Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

Future Internet: trends and challenges*

Jiao ZHANG†‡1,2,3, Tao HUANG1,2, Shuo WANG1,2, Yun-jie LIU1,2
1State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2Purple Mountain Laboratories, Nanjing 211100, China
3CETC Key Laboratory of Aerospace Information Applications, Shijiazhuang 050081, China
†E-mail: jiaozhang@bupt.edu.cn

Received July 23, 2018; Revision accepted Dec. 14, 2018; Crosschecked Aug. 15, 2019; Published online Sept. 5, 2019

Abstract: Traditional networks face many challenges due to the diversity of applications, such as cloud computing, Internet of Things, and the industrial Internet. Future Internet needs to address these challenges to improve network scalability, security, mobility, and quality of service. In this work, we survey the recently proposed architectures and the emerging technologies that meet these new demands. Some cases for these architectures and technologies are also presented. We propose an integrated framework called the service customized network which combines the strength of current architectures, and discuss some of the open challenges and opportunities for future Internet. We hope that this work can help readers quickly understand the problems and challenges in the current research and serves as a guide and motivation for future network research.

Key words: Future Internet; Network architecture; Service customized network
https://doi.org/10.1631/FITEE.1800445 **CLC number:** TP393

1 Introduction

Rapid advancement in computing technologies has enabled a great variety of applications and Internet services, ranging from cloud computing applications, smart-home applications, and industrial monitoring to medical health care. Specifically, mobile networks have undergone a tremendous growth, and mobile applications such as social networking, gaming, and augmented reality further contribute to the rapid growth of network traffic (Taleb et al., 2017). According to the results, network data traffic will grow by eight times from 2015 to 2020, and the number of machine connections will reach one billion by 2020 due to the development of the Internet of Things (IoT) and machine-type communications (MTC). Furthermore, new strategies and plans such as Internet Plus, Made in China 2025, and Industry 4.0, require the integration of the Internet with the economy, which further increases the demands and challenges for network technologies.

The traditional network has many problems and technical issues concerning performance, scalability, flexibility, and security, which make it difficult to meet the demands of applications and users. There are two main reasons for limited network development. On the one hand, there has been no breakthrough in the contemporary Internet architectures since the 1970s, and the scalability and flexibility of networks are limited by the "thin waist" Internet Protocol (IP) model, which is an end-to-end channel between two endpoints identified by IP addresses (Jacobson et al., 2009). On the other hand, because

---

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 61872401), Program of the CETC Key Laboratory of Aerospace Information Applications, China (No. SXX18629T022), and the Industrial Internet Innovation and Development Project 2018 from the Ministry of Industry and Information Technology of China

ORCID: Jiao ZHANG, http://orcid.org/0000-0001-5614-3420

the traditional equipment is not programmable and hard to upgrade, existing network infrastructures are struggling to support new network innovations, which makes it difficult to test new network architectures and develop new network protocols.

Therefore, the research on future Internet has become a hot and important research topic worldwide. Future Internet can be explained as new architectural designs with fundamental innovations for a next-generation Internet (Fisher, 2014). Many countries and organizations have proposed a variety of new network architectures such as the software-defined network (SDN), content-centric network (CCN), and expressive Internet architecture (XIA) to provide advanced network innovations, and proposed technologies such as network function virtualization (NFV) and edge computing to support customized demands. Some of these technologies have already been used in the data centers of Google, Amazon, Microsoft, etc., and in the telecommunications companies like AT&T, China Unicom, China Mobile, and China Telecom.

Table 1 shows the main differences between our survey and prior work. Prior surveys covered different aspects of the future Internet technologies. In particular, Bannour et al. (2018) presented a survey of the distributed control plane of SDN to increase the scalability. Li and Chen (2015) introduced the relationship between SDN and NFV. Taleb et al. (2017) provided an overview of edge computing for 5th generation (5G) wireless systems and summarized standardization activities. These advanced technologies were summarized separately in the literature. To the best of our knowledge, none of these surveys have focused on the various technologies and challenges for future Internet.

2 Increasing demands of services

The past few decades have witnessed a rapid development of the Internet services, ranging from cloud computing, IoT, and mobile applications to industry services. In this section, we summarize the new demands of these services, which are also the design goals for the future Internet.

2.1 Increasing performance

Latency and bandwidth are the most important performance metrics for networks. For example, the services provided by data centers can be classified into two categories. The first type provides online services including web search (Google search) and social networks (Facebook). This type of service is very sensitive to network latency. For example, Amazon has shown that a 100-ms network latency will decrease sales by 1%, and Google found that an extra 0.5 s of search latency will drop the traffic by 20% (Shalom, 2010). The other type offers resources including cloud computing (Microsoft Azure), big data (Hadoop), and online storage (Amazon AWS), which are very sensitive to network bandwidth. For instance, recent work has shown that the intermediate data transmission accounts for more than 50% of applications' completion time (Chowdhury et al., 2011).

In addition, as the number of users and the amount of data traffic increase, the wide area network and mobile network face challenges in latency and bandwidth. Specifically, the 5G network is expected to support data rates up to 10 Gb/s and latency at around 1 ms (Akyildiz et al., 2016). Thus, how to decrease latency and increase bandwidth becomes the basic requirements for the future Internet.

2.2 Programmable data plane

The traditional network is designed following a distributed pattern, and most of the equipment is dedicated hardware that is not programmable. This results in an inflexible data plane. For example, it is difficult to implement new protocols in traditional networks. This is because the architecture of the traditional network is not flexible enough and cannot support significant modifications. The design goal for IP is to support end-to-end communications. This means that new features and protocols can be deployed only in a "patched" fashion.

Furthermore, because the network nodes like switches and routers are not programmable, there is almost no practical way to experiment with new network protocols (McKeown et al., 2008). Hence, as more and more new protocols are developed, network nodes become very complex with a very low scalability. Therefore, to support a flexible data plane, the future Internet requires innovation in the network architecture and programmable network devices.

Table 1 Main differences between our survey and prior work

Survey	Covered aspects	Showing practical user case	Jointly considering new technologies
Bannour et al. (2018)	SDN control plane	No	No
Li and Chen (2015)	NFV, service chain	No	Partial (SDN)
Taleb et al. (2017)	Edge computing, 5G	No	Partial (SDN, NFV)
Our survey	Future Internet technologies (SDN, NFV, edge computing, CCN, etc.)	Yes	Yes

SDN: software-defined network; NFV: network function virtualization; CCN: content-centric network

2.3 Flexible management

Traditional networks rely on purpose-built network devices to provide diverse network services, like firewalls, load balancing, intrusion detection systems, and cache nodes. Network providers have to manage a great number of devices with different operating systems and software. This creates large overhead for the management of the network. On the one hand, it is very difficult to upgrade the software and functions of these devices (Li and Chen, 2015), and the network providers have to manually configure the devices because there is no centralized controller. On the other hand, one class of network services usually relys on more than one type of network device and function. Thus, when a customer requires a new service, the network providers may take a few days or even a few weeks to deploy the new service. As the number of services increases, managing these services will be a big problem. Therefore, the future Internet requires a flexible management architecture to reduce the management overhead.

2.4 Customized demand

In recent years, the number of over-the-top (OTT) (https://en.wikipedia.org/wiki/Over-the-top_media_services) applications including television (Apple TV), voice calling (Skype), and messaging (WeChat) has rapidly increased. OTT bypasses the telecommunication providers and allows content providers to directly sell audio, video, and other services to the consumer over the Internet. As a result, the telecommunication provider loses the control of the content and becomes a simple data transmission channel. Thus, the telecommunication provider has to change the pricing model for these services to one based on data usage, and introduces new economic models for better services.

Because different OTT applications have different requirements and demands for the network, the telecommunication provider can provide differentiated services to more efficiently use the network resources. Actually, service providers like Google have attempted to build their own network infrastructures to more flexibly use network resources to meet their business needs (Jain et al., 2013). Therefore, the future Internet should meet customized demands and provide high-quality network services for users who are willing to pay a high price.

2.5 Security, privacy, and trust

Network security, privacy, and trust have received much attention in recent years with the increasing data traffic in the network. Some authentications or logging messages include sensitive data and should be scrupulously handled. As the IoT, SDN, and NFV tend to be themes in the future Internet, the security and privacy issues face a greater challenge. For example, the massive number of IoT devices that manage sensitive information may be attacked. Furthermore, SDN and NFV not only make the network more flexible, but also put the network at more risk because most of the network nodes are programmable and changeable. Hence, the future Internet should handle the new threats and guarantee data safety.

3 Efforts on future Internet architectures and key technologies

The research related to the future Internet requires architectural design and various key technical innovations. In this section, we will cover a wide area of technologies that may address one or several of the above challenges.

3.1 Network architectures

To support network innovations, many new network architectures have been proposed, and we introduce only some very promising work in this subsection.

3.1.1 Software-defined networking

The idea of SDN was proposed by the Clean Slate Research Group of Stanford University in the Global Environment for Network Innovations (GENI) Project (www.geni.net). Experimental development of OpenFlow-based SDN was conducted in campus networks and backbone-sized networks (McKeown et al., 2008).

SDN separates the data plane and the control plane, and leverages a centralized controller to provide network programmability. A typical architecture of SDN is shown in Fig. 1. There are three layers:

1. Data plane

The SDN devices are programmable and support standard programming interfaces and protocols such as OpenFlow (McKeown et al., 2008). They do not have to support a large number of protocols, and accept only the instructions from the controller, which fully simplifies the data forwarding plane.

2. Control plane

The SDN control plane consists of a centralized controller and is responsible for monitoring global information and realizing network intelligence. Thus, the SDN controller greatly simplifies the control and operation of the network.

3. Application layer

An array of applications can be implemented in this layer to provide different network services such as quality of service (QoS) and security. The applications in this layer communicate with the control plane using northbound protocols and interfaces.

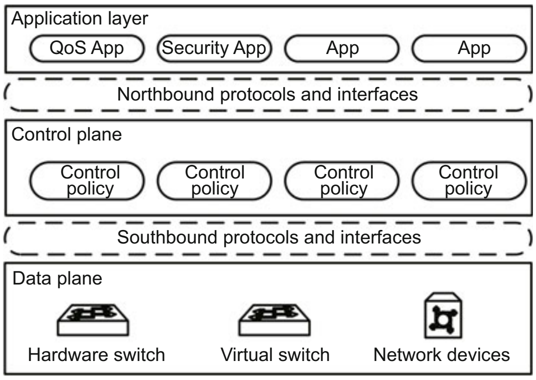

Fig. 1 Software-defined networking architecture

By providing programmable interfaces, SDN enables flexible network management and is a promising way to support new protocols and new network functions. However, SDN also faces some problems in practical implementations. The main problem is that the central controller becomes a bottleneck in the whole network. We believe this problem can be solved by layering the control plane and deploying a simple local controller on each switch. The design of hierarchical architectures and corresponding protocols between layers will become the new challenges.

3.1.2 Information centric networking

Currently, popular contents are transmitted repeatedly on the Internet, wasting resources and reducing the QoS. To reduce information redundancy, a kind of clean-slate network architecture called information-centric networking (ICN) has been proposed to change the traditional end-to-end communication model. Specifically, we choose named-data networking (NDN) (Zhang et al., 2014) as a typical representative to introduce ICN.

The hourglass model of the IP stack is retained in NDN, but the waist layer is changed with a hierarchical content-naming structure. This naming structure is similar to the uniform resource locator (URL) structure. Meanwhile, all the routing nodes of NDN are equipped with caches or repositories. Therefore, routing, forwarding, and caching are all implemented based on flexible naming. In addition, NDN integrates security in the data by cryptographically signing every data packet to ensure integrity and authenticity.

The main advantage of ICN is that it can provide efficient resource and data sharing, secure content delivery, and mobility support. Although the implementation of ICN faces non-technical problems, the roadmap to ICN encounters conflicts with current network architectures and lacks smooth evolution.

3.1.3 MobilityFirst

The design of existing IP networks is based on end-to-end connections and cannot support the increasing demand for mobile devices. Thus, to better meet the needs of mobility, the MobilityFirst Project (http://mobilityfirst.winlab.rutgers.edu) was started in 2010 with funding from the National Science Foundation's Future Internet Architecture (FIA) Program.

The major design goals of MobilityFirst are: (1) mobility as the norm with a dynamic host and network mobility at scale; (2) robustness with respect to intrinsic properties of wireless media; (3) trustworthiness in the form of enhanced security and privacy for both mobile networks and wired infrastructure; (4) usability features such as supporting context-aware pervasive mobile services, evolvable network services, manageability, and economic viability.

3.1.4 Knowledge-defined networking.

Knowledge-defined networking (KDN) (Mestres et al., 2017) has been proposed to implement a closed control loop that provides automation, recommendation, optimization, validation, and estimation. Fig. 2 shows the basic steps of the KDN control loop, which contains mainly the following elements:

1. The analytics platform aims to monitor the data plane and collect network information.

2. Machine learning algorithms use the data of the analytics platform to learn from the network and generate knowledge.

3. A northbound controller application programming interface (API) is used to automatically improve the network performance according to the discovered knowledge.

Besides KDN, machine learning has been used by some proposals to improve the network performance. For example, Pensieve (Mao et al., 2017) leverages reinforcement learning to predict the bitrate of an online video stream. In Valadarsky et al. (2017), the Internet routing algorithm was optimized using machine learning.

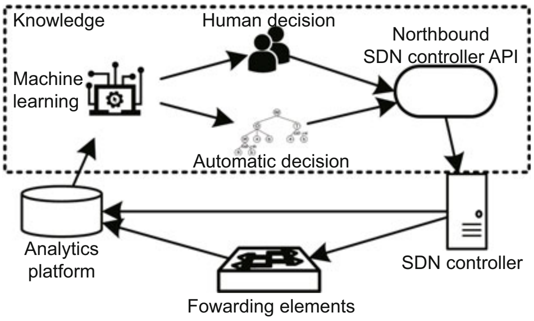

Fig. 2 Knowledge-defined networking operational loop

The growth of KDN benefits from the development of machine learning technology. However, the study of KDN is merely at the first stage and demands huge research effort. Some machine learning models and algorithms that are specially designed for network management scenarios are expected.

3.1.5 Other architectures

We omit the details about the following new network architectures:

1. Nebula (http://nebula-fia.org)

Nebula is a future Internet architecture that is intrinsically more secure and addresses threats to the emerging computer utility capabilities (called cloud computing) while meeting the challenges of flexibility, extensibility, and economic viability.

2. XIA (http://www.cs.cmu.edu/∼xia)

XIA addresses the growing diversity of network using models, the need for trustworthy communication, and the growing set of stakeholders who coordinate their activities to provide Internet services.

3.2 Key technologies

In addition to new network architectures, many new network technologies have been proposed, and we introduce some promising work in this subsection.

3.2.1 Virtualization of network functions

It is difficult to deploy new services in traditional networks. Thus, NFV (Li and Chen, 2015) has been proposed to benefit information technology (IT) virtualization evolution. NFV separates the network functions from the underlying hardware, and uses general software applications running on commercial servers or virtual machines to implement network functions. Currently, a lot of network functions have been implemented, for example, broadband network gateway (BNG), network address translation (NAT), content delivery network (CDN), and network accelerators.

The NFV architecture defined by the European Telecommunications Standards Institute (ETSI) is shown in Fig. 3. The NFV orchestrator controls the lifecycle of network services. Virtualized network functions (VNF) are deployed and executed

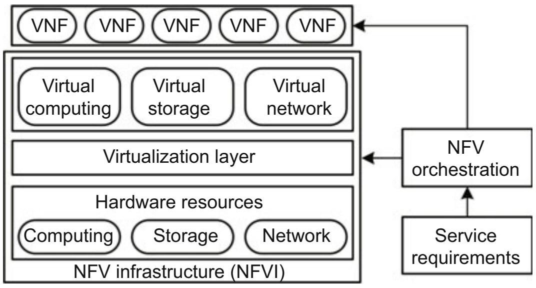

Fig. 3 Network function virtualization architecture

on the network function virtualization infrastructure (NFVI), which consists of commodity services.

The design philosophy of NFV is to reduce the operating expense (OPEX) and capital expenditure (CAPEX). However, the current development of NFV technology leaves much to be desired. On the one hand, flexibility introduces complexity in the NFV management and orchestration system, increasing the OPEX. On the other hand, a VNF cluster does not reduce the deployment cost compared with a dedicated hardware network function with the same performance.

3.2.2 P4

OpenFlow explicitly specifies the protocol headers on which it operates. However, the operation set of OpenFlow has grown from 12 to 41 fields, which increases the complexity of the specification while still not providing the flexibility to add new headers. Thus, P4 (Bosshart et al., 2014) has been proposed to solve the problem. P4 is a programming language designed to allow programming of packet forwarding planes. It has three main goals: (1) reconfigurability in the field (allowing programmers to change the way by which switches process packets); (2) protocol independence (switches are not tied to any protocol such as IP, Ethernet, Transmission Control Protocol (TCP), Virtual extensible Local Area Network (VxLAN), and Multi-Protocol Label Switching (MPLS)); (3) target independence (P4 can be compiled against many different types of execution machines and hardware).

Barefoot networks have designed the barefoot Tofino switch, which is the world's first end-user programmable Ethernet switch (https://barefootnetworks.com/products). Tofino has been built using a protocol-independent switch architecture (PISA) and is P4-programmable. The line rate of Tofino is up to 6.5 Tb/s.

3.2.3 Edge computing

Because data are increasingly produced at the edge of the network, edge computing was proposed to more efficiently process the data at the edge of the network (Shi et al., 2016). Specifically, with the rapid growth of data, the speed of the network becomes the bottleneck of cloud-based computing. If all the data are sent to the cloud for processing, the response time is too long. Thus, to reduce the response time and achieve efficient processing, the data need to be processed at the edge. Fig. 4 shows a simple edge computing example. The things are not only data consumers, but also data producers. Besides, the things can perform the computing tasks from the cloud. Today, edge computing is considered as an indispensable technology in 5G networks to meet the ultra-low-latency requirement.

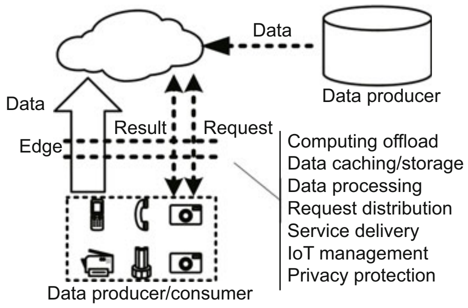

Fig. 4 Edge computing paradigm

4 Use case and applications

SDN, NFV, and other technologies have been used by service providers (Google and Facebook) and telecommunication companies (AT&T). Thus, in this section, we introduce some practical use cases.

4.1 Service providers

With the rapid development of the Internet, service providers now provide more and more applications, and their data centers are also rapidly growing in size. Data centers of companies such as Google, Microsoft, Tencent, and other Internet companies have reached the scale of tens of thousands of physical servers. In this situation, the traditional data center network architecture is unable to meet various demands. Thus, these providers leverage SDN to increase network resource usage.

For example, Google constructed B4 to connect its data centers to one another to replicate data in real time between individual campuses (SDxCentral, 2018). From 2010, Google deployed SDN in B4 and used a combination of Quagga open-source software along with OpenFlow to optimize the interconnected data center. B4 can support dynamic traffic engineering and greatly increase the bandwidth usage between data centers. In April 2014, Google announced the Andromeda virtualization platform based on SDN and NFV technologies. Google's Andromeda allows it to deliver the same capabilities available to its native applications all the way to containers and virtual machines running on the Google cloud platform. Recently, Google proposed Espresso and extended SDN to the peering edge of Google's network where it connects to other networks across the planet. The Espresso technology allows Google to dynamically choose the location, from which to serve individual users based on measurements of how network connections are performing in real time.

In addition, Facebook launched the Open Computing Project (OCP) in 2011 (www.opencompute.org). OCP focuses on redesigning hardware technology to efficiently support the growing demands on computing. In June 2014, Facebook announced a new open-source network switching technology, including the specification of a top-of-rack switch named Wedge and a Linux-based network operating system named FBOSS.

Companies from China have also used SDN. For instance, Tencent deployed an SDN traffic scheduling solution in its wide area network (WAN). In the solution, the distributed control layers were replaced by a centralized control system. The centralized control system uses the global routing algorithm to automatically adjust the load to reduce bandwidth costs. In addition, Baidu, Alibaba, and others have used the SDN technology in their data centers.

4.2 Telecommunication company

With the successful deployment of SDN and NFV in the data center, more and more telecom operators begin to use the new technologies.

For instance, AT&T proposed Domain 2.0 Plan, which is the model for AT&T's next-generation network, powered by technologies including SDN and NFV. By 2020, AT&T plans to virtualize and control over 75% of its network using this new software-defined architecture to meet the growing demands of data and video-hungry users. In addition, Central Office Re-architected as a Datacenter (CORD) (https://opencord.org), which combines NFV and SDN, was used by AT&T with the aim of transforming conventional central offices (COs) in operators' networks into something similar to data centers of cloud service providers.

China's operators have also made active efforts and exploration in the field of software-based networks and the data center cloud. For example, in April 2014, China Telecom and Huawei announced that they have jointly completed the commercial deployment of SDN, and successfully applied the SDN technology to the Internet data center network. In September 2015, China Unicom announced the CUBE-Net 2.0, a new generation of network architecture based on SDN and cloud technologies. In 2015, China Mobile completed the industry's first commercial test of the SDN + NFV data center and formally launched NovoNet, the next-generation innovation network.

5 Service customized network

In this section, we propose the service-customized network (SCN), which is an integrated framework of networking, caching, and computing. SCN combines the strength of SDN and CCN and has the potential to significantly improve the end-to-end performance of applications.

Currently, networking, caching, and computing are considered separately. For example, the design of SDN does not consider how to cache the data to reduce redundant traffic, while CCN considers only caching without solving network management problems. Furthermore, users have diverse needs, and the networking, caching, and computing resources need to be dynamically orchestrated to meet their requirements. For example, video service needs not only customized bandwidth and latency, but also caching resources to reduce traffic. In addition, the sensors may need caches in the network to store their data.

Integrating networking, caching, and computing is a new trend. Recently, technologies such as edge computing try to provide caching and computing at the edge of the network. However, network devices in edge computing are still separated from the caching and computing resources. Thus, we think that the network should provide differentiated services to customers so that customers can define the resources that they need.

To achieve this goal, SCN has been proposed. The architecture of SCN is shown in Fig. 5. SCN has two key features. On the one hand, the data plane of the integrated framework consists of the devices that are responsible for networking, caching, and computing. There is a centralized controller to manage the device to jointly leverage all its functions. Therefore, packets are sent to local caching or computing processing pipelines based on the policies of the centralized controller. On the other hand, SCN achieves virtualization in the control plane. Telecom operators can provide differentiated services by virtualizing the network built on one infrastructure. These virtual networks are allocated different network resources according to customer demand.

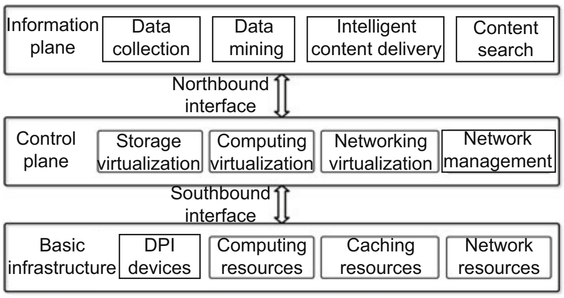

Fig. 5 Architecture of the service customized network

6 Open challenges and broader perspective

Because many new technologies such as machine learning, 4K video, and augmented reality/virtual reality (AR/VR) are increasingly used, these new technologies raise new demands and they may affect the design of the future Internet. In this section, we briefly introduce some cutting-edge technologies to meet the service demands described in Section 2.

6.1 End-to-end deterministic latency

Recall that new services such as Industry 4.0, smart city, autonomous driving, and the Internet of Vehicles require not only a high bandwidth but also a low latency. They usually require deterministic network behaviors such as deterministic or bounded latency. For example, the latency requirements of 4K video, AR/VR, and remote human-computer interaction are 15–35, 5–7, and 20 ms, respectively (Xu et al., 2011).

Recently, some researchers have proposed time-sensitive networking (TSN) (Wollschlaeger et al., 2017) and deterministic networking (DetNet) (Varga, 2017) to provide end-to-end latency guarantees across the Internet by changing the best-effort model of the current Internet. However, TSN can work only in LAN and cannot be used in WAN, whereas the standard and technology of DetNet are working in progress. Thus, how to control the latency and jitter is still a big challenge.

Considering the existing technologies, we think software-defined WAN (SD-WAN) is a promising way to solve the problem. For instance, in SD-WAN, the controller can control all flow paths. Thus, we can use multiple paths and split the WAN paths into two types: low priority and high priority. The latency of high-priority paths is guaranteed by limiting the total traffic on the path, while the latency of low-priority paths is not guaranteed without limiting the total traffic. Furthermore, to control the traffic, we should limit the rate of a flow in the source node instead of in the network.

6.2 Network artificial intelligence

Machine learning, as a new technology, has become one of the hottest topics in both academia and industry. Currently, machine learning brings challenges and opportunities to the future Internet in two aspects.

On the one hand, machine learning gives computers the ability to learn. It can be employed in a range of tasks and areas, such as image or video classification, speech recognition, and prediction. These applications need to analyze massive amounts of data from pictures and videos. These data need to be concurrently analyzed by distributed machines to improve the performance. Thus, the future Internet needs to better support the development of machine learning, and guarantee the fast transmission of its data. On the other hand, the size of the parameters in a very large neural network can be hundreds of megabytes. These parameters are used by distributed machines and need to be sent in microseconds. Thus, the future data center network should provide the lowest latency possible for large distributed systems.

In addition, machine learning offers some promising opportunities in network management. For example, machine learning technologies such as artificial neural network (ANN) and recurrent neural network (RNN) can learn communication patterns from history traces of flows and thus predict future traffic volumes. Using these technologies, we can build an intelligent controller to predict flow size and better schedule the network load. For instance, traffic data were collected from 10 cells in Melbourne for four weeks, and a reinforcement-learning-based scheduler has been proposed to better schedule IoT traffic by learning from the collected traffic data (Chinchali et al., 2018). In addition, Mao et al. (2016) leveraged reinforcement learning to schedule CPU resources according to different objectives.

Most solutions using machine learning are trained offline. However, in the online environment, because traffic may change dynamically, the machine learning algorithm should be trained online to adapt to traffic variations. In addition, the network management plane is usually distributed, whereas the machine learning algorithm is generally centralized. Thus, how to design distributed machine learning algorithms is a big challenge.

6.3 Management and orchestration

Service management and orchestration should be performed according to the real-time network status. The ideal situation is that the network can run in a closed loop by itself without human intervention and achieve optimal status during the running time. However, the new SDN/NFV-based architecture requires significant effort in designing the management and orchestration system. Particularly, multiple layers, multiple domains, and multiple vendors introduce more chaos in system design.

Currently, the management and orchestration of the network bring challenges and opportunities in two aspects. On the one hand, real-time network status data are required as basic elements. Hence, network-wide telemetry technologies are essential. Some recent work (Kim et al., 2015) adopts P4 to perform in-band network telemetry. On the other hand, the management algorithm, as the brain of the network, is crucial. We believe the evolution of the network brain can be divided into three steps. The first step is policy-based management, the second step is the method assisted by artificial intelligence (AI), and in the third step all the decisions will be made by AI.

7 Conclusions

We have surveyed future Internet technologies, which are becoming important research areas. First, we presented the main requirements of future Internet. Then we introduced some important architectures and technologies and examined the recent development and use of these new technologies. Finally, we discussed the research directions and open challenges for the future Internet.

Compliance with ethics guidelines

Jiao ZHANG, Tao HUANG, Shuo WANG, and Yun-jie LIU declare that they have no conflict of interest.

References

Akyildiz IF, Nie S, Lin SC, et al., 2016. 5G roadmap: 10 key enabling technologies. *Comput Netw*, 106:17-48. https://doi.org/10.1016/j.comnet.2016.06.010

Bannour F, Souihi S, Mellouk A, 2018. Distributed SDN control: survey, taxonomy, and challenges. *IEEE Commun Surv Tut*, 20(1):333-354. https://doi.org/10.1109/COMST.2017.2782482

Bosshart P, Daly D, Gibb G, et al., 2014. P4: programming protocol-independent packet processors. *ACM SIGCOMM Comput Commun Rev*, 44(3):87-95. https://doi.org/10.1145/2656877.2656890

Chinchali S, Hu P, Chu TS, et al., 2018. Cellular network traffic scheduling with deep reinforcement learning. Pros 32nd AAAI Conf on Artificial Intelligence, p.766-774.

Chowdhury M, Zaharia M, Ma J, et al., 2011. Managing data transfers in computer clusters with orchestra. Proc ACM SIGCOMM, p.98-109. https://doi.org/10.1145/2018436.2018448

Fisher D, 2014. A look behind the future Internet architectures efforts. *ACM SIGCOMM Comput Commun Rev*, 44(3):45-49. https://doi.org/10.1145/2656877.2656884

Jacobson V, Smetters DK, Thornton JD, et al., 2009. Networking named content. Proc 5th Int Conf on Emerging Networking Experiments and Technologies, p.1-12. https://doi.org/10.1145/1658939.1658941

Jain S, Kumar A, Mandal S, et al., 2013. B4: experience with a globally-deployed software defined WAN. Proc ACM SIGCOMM Conf on SIGCOMM, p.3-14. https://doi.org/10.1145/2486001.2486019

Kim C, Sivaraman A, Katta N, et al., 2015. In-band network telemetry via programmable dataplanes. Proc ACM SIGCOMM, p.1-2.

Li Y, Chen M, 2015. Software-defined network function virtualization: a survey. *IEEE Access*, 3:2542-2553. https://doi.org/10.1109/ACCESS.2015.2499271

Mao HZ, Alizadeh M, Menache I, et al., 2016. Resource management with deep reinforcement learning. Proc 15^{th} ACM Workshop on Hot Topics in Networks, p.50-56. https://doi.org/10.1145/3005745.3005750

Mao HZ, Netravali R, Alizadeh M, 2017. Neural adaptive video streaming with Pensieve. Proc Conf of the ACM Special Interest Group on Data Communication, p.197-210. https://doi.org/10.1145/3098822.3098843

McKeown N, Anderson T, Balakrishnan H, et al., 2008. OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Comput Commun Rev*, 38(2):69-74. https://doi.org/10.1145/1355734.1355746

Mestres A, Rodriguez-Natal A, Carner J, et al., 2017. Knowledge-defined networking. *ACM SIGCOMM Comput Commun Rev*, 47(3):2-10. https://doi.org/10.1145/3138808.3138810

SDxCentral, 2018. Google Brings SDN to the Public Internet. https://www.sdxcentral.com/articles/news/google-brings-sdn-public-internet/2017/04 [Accessed on July 23, 2018].

Shalom N, 2010. Amazon Found Every 100ms of Latency Cost Them 1% in Sales. https://blog.gigaspaces.com/amazon-found-every-100ms-of-latency-cost-them-1-in-sales [Accessed on July 23, 2018].

Shi WS, Cao J, Zhang Q, et al., 2016. Edge computing: vision and challenges. *IEEE Int Things J*, 3(5):637-646. https://doi.org/10.1109/JIOT.2016.2579198

Taleb T, Samdanis K, Mada B, et al., 2017. On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration. *IEEE Commun Surv Tut*, 19(3):1657-1681. https://doi.org/10.1109/COMST.2017.2705720

Valadarsky A, Schapira M, Shahaf D, et al., 2017. Learning to route. Proc 16^{th} ACM Workshop on Hot Topics in Networks, p.185-191. https://doi.org/10.1145/3152434.3152441

Varga B, 2017. DetNet Service Model: draft-varga-detnet-service-model-02. https://datatracker.ietf.org/doc/html/draft-varga-detnet-service-model-02 [Accessed on July 23, 2018].

Wollschlaeger M, Sauter T, Jasperneite J, 2017. The future of industrial communication: automation networks in the era of the Internet of Things and Industry 4.0. *IEEE Ind Electron Mag*, 11(1):17-27. https://doi.org/10.1109/MIE.2017.2649104

Xu XW, Pan YC, Lwin PPMY, et al., 2011. 3D holographic display and its data transmission requirement. Proc Int Conf on Information Photonics and Optical Communications, p.1-4. https://doi.org/10.1109/IPOC.2011.6122872

Zhang LX, Afanasyev A, Burke J, et al., 2014. Named data networking. *ACM SIGCOMM Comput Commun Rev*, 44(3):66-73. https://doi.org/10.1145/2656877.2656887