

1

Introduction

Internet has truly become the service of the digitized society now and delivers a broad range of services such as banking, e-commerce, social networking, media, content storage, and much more. Currently, there is a strong trend to penetrate coverage and usage of Internet by going mobile. The usage by individual is steadily increasing both in time of use and bandwidth demand of application in use. Still there is 80% of the global population that lacks access to Internet. Clearly, the technologies building Internet need to evolve in order to facilitate the steady growth by cost-efficient and sustainable means. Moreover, it is commonly recognized in the technical literature that the Internet has constraints in terms of mobility, quality of service, security, and scalability (e.g., due to IP address starvation and semantic overloading of IP addresses) even if patches exist for fixing any particular problem.

Over the past decades, the telecommunications industry has migrated from legacy telephony networks to telephony networks based on an IP network. This shift allows the mobile network operators to leverage the high bandwidth, multiplexing, innovative products, and services that have been long deployed and tested in IP network and then stimulates a new wave of revenue generation. Since the emergence of cellular data networks, the volume of data traffic carried by cellular networks has been growing continuously due to the innovation of mobile devices, mobile applications, the rapid increase in subscriber size, and cellular communication bandwidth. The trend of cellular data growth will continue to accelerate as technology and application availabilities further improve. Indeed, in 2009, mobile data traffic exceeded mobile voice traffic for the first time [2, 3], starting the new paradigm of mobile networks. The dominant usage of mobile network has shifted from low bandwidth voice and messaging traffic

to a diverse type of data traffic, ranging from web browsing, video streaming, and even online gaming. Existing report [4] has shown that the mobile data traffic is increasing at a rate of 60% per year. By the end of 2016, mobile traffic has surpassed the traffic generated in the wireline network [5, 6]. It is predicted that by the year of 2018, the yearly IP traffic globally will be 64 times of that in 2005.

The demand on mobile network infrastructure will be further increased as the market of Internet of things (IoT) grows. IoT devices and their applications that function without direct human intervention are rapidly becoming an integral part of our lives. IoT devices and applications have wide use cases in a variety of areas, including telehealth, shipping and logistics, utility and environmental monitoring, industrial automation, and asset tracking. It is predicted to be 25 billion devices – including fixed/mobile personal devices and IoT devices – connected to IP networks [5] by the year of 2020 with a seven trillion dollar market value. The Machine-to-Machine (M2M) communications will generate 80% more data than the human directly generated data in the year of 2018. The prosperity of IoT industry requires great support of the wide area wireless communication infrastructure, in particular, the cellular data networks.

To cope with the explosive cellular data volume growth and best serve their customers, cellular network operators need to design and manage cellular core network architectures accordingly. There are two main challenges that the cellular network faces. The first challenge is the increasing cost of network operation. While the most advanced cellular technologies today is the 4G technology such as long term evolution (LTE). However, the coexistence of multiple generations of cellular networks is unavoidable today. This is because the subscribers may gradually upgrade their devices over the years. Thus, the operators need to support multiple technologies over a long period of time and manage multiple networks simultaneously. The operational cost may not be covered by the revenue growth. The second challenge is the high demand of network capacity expansion. Many techniques have been proposed to expand the cellular network capacity such as multi-antenna technologies and Wi-Fi offloading architecture. Adding spectrum and deploying small/femto cells have also been used together to further expand the network capacity, which is, however, expensive and not easy to deploy. Moreover, the changes of infrastructure still cannot keep up with the exponential growth of traffic demand.

The current cellular infrastructure is not capable of addressing the explosive need of data demand. The main reason is that the resource is configured and allocated in a rather static manner. The resources are not utilized in an efficient way. The traffic demand, however, are highly dynamic, exhibiting a time-of-day phenomenon [5, 6]. Flash crowds happen often due to popular events. To fundamentally meet the traffic demand, blindly increasing the network capacity is not enough. We need to find ways to better utilize existing capacity.

The traditional business model of cellular carriers was based on revenues for telephony. The Internet was an over-the-top service, priced by online minutes or data volumes during the late 1990s. This has changed completely to a flat-rate-based model for Internet access, and operators deploy broadband networks to cover demands of the digital society, struggling to return revenues needed to deploy even higher speed networks. Governance and regulations have further limited profitability of operators. In order to increase revenues, operators are deploying a number of service-centric networks on top of the broadband infrastructure such as IPTV, demanding new functionality of carrier network. On the cost side, carrier operators would like to reduce the capital and operational cost significantly. In the following, we discuss the opportunities of leveraging existing cloud technologies for better resource utilization. We will review the related technologies and then outline the rest of this book.

1.1 Cloud-Enabled 5G: SDN and NFV

The Internet has successfully been growing for more than 20 years; the growth in demand has so far been met by introducing even larger and larger routers. This has been beneficial and to scale in public networks. However, in order to meet today's steadily growing demand for Internet access and other packet-based services, there is a present need to deploy more efficient packet networks also within metro and aggregation network domain. The attempt to copy the approach from the coarsely populated, but large core network sites and migrate to metro and aggregation network sites may not be the most cost optimal approach. It may be time now to split the router architecture in similar ways as was done in the traditional mobile core network, in order to penetrate the highly dense metro/aggregation networks.

Splitting the router control and forwarding plane forms the initial idea of software-defined networking (SDN).

The initial idea was born to decouple the routing intelligence software from simple forwarding hardware allowing, particularly for academic research networks and test beds, fast prototyping and evaluation of new control theories and algorithms [7]. It was part of the Clean Slate Internet Design initiative of Stanford University [8]. The target is to develop a system that is amenable to high-performance and low-cost implementations and capable of supporting a broad range of research, can isolate experimental traffic from production, and is consistent with vendors' need for closed platforms.

The key technical idea of SDN is to provide an open control interface to the operating system of the network device without compromising the details of the implementation, an important business aspect for equipment manufacturers. This is enabled by support of OpenFlow [8] in the operating system and is based on the Ternary Content Addressable Memory (TCAM)-based flow tables, most routers and switches make use of. In a classical router or switch, the fast packet forwarding data path and the high-level routing decisions in the control path occur on the same device. An OpenFlow-enabled switch separates these two functions. The data path portion still resides in the switch, while high-level routing decisions are moved to a flow controller, typically a standard server. The OpenFlow Switch and Controller communicate via the OpenFlow protocol, which defines operation and management (OAM) messages.

Besides this technical view, this split design will enable a cost reduction and new market opportunities by the basic principle of modularization. This is of high importance for supporting flexible network innovations because the development cycles of hardware and software components are extremely different, and the modularization supports a decoupling of the innovations from a market perspective. The right layering approach will enable high market volumes for specific modules (software or hardware).

The introduction of the SDN concept into real networks would have a profound impact on the way in which networks are built and operated. In order to understand and evaluate the practical implications of the general concept, it would be beneficial to first test it in research networks. Feedback from the experimental implementation will be crucial in improving the overall concept and allow taking

the concept to further applications in networking. First trials are currently under way in selected US universities, which focus on the easy management and reconfiguration of research networks, for example, for applications in the field of Clean Slate research.

While SDN brings innovative evolution to network routers and switches, the network comprises other types of devices besides routers and switches. Network operators enforce network policies using a combination of switches and network functions (NF). Policies may be complex, such as ensuring that unauthorized users are prevented from accessing sensitive servers or malicious traffic is eliminated from the network. To do this, an operator could use a stateful firewall to ensure that only traffic initiated from within the network is permitted and in doing so protect users from malicious traffic. Indeed, today's networks heavily rely on a wide spectrum of NFs. The diversity and complexity of NFs have been further expanded as the proliferation of wireless devices and mobile applications. NFs offer a variety of valuable benefits, ranging from improving security (e.g., firewalls, intrusion detection systems, and deep packet inspection), improving performance (e.g., proxies, caches) and reducing bandwidth costs (e.g., WAN optimizers, video transcoder). However, despite their benefits, NFs come with high infrastructure and management costs. One important reason is their complex and specialized processing. As a direct result of this complexity, configuration errors are common – configuration errors comprise as much as 65% of the network outages [9]. Other reasons of their complexity come from the lack of standardized management tools across different devices and vendors. Moreover, there is a need to consider policy interactions between these appliances and other network infrastructure, which cannot be easily troubleshot.

To facilitate programmability and flexibility of NFs, in 2012, operators initiated a new concept, called network functions virtualization (NFV) within the European Telecommunications Standards Institute (ETSI) consortium [10]. Instead of building NFs in the form of proprietary hardware boxes, NFV calls for the virtualization of them. Using virtualization and cloud technologies, it allows legacy NFs to be flexibly deployed in the form of software on commodity servers. Sharing the same spirit of splitting the router's control plane from forwarding plane, the decoupling of NF software from the hardware facilitates a faster pace for innovations and shorter development cycles, and result in shorter time to market of new services.

1.1.1 Benefits

One important benefit of the SDN and NFV is the potential to open new business opportunities in network architecture, related systems, and hardware. Today, when a new network protocol or network application is invented, the entire industry needs to go through the tedious standardization process by established organizations such as the Internet Engineering Task Force (IETF), International Telecommunication Union (ITU), and Institute of Electrical and Electronics Engineers (IEEE), to name but a few. After the standardization process, the development cycle and extensive testing activities may last 18–24 months, ensuring the performance, reliability, and interoperability with other equipments. While the seamless interconnectivity plays an important role in the success of the Internet and its applications, however, the long duration of standardization process and development cycle has made today's networks difficult to change. This is particularly true for research and innovations related to test and validate novel networking approaches, for example, new or revised routing or forwarding schemes. Such modifications currently require the close collaboration between operators and system manufactures, and hence demand substantial resources in terms of time and manpower.

The essence behind SDN and NFV is the decoupling of functionality. Such decoupling offers the opportunity to change the current limitations. In SDN, modifications to the control plane can be independently done, while the switching hardware remain unaltered. In NFV, the separation of software and hardware of NFs allows each part to grow and innovate independently. The modularity design choice in SDN and NFV comes with a number of economic advantages, especially in the following areas:

- Lowering the entry barriers for control plane product and switching plane gears
- Independent resource/budget allocation of each module
- Providing high extensibility because each module has clear interface with the rest
- Enabling rapid prototyping, instead of waiting for every component to be ready
- Allowing new players to enter the market. There are mutual benefits to both traditional equipment vendors and network operators, enabling small players to grow

- Increasing the speed of innovation. One key obstacle to any network innovation deployment is the incremental deployability and incentives for the first movers. Seamless interoperability usually slows down the deployment process and decreases the incentives for first adopters. With the modular design, first movers can gain technological advantages more easily when deploying new networks and services.

1.1.2 Challenges

While SDN and NFV have great potentials as discussed earlier, applying them in production networks faces several key challenges.

- *Scalability*: The SDN and NFV architectures need to scale to support millions of subscribers, increasing traffic demand and future growth. It needs to be able to meet the necessity of adding more controllers, switches, and NFs to a network. There is a potential need for a hierarchy of control components, or a chain of NFs, and might take a peer-to-peer approach to further expand the scalability.
- *Performance*: Carrier network has high requirement on performance than enterprise or data center networks. The links are usually tens to hundreds of Gigabits per second capacity. The customers' perceived performance should meet the predefined service level agreement (SLA). In SDN, how to deal with the response time of the controller considering the need to configure flows and how to provide sufficient bandwidth using a combination of gears are important questions to be answered. In NFV, how to control and reduce the additional overhead introduced by traversing several NFs and how to hide the overhead of virtualization techniques are critical to meeting the performance requirement.
- *Openness*: One of the key questions to enable openness is at what level of abstraction should be defined for different interfaces. The trade-offs between modularity and extensibility need to be carefully examined. For example, how one can add additional features to interfaces without affecting other parts of the system.
- *Internetworking and interoperability*: On the one hand, the network will gradually change. Seamlessly interconnecting SDN/NFV networks with traditional legacy networks is critical to its deployment. The devices supporting the new technologies should be able to communicate with legacy devices using traditional protocols, and the performance should not be compromised in this process.

- *Sustainability*: Energy efficiency has drawn significant attentions in the carrier network recently, both for the sake of reducing cost and for meeting the governance regulation requirements. The ability to switch off functionality and hardware, which is currently not in use, and how to make most efficient use of available processing capabilities are the key enablers to an energy-efficient network.
- *Resiliency*: In evaluating a network design, the network resilience is an important factor, as a failure of a few milliseconds may easily result in terabyte data losses on high-speed links. In traditional networks, where both control and data packets are transmitted on the same link, the control and data information are equally affected when a failure happens. In the SDN and NFV deployment, we need to come up with strategies to handle individual components' failures and to perform failure recovery to meet the carrier grade's five nines availability requirement.

1.2 Supporting Technologies

In this section, we briefly highlight the high-level ideas of related technologies. We will go into details of each of them in the rest of this book.

1.2.1 Cloud Computing

Cloud computing is transforming the way people use computers and how application services are run. While it has been widely used as the platform of choice for many web services, it is also becoming the hosting platform of network services. Virtualization is the key underlying technology enabling cloud providers to host services for a large number of customers. With virtualization, the physical resources and the applications are loosely coupled. The cloud customer is able to dynamically provision infrastructure to meet the current demand by leasing resources from the cloud infrastructure provider. On the other hand, the cloud provider can leverage economies of scale to provide dynamic, on-demand infrastructure in the best cost-effective manner. Using virtualization techniques [11], multiple virtual machines from different customers share the same physical servers. This is called *multitenancy*, which allows independent customers to lease resources from the cloud provider. Virtualization initially started with host virtualization, such as virtual machines

(VMs) or container. Later, the same concept has been extended to network virtualization, network functions virtualization (NFV), and so on. These technologies move the networking industry from today's manual configuration to more automated and scalable solutions. They are complementary approaches that solve different subsets of network mobility problem. SDN and NVF are among other initiatives to move from the traditional cellular infrastructure toward a cloud-based infrastructure across multiple areas including radio access network, core network, backhaul, and operational/business support systems (OSS/BSS). We elaborate them more as follows.

1.2.2 Network Virtualization

Network virtualization gives each customer tenant its own network topology and control over the flow of its traffic. For sharing computing resources, people have been widely used the virtual machine as the standard abstraction. For network virtualization, however, the right abstraction is still a subject of ongoing debate. There are a diverse set of solutions, which differ in the level of concrete details they expose to the individual tenants. In the widely used Amazon EC2, the network is abstracted in the simplest form. All of a tenant's virtual machines are connected with each other using best effort network fabric. VMware [12] offers a simple "one big switch" abstraction, where different rules such as access control or traffic engineering can be programmed by the tenants. The rules are implemented at the network edge (in the virtual switches inside the host operating systems). As more applications move to the cloud, providers must go beyond simple sharing of network bandwidth to support a wider range of abstractions. Recently, SDN has become a natural platform for network virtualization, thanks to the standard interface between controller applications and switch forwarding tables. However, supporting a large number of tenants with different topologies and controller applications raises scalability challenges.

1.2.3 Network Functions Virtualization

Network Functions Virtualization is a recent trend prompted by technology availability that makes high-performance packet processing now possible on commodity systems reasonably well, and service providers' desire for virtualization of network functions such as routers, firewalls, network address translation (NAT), and so on

to avoid vendor lock-in and ability to choose virtualized network functions from different vendors as per the requirements. NFV not only allows service providers the ability to compose network functions but also the elasticity to flex compute and storage resources required by different VNFs depending on the traffic patterns and distributions. It provides a new way to design, deploy, and manage network services. It decouples the network functions from purpose-built hardware, so they can run in software. Most of the current focus in this space has been on virtualizing more and more network functions and cloud-like orchestration systems for managing deployed VNFs.

1.2.4 Software-Defined Networking

SDN is a new approach to designing, building, and managing networks, which enables the separation of the network's *control plane* and *data plane*, which makes it easier to optimize each. SDN has the potential to make significant improvements to service request response times, security, and reliability.

In SDN, the controllers collect information from switches, and compute and distribute the appropriate forwarding decisions to switches. Controllers and switches use a protocol to communicate and exchange information. An example of such protocol is OpenFlow [13], which provides an open and standard method for a switch to communicate with a controller, and has drawn significant interests from both academic and industry.

In summary, NV, NFV, and SDN each provide a software-based approach to networking, in order to make networks more scalable and innovative. Hence, some common beliefs guide the development of each. For example, they each aim to move functionality to software, use general-purpose hardware in lieu of purpose-built hardware, and support more efficient network services. Nevertheless, note that SDN, NV, and NFV are independent, though mutually beneficial.

1.3 Outline of Chapters

This book is divided into five chapters and provides information on the different technologies that have been examined in the design of 5G next-generation networks.

In Chapter 1, we introduce the background of carrier networks, the challenges of cellular network operations, and the requirements

for future network moving into cloud. We discuss the advantages of considering SDN and NFV in the telecom communication networks. We also outline the developing technology that are relevant to 5G cloudification design.

Chapter 2 reviews various wired network virtualization technologies and wireless virtualization. It then studies the state of the art in cloud computing, the service models, and its relationship with communication networks.

In Chapter 3, we provide a survey of the existing NFV technologies. We first present its motivation, use cases, and architecture. We then focus on its key use case, the service function chaining, and the techniques and algorithms.

In Chapter 4, we provide a review of the SDN technology and business drivers, describe the high-level SDN architecture and principles, and give three scenarios of its use cases in mobile access aggregation networks and the cloud networks. Furthermore, we provide discussions on the design implementation considerations of SDN in the mobile networks and the cloud, in comparison with traditional networks.

In Chapter 5, we review the case studies of NFV in the next-generation 5G network. In particular, we discuss several use cases of SDN and NFV in the packet core network and in customer premise equipment (mobile edge networks). In both case studies, we discuss the challenges and the opportunities.

The main objective of Chapter 6 is to set the stage of existing activities in the industry from both standardization and open source consortium point of view. The standards movement drives the future implementation and use cases of these technologies. Given telecom industry has traditionally taken the route of standardization followed by implementation, IETF, ETSI, and third generation partnership project (3GPP) play an important role in the future design of SDN and NFV. On the other hand, leveraging the experience of other application services, more and more network services are moving toward open source contribution approach to accelerate the functionality development cycle. Recently, there are open source communities formed in the network service area. In this chapter, we summarize their activities, relationships, limitations, and future directions.

