

Chapter 2

Virtualization and Cloud Computing

In an effort to move the networking industry from today's manual configuration to embrace automated solutions that are coordinated with the rest of the infrastructure, there have been several emerging technologies in the past few years, chief among them are *network virtualization* (NV), *network functions virtualization* (NFV), and *software-defined networking* (SDN).¹ Broadly speaking, all these three solutions are designed to make networking more automated and scalable to support virtualized and cloud environments. These technologies are software-driven schemes that promise to change service and application delivery methods, so as to increase network agility. They are different but complementary approaches (with some overlapping terminology) to provide network programmability. In other words, they solve different subsets of the macro issue of *network mobility*.

It is important to mention that, despite the fact that SDN, NV, and NFV are mutually beneficial, they are not dependent on one another. That is, NFV and NV can be implemented without an SDN being required and vice versa, but SDN makes NFV and NV more compelling and vice-versa.

In this chapter, we review the state of the art in network virtualization and investigate the challenges that must be addressed to realize a viable network virtualization environment.

2.1 What Is Virtualization?

In computing, *virtualization* is the process of abstracting computing resources such that multiple applications can share a single physical hardware. Put differently, virtualization refers to the creation of a *virtual*, rather than actual, version of a resource. The canonical example of virtualization is “server virtualization,” in which certain attributes of a physical server are decoupled (abstracted) and reproduced in

¹The first two technologies are covered in this chapter whereas SDN is discussed in the next chapter.

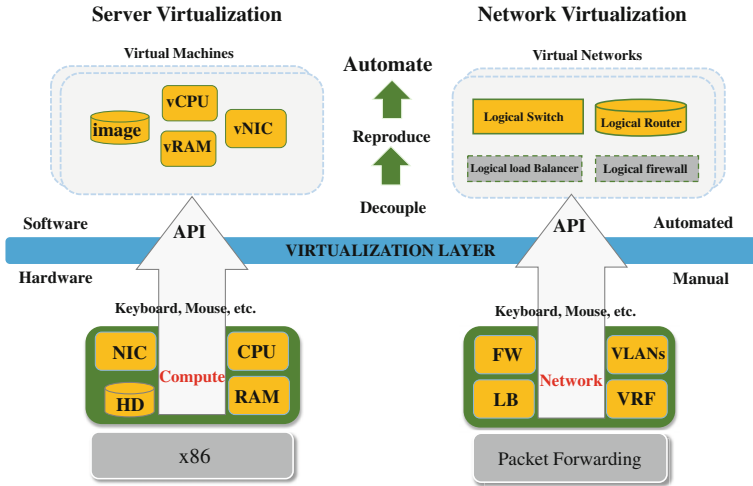


Fig. 2.1 Server virtualization: the physical hardware resources are mapped to multiple virtual machines, each with its own CPU, memory, disks, and I/O devices

a hypervisor (virtualization software) as vCPU, vRAM, vNIC, etc.; these are then assembled arbitrarily to produce a virtual server, in few seconds.

Computing resources are not the only resources that are virtualized; *storage* can be virtualized too. Through virtualization either one resource is shared among multiple users or multiple resources, e.g., storages, are aggregated and presented as one or more high capacity resource that can be used by one or multiple users. In any of those cases, the user has the illusion of sole ownership.

Besides computing and storage, a *network* may be virtualized too. That is, the notion of abstraction can be extended from computing resources and storage to the fundamental components of the networks, i.e., nodes and links. Therefore, in a broader context, virtualization refers to the creation of a *virtual* version of a resource, such as an operating system, a storage device, or network resources.

Server and desktop virtualization is a mature technology now. Virtualization software, e.g., VMware Workstation [34], maps the physical hardware resources to the virtual machines that encapsulate an operating system and its applications, as shown in Fig. 2.1. Each virtual machine fully equivalent of a standard x86 machine, as it has its own central processing unit (CPU), memory, disks, and I/O devices.

Thus far, it should be clear that with virtualization a system pretends to be more than one of the same system. In the following, we will see why this property is important and how it can make networking more programmable and agile [35].

2.2 Why Virtualization?

There are three major technical benefits of improving availability, enabling mobility, and improving utilization, but the benefits do not stop there. They extend to simplifying the IT architecture, and ultimately to accurately aligning billing with consumption.

The basic motivation for virtualization is to efficiently share resources among multiple users. This is similar to multitasking operating systems where, rather than doing one task at a time, unused computing power is used to run another task. Consider an organization that has many servers all doing single or a small cluster of related tasks. Without losing the security of isolated environments, virtualization allows these servers to be replaced by a single physical machine which hosts a number of virtual servers. Similarly, storage aggregation enhances the overall manageability of storage and provides better sharing of storage resources.

Migration is another big advantage of virtualization. It comes in handy if an upgrade is required or when the hardware is faulty because it is fairly simple to migrate a virtual machine from one physical machine to another. Therefore, increasing backup capability is another compelling reason for virtualization. If a server crashes, the data on that server can be set to be automatically transferred to another server in the network. Such a redundancy increases availability, too. What is more, a virtual machine offers a much greater degree of isolation [36].

Saving on physical machine costs, reduced energy consumption, and smaller physical space requirement are among other notable advantages of virtualization. There are also business benefits for a virtualized enterprise, including flexible sourcing, self-service consumption, and consumption-based billing. Most of those advantages are applicable to NV and/or NFV. These two paradigms, however, provide other unique advantages that will be discussed in detail, later in this chapter. Among them are, improving security [37, 38], accelerating time-to-market, and extending accessibility.

2.3 Network Virtualization

Network virtualization refers to the technology that enables partitioning or aggregating a collection of network resources and presenting them to various users in a way that each user experiences an isolated and unique view of the physical network [39–41]. The abstraction of network resources may include fundamental resources (i.e., links and nodes) or derived resources (topologies) [39]. This technology may virtualize a network device (e.g., a router or network interface card (NIC)) a link (physical channel, data path, etc.), or a network. As a result, similar to server virtualization which reproduces vCPU, vRAM, etc., network virtualization software may reproduce logical channel (L1), logical switches, logical routers (L2–L3), and

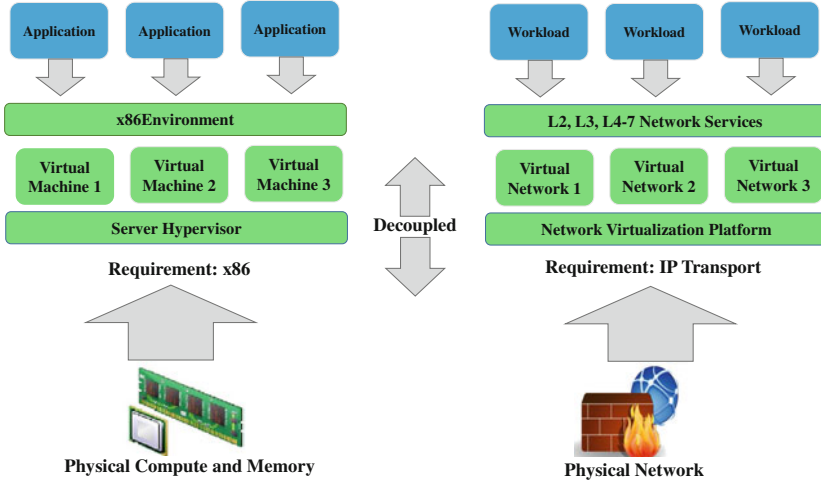


Fig. 2.2 Network virtualization versus server virtualization

more. These logical resources, along with L4–L7 services,² can be assembled in an arbitrary topology, presenting a complete L1–L7 virtual network topology. Thanks to network virtualization, multiple logical networks can coexist and share a physical network.

Network virtualization decouples the roles of the traditional Internet service providers (ISPs) into infrastructure providers (InPs) and service providers (SPs) [40]. InPs and SPs are two independent entities: the former manages the physical infrastructure whereas the latter creates virtual networks by aggregating resources from one or multiple InPs and offers end-to-end services. This decoupling will proliferate deployment of coexisting heterogeneous networks free of the inherent limitations of the existing Internet [40–42]. It is also a way to automate the network to improve networking administrators’ responsiveness to change. Indeed, it is hard to keep up with too many requests for network configuration changes, that can take days or weeks to handle. By allowing multiple heterogeneous networks to cohabit on a single physical architecture, network virtualization increases flexibility, security, and manageability of networks (Fig. 2.2).

In view of the great degree of flexibility and manageability it offers, network virtualization has become a popular topic of interest, both in academia and industry, during recent years. However, the term network virtualization is somewhat overloaded and several definitions, from different perspectives, can be found in the literature (cf. [40, 43, 44]). The concept of multiple coexisting logical networks over a shared physical network has frequently appeared in the networking literature under several different names. Chowdhury and Boutaba [40, 41] classify them into four

²L4–L7 services also can be virtualized to produce logical load balancers or logical firewalls, for example. This is referred to as *network functions virtualization* and will be discussed in Sect. 2.4.

main categories of virtual local area networks (VLANs), virtual private networks (VPNs), overlay networks, and active and programmable networks, whereas in a recent paper [39], Wang et al. divide them into three main groups of VPNs, overlays, and virtual sharing networks (VSNs). We follow the later one due to its clearer and simpler calcification.

Broadly speaking, three types of commercial virtual networks exist which are described in the following.

2.3.1 *Overlay Networks*

An *overlay network*³ is a logical network that runs independently on top a physical network (underlay). Overlay networks do not cause any changes to the underlying network. Peer-to-peer (P2P) networks, virtual private networks (VPNs), and voice over IP (VoIP) services such as Skype are examples of overlay networks [40, 41, 45]. Today, most overlay networks run on top of the public Internet, while the Internet itself began as an overlay running over the physical infrastructure of the public switched telephone network (PSTN). The Internet started by connecting a series of computers via the phone lines to share files and information between governmental offices and research agencies. Adding to the underlying voice-based telecommunications network, the Internet layer allowed data packets transmission across the public telephone system, without changing it.

P2P networks are an important class of overlay networks [46]; they use standard Internet protocols to prioritize data transmission between two or more remote computers in order to create direct connections to remote computers, for file sharing. P2P networks use the physical network's topology, but outsource data prioritization and workload to software settings and memory allocation.

Although there are various implementations of overlays at different layers of the network stack, most of them have been implemented in the application layer on top of IP, and thus, they are restricted to the inherent limitations of the existing Internet.

2.3.2 *Virtual Private Networks*

Many companies have offices spread across the country or around the globe, and they need to expand their private network beyond their immediate geographic area, so as to keep fast and reliable communications among their offices. Until recently, such a communication has meant the use of *leased lines* to deploy a *wide area network* (WAN). A WAN is preferred to a public network (e.g., the Internet) for its reliability, performance, and security. But, maintaining a WAN is expensive, especially when leased lines are required. What is more, leased lines are not a viable solution if

³Here, the network refers to a telecommunication or computer network.

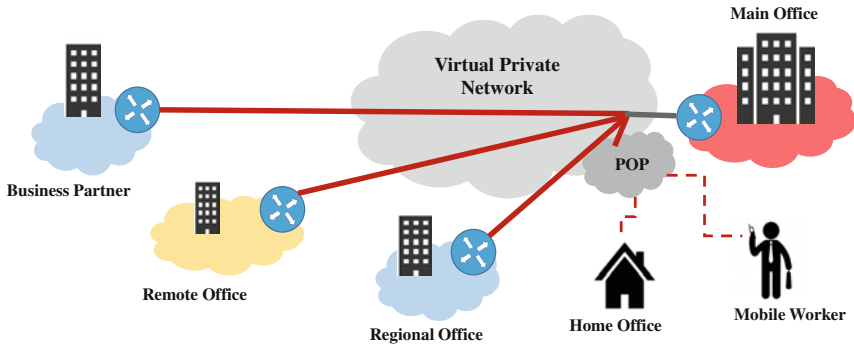


Fig. 2.3 Typical VPN topology: a private network deployed using a public network (usually the Internet) to securely connect remote sites/users together

part of the employees need to access the corporate network remotely, from home, from the road, or from other organizations. On account of ubiquitous Internet, many companies create their own virtual private networks (VPNs) to accommodate the needs of the remote (mobile) workforce and distant offices.

A virtual private network (VPN) is an assembly of two or more private networks or individual users that uses secured tunnels over a public telecommunication infrastructure, such as the Internet, for connection. A VPN is commonly used to provide distributed offices or individual users with secure access to their organization's network. This is illustrated in Fig. 2.3. It is *virtually* private as it uses *public* infrastructure to provide remote access; this access is, however, secure as if the organization uses its private (owned or leased) lines for remote connection. VPNs are meant to provide the organizations with the same capabilities as WANs but at a much lower cost. They can be remote access (connecting an individual user to a network) or site-to-site (connecting two networks together) [47].

Although VPNs use the shared public infrastructure, they maintain privacy through security procedures and *tunneling*⁴ protocols such as secure socket tunneling protocol (SSTP), point-to-point tunneling protocol (PPTP), and layer two tunneling protocol (L2TP). These protocols encrypt the data and send it through a "tunnel" that cannot be entered by data that is not properly encrypted. It is worth mentioning that there is another level of encryption for the originating and receiving network addresses.

A well-designed VPN needs to incorporate security, reliability, scalability, network management, and policy management [47]. Such a network not only extends geographic connectivity, but can improve security, simplify network topology, reduce

⁴Tunneling is a mechanism used to send unsupported protocols across different networks; it allows for the secure movement of data from one network to another. Specifically, tunneling refers to the transmission of data intended to be used only within a private network through a public network such that the routing nodes in the public network are oblivious to the fact that the transmission is part of a private network [48].

transit time, and transportation costs for remote users, provide global networking opportunities and telecommuter support, reduce operational costs compared to traditional WANs, and improve productivity [47]. The above benefits motivate the organizations to deploy VPNs; there are other reasons, primarily for individuals, to start using a VPN [49], for example,

- One can use a VPN to connect securely to a remote network via the Internet. Most companies and nonprofit organizations, including universities, maintain VPNs so that employees can access files, applications, and other resources, from home or from the road, without compromising security.⁵
- Where online privacy is a concern, connecting to a VPN is a smart, simple security practice, when you are on a public or untrusted network, such as a Wi-Fi in a hotel or coffee shop. It helps prevent others who may be trying to capture your passwords.
- VPNs turn out to be very useful in circumventing regional restrictions (censorship) on certain websites. They can also be used for recreational purposes; for example, one can connect to a US VPN to access the US-only websites outside the US. Note that, many media websites (CNN, Fox, Netflix, etc.) may impose a geographical restriction on viewing their content and online videos.
- A VPN comes in handy to virtually “stay home away from home.” Meaning that, one can virtually reside in a specific country from abroad with an IP of that country. It happens that we need to access our geographically restricted accounts, such as online banking and state websites when traveling abroad. Such websites often restrict all access from abroad, for security or other reasons, which can be very inconvenient.

Based on the layer at which the VPN service provider’s interchange VPN reachability information with customer sites, VPNs are classified into three types: layer 1 VPN (L1VPN), layer 2 VPN (L2VPN), and layer 3 VPN (L3VPN) [39, 40, 50, 51]. While L1VPN technology is under development, the other two technologies are mature and have been widely deployed. Also, based on their networking requirements, enterprises can connect their corporate locations together in many different ways. These networking services can typically be viewed from three perspectives, which are demarcation point (or enterprise/service provider handoff), the local loop (or access circuit), and the service core. Choosing layer 2 or layer 3 VPN will make a different impact on these three network services [52].

In an L2VPN, the service provider’s network is virtualized as a layer 2 switch whereas it is virtualized as a layer 3 router in an L3VPN [39]. In the former, the customer sites are responsible for building their own routing infrastructure. Put differently, in an L3VPN, the service provider participates in the customer’s layer 3 routing, while in an L2VPN it interconnects customer sites using layer 2 technology.

As listed in Tables 2.1 and 2.2, both Layer 2 and Layer 3 services have their advantages and disadvantages. These are basically related to the differences of router and switch in computer networking; some of them are highlighted in Table 2.3.

⁵You can also set up your own VPN to safely access your secure home network while you are on the road.

Table 2.1 Layer 2 VPNs: advantages and disadvantages

Advantages
Highly flexible, granular, and scalable bandwidth
Transparent interface—no router hardware investment is required
Low latency - switched as opposed to routed
Ease of deployment—no configuration required for new sites
Enterprises have complete control over their own routing
Disadvantages
Layer 2 networks are susceptible to broadcast storms—due to no router hardware
No visibility from the service provider— monitoring services can be difficult
Extra administrative overhead of IP allocations—because of flat subnet

Table 2.2 Layer 3 VPNs: advantages and disadvantages

Advantages
Extremely scalable for fast deployment
Readiness for voice and data convergence
“any to any” connectivity—a shorter hop count between two local sites
Enterprises leverage the service provider’s technical expertise for routing
Disadvantages
Increased costs—due to requiring customer router hardware
Class of service and quality of service usually incur additional fees
IP addressing modifications would have to be submitted to the service provider

2.3.3 Virtual Sharing Networks

VPNs and overlays are not the only types of virtual networks implemented so far; there exist other networks that do not fall into these two categories. Virtual *local area networks* (Virtual LAN’s) are examples of these networks. While properly segmenting multiple network instances, such technologies commonly support sharing of physical resources among them. The term *virtual sharing networks* (VSNs) has recently been suggested for these types of networks [39].

Originally defined as a network of computers located within the same area, today LANs are identified by a single *broadcast domain* in which the information broadcasted by a user is received by every other user on that LAN while it is prevented from leaving the LAN by using a router. The formation of broadcast domains in LANs depends on the physical connection of the devices in the network. Virtual LANs (VLANs) were developed to allow a network manager to logically segment a LAN into different broadcast domains. Thus, VLANs share the same physical LAN infrastructure but they belong to different broadcast domains. Since it is a logical, rather than a physical, segmentation, it does not require the workstations to be

Table 2.3 Router versus switch

	Router	Switch
Definition	A router is a network device that connects two or more networks together and forwards packets from one network to another	A switch is a device that connects many devices together on a computer network. It is more advanced than a hub
OSI Layer	Network Layer (L3) devices	Data Link Layer. Network switches are L3 devices
Data form	Packet	Frame (L2 switch)/Frame and Packet (L3 switch)
Address used for data transmission	IP address	MAC address
Table	Stores IP addresses in routing table and keeps them on its own	A network switch stores MAC addresses in a lookup table
Transmission type	At initial level broadcast then unicast and multicast	First broadcast; then unicast and multicast as needed
Routing decision	Takes faster routing decision	Takes more time for complicated routing decision
Used to connect	Two or more networks	Two or more nodes in the same or different network

physically located together. They can be on different floors of a building, or even in different buildings. Further, broadcast domain in a VLAN can be defined without using routers; instead, bridging software is used to define which workstations belong to the broadcast domain, and routers are only used to communicate between two VLAN's.

The sharing and segmentation concept of the VLAN can be generalized to a broader set of networks, collectively called *virtual sharing networks* (VSNs). The key requirement for such networks is to share a physical infrastructure while being properly segmented [39]. For example, a large corporate may have different networks with specific permission for guests, employees, and administrators, yet all sharing the same access points, switches, router, and servers.

2.3.4 Relation Between Virtual Networks

A virtual network can be considered “virtual” from different perspectives, so its type may change simply by changing the perspective. An overlay network is virtual as it is separated from the underlying physical network; a VPN is virtual since it is distinct from the public network; VSNs are virtual because multiple segmented networks share a same physical infrastructure. With these views, VPN can be considered an

overlay network, as the tunnels used for connection are separate and external to the private network and used to extend the functionality and accessibility of the primary physical network. Likewise, overlay networks sharing the same underlay become VSN. However, it should be noted that overlay, VPN, and VSN respectively emphasize on new services, connectivity, and resource sharing.

In summary, network virtualization is an overlay; that is, to connect two domains in a network, it creates a tunnel through an existing network rather than physically connecting them. It saves administrators from having to physically wire up each new domain connection; especially, they need not change what they have already done; they make changes on top of an existing infrastructure.

While NV creates tunnels through a network, the next step to automate the network is to put services, such as firewall, on tunnels. This is what NFV offers, and is explained in the following section.

2.4 Network Functions Virtualization

There are increasing variety of proprietary hardware appliances to launch different services in telecommunication networks. Launching a new network service yet often requires another appliance, implying further space and power to accommodate these boxes, in addition to increased integration and deployment complexity. Further, as innovation accelerates, lifecycles of these hardware-based appliances becomes shorter and shorter, meaning that the return on investment is reduced. The above problems could be addressed if the services are run in *software*. Thus, enlightened by virtualization, the following question would arise: If administrators can set up a virtual machine by a click, why shouldn't they launch a service in a similar fashion?

Network functions virtualization⁶ (NFV) decouples *network functions* from proprietary hardware appliances, to overcome the above deficiencies (Fig. 2.4). It offers a new way to architect and implement layer 4 through layer 7 network functions in order to run computationally intensive network services in software, that can be moved to standard hardware. Through decoupling layer 4–7 network functions, such as firewall, intrusion detection, and even load balancing, from proprietary hardware appliances, and implementing them in software, NFV offers a cost effective, and more efficient way to design, deploy, and manage networking services. Network functions virtualization is targeted mainly at the carrier or service provider market, and it enables operators to [53, 54]:

⁶The Network Functions Virtualization Industry Specification Group (NFV ISG) was initiated under the auspices of the European Telecommunications Standards Institute (ETSI). NFV ISG first met in January 2013, and will sunset two years after [53]; it included over 150 companies in 2013. The NFV ISG objective is not to produce standards but to achieve industry consensus on business and technical requirements for NFV. A more detailed version of [53] is expected to be released in the second half of 2014.

- **Reduce CapEx:** NFV reduces the need to purchase purpose-built hardware by using commercial off-the-shelf hardware which is typically less expensive than purpose-built, manufacturer-designed hardware. By shifting more components to a common physical infrastructure, operators save more. Also, through supporting pay-as-you-grow models, which eliminate wasteful overprovisioning, operators can save even more.
- **Reduce OpEx:** NFV reduces space, power, and cooling requirements of equipment since all services utilize a common hardware. Further, it simplifies the roll out and management of network services as there is no need to support multiple hardware models from different vendors.
- **Accelerate time-to-market:** NFV reduces the time required to deploy new network services. This in turn improves return on investment of new services. In addition, by reducing time-to-market, it lowers the risks associated with rolling out new services to meet the needs of customers and seize new market opportunities.
- **Increase flexibility:** NFV simplifies the addition of new applications and services as well as the removal of existing ones, to address the constantly changing demands and evolving business models. It supports innovation by enabling services to be delivered in software that can run on a range of industry-standard server hardware.

It should be noted that operators need to evolve their infrastructures as well as their operations/business management practices to fully benefit from NVE. The virtualized telecommunications network infrastructure requires more stringent reliability, availability, and latency in comparison to the cloud as it is currently used in the IT world [4].

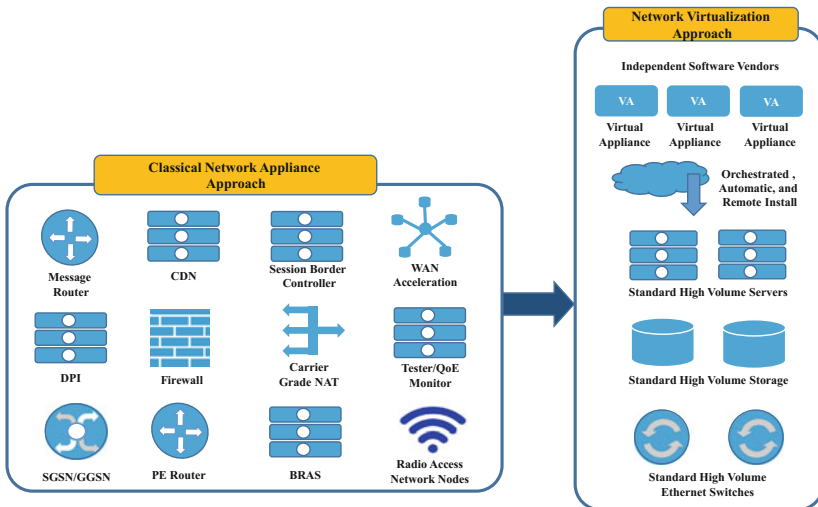


Fig. 2.4 The ETSI vision of network functions virtualization [3]

2.4.1 What to Virtualize?

In the mobile network, i.e., evolved packet core (EPC), IP multimedia subsystem (IMS), and RAN, we can, for example, virtualize mobility management entity (MME), serving gateway (SGW), packet data networks gateway (PGW), radio network controller (RNC), and base stations network functions. A virtualized EPC (vEPC) automates the authentication and management of subscribers and their services, whereas a virtualized IMS (vIMS) can deliver a portfolio of multimedia services over IP networks. The EPC and IMS network functions can be unified on the same hardware pool. Base station functions, e.g., PHY/MAC/network stacks that handle different wireless standards (2G, 3G, LTE, etc.) can share the centralized hardware resources and achieve dynamic resource allocation [53].

In general, the benefits of virtualizing network functions fall into two main categories, i.e., *cost saving* and *automation* gains and these benefits vary from system to another [4]. An assessment of the NFV benefits is shown in Fig. 2.5. While in many cases operators will benefit from virtualizing network functions, there are a few exceptions. For example, virtualizing high-performance routers or Ethernet switches is not expected to result in cost saving. Further, virtualizing products primarily focused on packet forwarding may or may not be cost effective, depending on their deployment and the ratio of control versus data plane traffics [4]. Nevertheless, even when there are no cost savings, virtualization could be justified, at times, by automation gain.

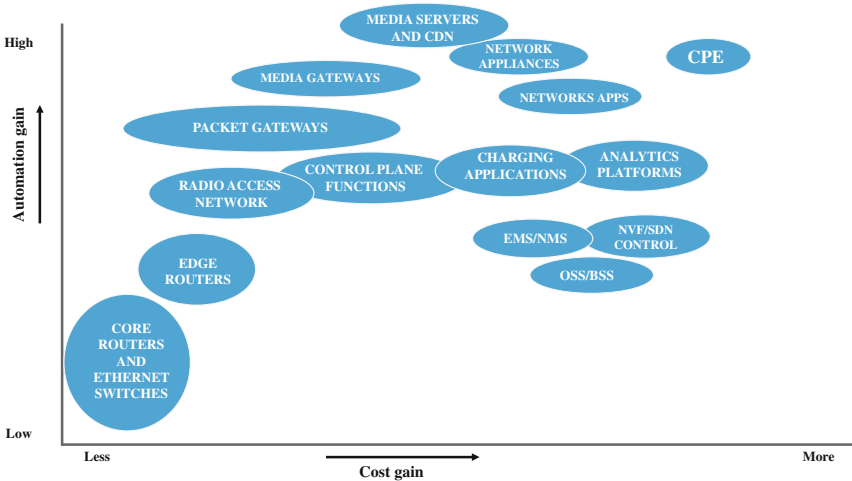


Fig. 2.5 Network functions virtualization: an assessment of the benefits [4]

2.5 Wireless Virtualization

As a natural extension of wired network virtualization, wireless networks virtualization is motivated by the observed benefits of that in wired networks. However, while virtualization of wired networks and computing systems has become a trend, much less virtualization has occurred in infrastructure based wireless networks [55]. Yet, the idea of virtualizing wireless access has recently attracted substantial attention in both academia and industry. It is one of the frontier research areas in computer science [39, 55, 56].

Wireless virtualization may refer to wireless access virtualization, wireless infrastructure virtualization, wireless network virtualization, or even mobile network virtualization [39, 56]. It is about the abstraction and sharing of wireless resources and wireless network devices among multiple users while keeping them isolated. Wireless resources may include low-level PHY resources (e.g., frequency, time, and space) or wireless equipment (e.g., a base station (BS)⁷), a network device (e.g., a router), a network, or a client hardware (e.g., wireless NIC). Thus, similar to wired network virtualization, wireless network virtualization software may reproduce logical channel and logical RAN (L1) in addition to logical switches and logical routers (L2–L3).

The motivations for virtualizing wireless networks are very similar, but not limited, to those of wired networks. First, as an extension of wired network virtualization, wireless virtualization can potentially enable separation of traffic to increase flexibility (e.g., in terms of QoS), improve security, and facilitate manageability of networks. Powerful network management mechanisms are particularly important in emerging heterogeneous networks. Second, it has a great potential to increase the utilization of wireless networks. This is important from both infrastructure and spectrum virtualization points of view. The former opens up the doors for the concept of infrastructure as a service (IaaS) so that one operator can use its own or other operators underutilized equipment (e.g., BSs on the outskirts) in the congested sites, for example in downtown. Spectrum virtualization can also provide better utilization; it may even bring more gain and is more valuable as spectrum is a scarce resource. Third, by decoupling the logical and physical infrastructures, wireless virtualization promotes mobile virtual network operators (MVNOs⁸). This allows decoupling operators from the cost of infrastructure ownership (capital and operation expenditures). Fourth, wireless virtualization provides easier migration to newer products and will likely support the emergence of new services. Last but not the least, it is a key enabler for cloud radio access network, which is expected to help operators reduce TCO and become greener.

⁷A single physical BS can be abstracted to support multiple mobile operators and allow individual control of each RAN by having a separate vBS configured for each operator.

⁸MVNOs [57] are a new breed of wireless network operators who may not own the wireless infrastructure or spectrum, but give a virtual appearance of owning a wireless network. Basically, MVNOs resell the services of big operators, usually lower prices and with more flexible plans. Virgin Mobile is an example for MVNO.

Depending on the type of the resources being virtualized and the objective of virtualization, three different generic frameworks can be identified for wireless virtualization [56]:

1. **Flow-based virtualization** deals with the isolation, scheduling, management and service differentiation between *traffic flows*, streams of data sharing a common signature. It is inspired by the flow-based SDN and network virtualization but in the realm of wireless networks and technologies. Thus, it requires wireless-specific functionalities such as the radio resource blocks scheduler to support quality of service (QoS) and service-level agreement (SLA) over the traffic flows.
2. **Protocol-based virtualization** allows to isolate, customize, and manage multiple wireless protocol stacks on a single radio hardware, which is not possible in flow-based virtualization. This means that MAC and PHY resources are being virtualized. Consequently, each tenant can have their own MAC and PHY configuration parameters while such a differentiation is not possible in a flow-based virtualization. The wireless network interface card (NIC)⁹ virtualization [60, 61] where IEEE 802.11 is virtualized by means of the 802.11 wireless NIC, falls into this category.
3. **RF front end and spectrum-based virtualization** is the deepest level of virtualization which focuses on the abstraction and dynamic allocation of the spectrum. Also, it decouples the RF frontend from the protocols and allows a single front end to be used by multiple virtual nodes or a single user to use multiple virtual frontends. The spectrum allocation in the spectrum-based virtualization differs from that of the flow-based virtualization for its broader scope and potential to use noncontiguous bands as well as the spectrum allocated to different standards.

As noted, the depth of virtualization is different in these three frameworks and they are complementary to each other. From an implementation perspective, the flow-based virtualization is the most feasible approach with immediate benefits. It connects virtual resources and provides a more flexible and efficient traffic management. In all three cases, a flow-based virtualization is required to integrate the data. However, in the flow-based approach, the depth of virtualization is not sufficient for more advanced wireless communication techniques, such as the *coordinated multipoint* transmission and reception [56, 62, 63].

As a potential enabler for future radio access network, wireless virtualization is gaining increasing attention. However, virtualization of wireless networks, especially efficient spectrum virtualization is far more complicated than that of a wired network. It faces some unique challenges that are not seen in wired networks and data centers. Virtualization of the wireless link is the biggest challenge in this domain [56, 64, 65]. Some other key issues in wireless virtualization are:

⁹By means of a wireless NIC, which is basically a Wi-Fi card, a computer workstation can be configured to act as an 802.11 access point. As a result, 802.11 virtualization techniques can be applied to the 802.11 wireless NIC. Virtualization of WLAN, known as VirtualWi-Fi (previously MultiNet [58, 59]) is a relatively old technology. It abstracts a single WLAN card as multiple virtual WLAN cards, each to connect to a different wireless network. Therefore, it allows a user to simultaneously connect his machine to multiple wireless networks using a single WLAN card.

- **Isolation:** *Isolation* is necessary to guarantee that each operator can make independent decision on their resources [66]. Also, since resources are shared in a virtualized environment, there must be effective techniques for ensuring that the resource usage of one user has little impact on others. In wired networks, this may only occur when every user is not provided with distinct resources, mainly due to resources insufficiency. *Overprovisioning* can solve the issue in such cases. It is not, however, a viable solution in wireless virtualization because *spectrum*, the key wireless resource, is scarce. To fulfill such requirements, sophisticated dynamic resource partitioning and sharing models are required.
- **Network management.** Wireless networks are composed of various radio access technologies (RATs), e.g., 3G, 4G, and Wi-Fi. Similarly, a single wireless device is capable of accessing to multi-RAT. In such a multi-RAT environment, resource sharing is not straightforward. In contrast to network virtualization technologies which are mainly based on Ethernet, wireless virtualization must penetrate deeper into the MAC and PHY layers. Further, even in a single RAT environment, slicing and sharing is not easy because wireless channels are very dynamic in nature and an efficient slicing may require dynamic or cognitive spectrum sharing methods [67]. Hence, *dynamic network virtualization* algorithms must be considered.
- **Interference:** Wireless networks are highly prone to interference and their performance is limited by that. Interference is out there, particularly, in dense, urban area. This must be considered in slicing radio resources since it is not easy to isolate and disjoint subspaces. Especially, in a multi-RAT environment, if different spectrum bands of various RATs are shared and abstracted together, interference becomes even a bigger issue because interference between different RAT needs to be taken into account too. For example, a slice from WiFi unlicensed spectrum could be assigned to an LTE user, causing unforeseen interference between LTE and WiFi networks.
- **Latency:** [66] Current wireless standards impose very strict latency, in order to meet real-time applications requirement [68]. This mandate 5–15 ms round-trip latency in layer 1 and layer 2 of today's wireless standards and will be more stringent in the next generation (5G) [69].

There are also other concerns like synchronization, jitter [70], and security [71].

2.5.1 State of the Art in Wireless Virtualization

Technical advances will be discussed in Sect. 7.6. Here we consider the state of research in this field.

2.6 Cloud Computing

Cloud computing refers to delivering computing resource as a service over the Internet, on a pay-as-you-go pricing. This type of computing relies on sharing a pool of physical and/or virtual resources, rather than deploying local or personal hardware and software. The name “cloud” was inspired by the cloud symbol that has often used to represent the Internet in diagrams. Thanks to cloud computing, wherever you go your data goes with you.¹⁰ Today, many large and small businesses use cloud computing, either directly or indirectly. The big players in the cloud space are: Amazon (AWS), Microsoft (Azure), Google (Google Cloud Platform), and Rackspace (OpenStack).

But, what explains the wide use of cloud computing among businesses? Costs reduction is probably the main driver. Cloud computing helps businesses reduce overall IT costs in multiple ways. First, cloud providers enjoy massive *economies of scale*. Effective use of physical resources due to *statistical multiplexing* brings prices lower, 5–7 times [72]. Then, multiple pricing models, especially, pay-per-use model, allow customers to optimize costs.¹¹ Cloud computing brings down IT labor costs and gives access to a full-featured platform at a fraction of the cost of traditional infrastructure. Universal access is another advantage of cloud computing. It allows remote employees to access applications and work via the Internet. Other important benefits include a choice of applications, flexible capacity, up to date software, potential for greener communication, and speed and agility. With flexible capacity, the organizations need not be concerned about over/under-provisioning for a service. When there is a load surge they can enjoy the infinite computing capacity on demand, and get results as quickly as their program scales, since there is no price difference in using 1000 servers for an hour or one server for 1000h [72].

2.6.1 Cloud Services Models

Broadly speaking, public cloud services are divided into three categories: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). In general, X as a Service (XaaS) is a collective term used to refer to any

¹⁰Today, many people actually use cloud even before they knew it. The photos that you store on your social networking sites (e.g., Facebook) or any kind of file you store and view in online file storage sites (Dropbox, Google Drive, etc.) are stored on their servers, which can be accessed from anywhere by simply logging in with your account information. In addition, you may have used or heard about Google Docs, where you can create, store, and share documents (Word) and spreadsheets (Excel) on their server, once you have a Gmail id. It is also the same business model for emails services (Gmail, Yahoo mail, etc.) as you can log in to access your emails anywhere you want.

¹¹While most major cloud service providers such as Azure, AWS, and Rackspace have an hourly usage pricing model, since March 2014 Google Compute Engine has started providing a per-minute pricing model.

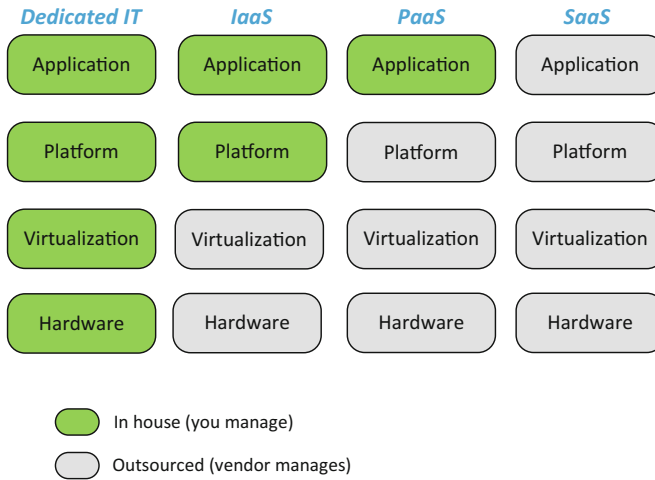


Fig. 2.6 Dedicated hosting versus cloud computing (purchasing IaaS, PaaS, and SaaS)

services that are delivered over the Internet, rather than locally. XaaS presents the essence of cloud computing and new variants of XaaS emerge regularly.¹² Yet, the three basic models (IaaS, PaaS, and SaaS) suffice for a proper understanding of cloud computing. This three service models form a service growth model, from IaaS through PaaS to SaaS, as illustrated in Fig. 2.6, in which the following layers can be identified [73]:

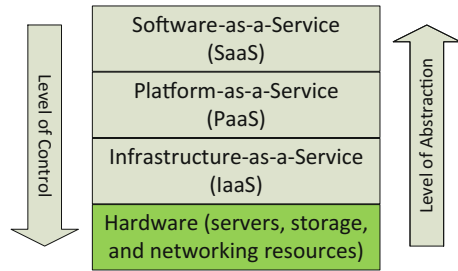
- **Application** denotes the software for the customer.
- **Platform** includes runtime environment (e.g., .NET, PHP), middleware, and operating system in which software is run.
- **Virtualization** refers to the virtualization software (hypervisor) which creates multiple virtual environments based on the physical hardware.
- **Hardware** is the equipment (servers, storage, and network resources).

As can be seen in Fig. 2.6 and more clearly in Fig. 2.7, the first growth phase is the use of IaaS. IaaS, PaaS, and SaaS are different logical layers in the stack, as visualized in Fig. 2.7. The level of abstraction/control increases as we move up/down the stack. Here is a brief explanation of each service model.

- **Infrastructure as a Service (IaaS):** In this case, computing resources (compute, storage, and network) are exposed as a capability and the clients put together their own infrastructure. For example, they decide on the operating system, the amount of storage, and the configuration of network components such as firewalls. The clients do not own, manage, or control the underlying infrastructure; instead, they rent it, as a service. As can be seen in Fig. 2.6, the hardware and virtualization

¹²Other examples of XaaS are Storage as a Service (SaaS), Desktop as a Service (DaaS), Network as a Service (NaaS), and Monitoring as a Service (MaaS).

Fig. 2.7 Different service models or layers in the cloud stack



move to the cloud. This eliminates the need for customers to set up and maintain their own physical resources. Service provider supplies virtual hardware resources (e.g., CPU, memory, storage, load balancer, virtual LANs, etc.). An example is Amazon elastic cloud compute (EC2) [74], which provides resizable compute capacity along with the needs of the customers. The pay-per-use model makes it possible to not have to make more expenditure than is strictly necessary.

- **Platform as a Service (PaaS):** In this solution, programming platforms and developing tools (such as Java, .NET, or Python) and/or building blocks and APIs for building and running applications in the cloud are provided as a capability. The customer has control over the applications and some of the configuration of the platform environment but not over the infrastructure; this is the main difference between PaaS and IaaS. Hence, unlike IaaS where users select their operating system, application software, server size, etc., and maintain complete responsibility for the maintenance of the system, with PaaS operating system updates, versions, and patches are controlled and implemented by the vendor.

Facebook is probably the most well-known PaaS. Web hosting is another example of PaaS, where web hosting provider provides an environment with a programming language such as PHP and database options in addition to hypertext transfer protocol (HTTP) which allow a personal website to be developed. Some of the biggest names in PaaS include Amazon Elastic Cloud Computing, Microsoft Azure, and Google App Engine, and Force.com [75, Chap. 8].

- **Software as a Service (SaaS):** With SaaS, the customer uses applications, both general (such as word processing, email, and spreadsheet) and specialized (such as customer relationship management and enterprise resource management) that are running on cloud infrastructure. The applications are accessible to the customers, at any time, from any location, and with any device, through a simple interface such as a web browser. As it is shown in Fig. 2.6, all layers are outsourced in SaaS. This is the ultimate level of abstraction and the consumer only needs to focus on administering users to the system. Then again, the users can influence configuration only in a limited manner, e.g., the language setups and look-and-feel settings [73].

2.6.2 Types of Clouds

Originally synonymous with public clouds, today cloud computing breaks down into three primary forms: *public*, *private*, and *hybrid* clouds.¹³ Each type has its own use cases and comes with its advantages and disadvantages.

Public cloud is the most recognizable form of cloud computing to many consumers. In a public cloud, resources are provided as a service in a virtualized environment, constructed using a pool of shared physical resources, and accessible over the Internet, typically on a pay-as-you-use model. These clouds are more suited to companies that need to test and develop application code and bring a service to market quickly, need incremental capacity, have less regulatory hurdles to overcome, are doing collaborative projects, or are looking to outsource part of their IT requirements. Despite their proliferation, a number of concerns have arisen about public clouds, including security, privacy, and interoperability. What is more, when internal computing resources are already available, exclusive use of public clouds means wasting prior investments. For these reasons, private and hybrid clouds have emerged, to make the environments secure and affordable.

Private clouds, in a sense, can be defined in contrast to public clouds. While a public cloud provides services to multiple clients, a private cloud, as the name suggests, ring-fence the pool of resources, creating a distinct cloud platform that can be accessed only by a single organization. Hence, in a private cloud, services and infrastructure are maintained on a private network. Private clouds offer the highest level of security and control. On the other hand, they require the organization to purchase and maintain its own infrastructure and software, which reduces the cost efficiency. Besides, they require a high level of engagement from both management and IT departments to virtualize the business environment. Such a cloud is suited to businesses that have highly critical applications, must comply with strict regulations, or must conform to strict security and data privacy issues.

A hybrid cloud comprises both private and public cloud services. Hence, it is suited to companies that want the ability to move between them to get the best of both the worlds. For example, an organization may run applications primarily on a private cloud but rely on a public cloud to accommodate spikes in usage. Likewise, an organization can maximize efficiency by employing public cloud services for nonsensitive operations while relying on a private cloud only when it is necessary. Meanwhile, they need to ensure that all platforms are seamlessly integrated. Hybrid clouds are particularly well suited for E-commerce since their sites must respond to fluctuating traffic on a daily and seasonal basis. On the downside, the organization has to keep track of multiple different security platforms and ensure that they can communicate with each other. Regardless of its drawbacks, the hybrid cloud appears to be the best option for many organizations.

¹³Some add a fourth type of cloud, called *community cloud* [75]. It refers to an infrastructure that is shared by multiple organizations and supports a specific community. The healthcare industry is an example of an industry that is employing the community cloud concept.

Table 2.4 Cloud computing: benefits and risks

Cloud type	Benefits	Drawbacks
Public	• Low investment in the short run (pay-as-you-use)	• Security: multi-tenancy and transfers over the Internet [76]
	• Highly scalable	• Privacy and reliability [76]
	• Quicker service to market	
Private	• More control and reliability	• Higher cost: heavy investment in hardware, administration and maintenance
	• Higher security	• Must comply with strict regulations
	• Higher performance	
Hybrid	• Operational flexibility: can leverage both public and private cloud	• Security, privacy, and integrity concerns
	• Scalability: run bursty workloads on the public cloud	
	• Cost effective	

In Table 2.4, we enlist the main benefits and risks associated with each type of clouds. Understandably, security is one of the main issues in cloud computing. There are many obstacles as well as opportunities for cloud computing. Availability and security are among the main concerns [72, 77].

2.6.3 Virtualization Versus Cloud Computing

By now the reader should have realized the connection between virtualization and cloud computing. Broadly speaking, these two technologies share a common bond: they are both meant to increase efficiencies and reduce costs. They are quite different though. Virtualization is one of the elements that forms cloud computing. It is the software that manipulates hardware, while cloud computing is a service that results from that manipulation [78].

Observing that cloud computing is built on a virtualized infrastructure, one can deduct that if an organization have already invested in virtualization, they may bring in cloud to further increase the computing efficiency. Then, the cloud could work on top of the current virtualized infrastructure; it also helps in the delivery of current network as a service. Put differently, cloud computing makes use of virtualized resources at a different level, where the resources can be accessed as a service, and in an on-demand manner. Conversely, any organization considering adoption of a private cloud must work on virtualization, too.

Organizations can improve the computing resources efficiency through virtualization; however, they cannot get rid of provisioning. An administrator is still required to provision the virtual machines for the users. Cloud computing removes the need for manual provisioning. It offers a new way for IT services delivery by providing a customer interface to automated, self-service catalogs of standard services, and by using autoscaling to respond to increase or decrease in users demand [79].

2.7 Summary

Network overlay (using encapsulation and tunneling techniques) is one way to implement virtual networks. This approach is network agnostic, but it cannot reserve resources such as bandwidth. In addition, it does not guarantee service quality, hence it can result in degraded application performance.

Using software-defined networking (SDN) is another way to implement network virtualization. For example, one can define the virtual networks in the flow tables of the SDN switches. The SDN approach to network virtualization overcomes the limitations of the above approach. It also brings the ability to do more granular traffic routing and to gather more intelligence about the infrastructure.

Cloud Mobile Networks

From RAN to EPC

Vaezi, M.; Zhang, Y.

2017, XVII, 117 p. 34 illus., Hardcover

ISBN: 978-3-319-54495-3