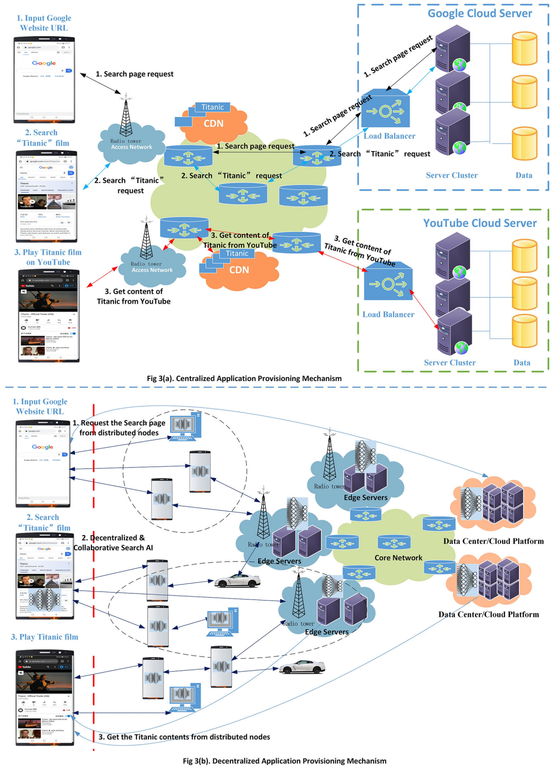

Figure 3. Differences between existing centralized and future decentralized application provisioning approaches.

Comparison of Centralized and Decentralized Application Solutions

In Figure 3, we use a mobile search application as an example to illustrate the significant changes in existing centralized and future decentralized application provisioning mechanisms.

Figure 3(a) shows the existing centralized Google mobile search application mechanism. It can be seen that the application is mainly processed via a collaboration between the mobile browser and dedicated cloud application servers, and the network is only responsible for the transmission of information. When the user inputs the Google URL, the mobile browser will query the IP address corresponding to the URL using the DNS service and sends the Web page request to Google's dedicated cloud server, which returns the search page. After the user inputs some search content ("Titanic" in this example), the browser will send this search HTTP request to the Google search server, which generates the results pages. The user may then click a link to to play "Titanic" on the YouTube website. The browser will obtain the content from the YouTube cloud servers or nearby CDN networks.

In contrast to this existing centralized mechanism, a decentralized mechanism will be very different in terms of the data storage, server architecture, and communication protocols used, as shown in Figure 3(b). There are no dedicated cloud application and database servers, and the whole network acts as a decentralized communication, computing, and storage infrastructure. The browser will get the search Web page from a distributed file system in a peer-to-peer way, using a distributed hash table. Segments of Web pages may be located on nearby mobile phones, PCs, edge, or cloud server nodes. After the user inputs the search content "Titanic," the AI-enhanced browser can process and analyze the search input by itself, using a lightweight AI model for natural language processing, and can collaborate with ambient AI-enhanced devices to generate the search results pages. When the user clicks the YouTube link in the search results page, the media player will fetch content segments from a distributed peer-to-peer network.

CONCLUSION AND OPEN ISSUES FOR FUTURE RESEARCH

Decentralization has become a likely trend for a future 6G network. In this article, we mainly focus on the potential disruptive changes to application provisioning mechanisms in the 6G era. By analyzing the issues faced by the existing centralized infrastructure, we propose some insights for a decentralized application provisioning mechanism for the future 6G network.

However, until now, there have been no comprehensive discussions of 6G from this perspective. Several issues are still open, and we describe these here to provide readers with inspiration to address these issues.

1) *Decentralized operating system for ubiquitous computing:* In view of the IoE application scenarios envisaged for 6G, it will be necessary to develop a decentralized operating system for a dynamic, autonomous, collaborative network, which can efficiently enable peer-to-peer communication, decentralized data storage and access, on-demand service migration and deployment, and flexible adaptation of

heterogenous devices such as servers, mobile phones, TV sets, vehicle-mounted systems, and other IoT devices.

2) *Collective decision making by decentralized AI:* Decentralized AI has become one of the most promising trends for the next phase of AI. With the aid of D2D and MEC, decentralized and collaborative AI services among distributed network nodes will become an important enabling technology for 6G. The issues of how to integrate these scattered AI capabilities over distributed nodes and find the optimal combination of services to provide the best experiences for users are worthy of in-depth study and exploration. This will involve coordination between and decisions by multiple intelligent agents, and thus constitutes a collective decision-making issue.
3) *Disruptive influences of the decentralization network and service model:* A decentralization model will bring about disruptive impacts on existing application provisioning mechanisms in terms of business models, products, services, and ecosystem roles. It will inevitably weaken the authority of central entities and will affect the commercial interests of the existing Internet giants. At the same time, it will also create impacts on the infrastructure governance of telecom network operators. The problem of how to effectively activate and coordinate multiple stakeholders (individual users and other enterprises) to participate in the provisioning of network resources in a future 6G ecosystem is a newly emerging issue. It is therefore necessary to explore the strong potential influence on the operation of network infrastructure.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Project under Grant 2018YFE0205503.

REFERENCES

1. Z. Zhang et al., “6G wireless networks: Vision, requirements, architecture, and key technologies,” *IEEE Vehicular Technol. Mag.*, vol. 24, no 40, pp. 28–41, Sep. 2019.
2. K. B. Letaief *et al.*, “The roadmap to 6G: AI empowered wireless networks,” *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
3. P. Zhang *et al.*, “Technology prospect of 6G mobile communications,” *J. Commun.*, vol. 40, no. 1, pp. 141–148, 2019.
4. B. Zhang *et al.*, “6G technologies: key drivers, core requirements, system architectures, and enabling technologies,” *IEEE Vehicular Technol. Mag.*, vol. 14, no 3, pp.18–27, Sep. 2019.
5. M. Gusev and S. Dustdar, “Going back to the roots—The evolution of edge computing, an IoT perspective,” *IEEE Internet Comput.*, vol. 22, no. 2, pp. 5–15, Mar./Apr. 2018.
6. X. Q. Qiao *et al.*, “Web AR: A promising future for mobile augmented reality—State of the art, challenges, and insights,” *Proc. IEEE*, vol. 107, no. 4, pp. 651–666, Apr. 2019.
7. X. Wang and Y. Han, “In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning,” *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep.–Oct. 2019.
8. X. Q. Qiao *et al.*, “A new era for web AR with mobile edge computing,” *IEEE Internet Comput.*, vol. 22, no. 4, pp. 46–55, Jul./Aug. 2018.

Xiuquan Qiao is a full professor with State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. His main research interests focus on 5G/6G networks, augmented reality, edge computing, and services computing. Contact him at qiaoxq@bupt.edu.cn.

Yakun Huang is a Ph.D. candidate at Beijing University of Posts and Telecommunications. Contact him at hyk_it@foxmail.com.

Schahram Dustdar is a full professor of computer science (Informatics) with a focus on Internet Technologies heading the Distributed Systems Group, TU Wien, Austria. He is a member of the Academia Europaea: The Academy of Europe. He is a Fellow of the IEEE. Contact him at dustdar@dsg.tuwien.ac.at.

Junliang Chen is a professor of Beijing University of Posts and Telecommunications, Beijing, China. He is a member of both the Chinese Academy of Sciences and the Chinese Academy of Engineering. He is a Senior Member of the IEEE. Contact him at chjl@bupt.edu.cn.