Towards Future Mobile Network: Harnessing the Transformative Nature of NFV and SDN

DK Lee, Jong-Han Park, and Jin-Hyo Park
SK Telecom

Abstract

Mobile operators today face yet another critical challenge as technology lifecycle becomes increasingly short and also as heterogeneous and complex network becomes exceedingly expensive and difficult to manage. With extremely competitive market and demanding users, the overall revenue structure is expected to get worse.

A network architecture based on software-defined networking (SDN) and virtualization techniques gives operators greater opportunity to build cost-effective and efficient alternative to the legacy. In this work we review our Carrier Cloud as a future mobile network infrastructure that exploits both SDN and NFV in order to increase the operator agility, reduce the cost, and even disrupt the vendor landscape. This new architecture will not be fully adopted by the conservative operators at once. Technological hurdles have to be overcome, and a clear understanding of operational differences must be preceded.

Ⅰ. Introduction

About 30 years ago when the mobile network is first introduced in Korea, the network and its architecture was mainly designed for voice-oriented services, which initially looked like a mobile counterpart of the landline telephone service. As the communication paradigm radically shifted from voice-oriented services to data-oriented services overtime, the mobile network also had to reflect this transition in its architecture through evolution from the 2nd generation mobile network (2G) to the 3rd generation mobile network (3G) and recently to the Long Term Evolution (LTE, 4G) architecture. Today in LTE-Advanced (LTE-A) with Carrier Aggregation technology, the maximum bandwidth that the network can provide in the commercialized environment is up to 150 Mbps [4], which is more than 4,600 times larger bandwidth than what the initial mobile data network with GPRS could support.

On the flip side of this seemingly inspiring advancement in technology are the mobile network providers, who have been investing tremendous time, money, and efforts to first understand the technology in time and finally deploy it in the commercialized network. This process of verifying and deploying the new technologies and architectures have become increasingly cumbersome as the number of technologies already up and running in the network is quite high and also the lifespan of a given technology is also becoming shorter. This scalability issue in cost and management eventually led the providers to seek a more effective approach in accommodating new technologies and architectures in their network.

Recently, a number of new technologies emerged from both academia and industry that may have a profound impact (if not already) to mobile service providers. In this paper, we introduce two recent technologies named Network Functions Virtualization (NFV) [1,2] and Software Defined Networking (SDN) [3,7], why we believe they are essential in the future mobile network, how we think these two technologies can help us define our future mobile network by briefly describing our future mobile network vision, and finally the benefits of the transformative network built on top of the two technologies, which we envision.

This paper is organized as follows. First we provide a quick review of the mobile network and its past evolutions and concerns during the past evolutions (Section II). We then talk about the recent technology trends, introducing the two main technologies: NFV and SDN and describe our vision network in more detail and the incremental steps that we believe are necessary before finally implementing the vision network (Section III). In Section IV, we conclude.

Ⅱ. Mobile network, its past evolutions, and concerns

Ever since the advent of the packet data network (in both wired and wireless domain), there have always been contents and killer applications, using up all the available bandwidth that the underlying network architecture and technology could provide. This ever-increasing demand for higher traffic volume and speed posed many challenges on the mobile network architecture and drove many revolutionary changes in the mobile network architecture, including radio access network, overall traffic paths, value-add services, and etc. In this section, we quickly review the history of mobile network, how it evolved so far, and the approaches used to accommodate the evolutions. We then describe the challenges faced by today's mobile network and why the traditional approaches may no longer be well-suited for the future evolutions and changes.

1. Mobile network and its evolutions

Starting with GPRS (48Kbps), the mobility data network has seen rapid advancements to new major technologies and architectures such as EDGE (237Kbps), UMTS (3.6Mbps), HSPA 7.2 (7.2Mbps), HSPA+ (21.6Mbps), and recently LTE (150Mbps).

In the past, the evolution towards the succeeding technology had been relatively more manageable than what it is today. There are many reasons, and we list some of them in this section. First, the mobility data speed supported at the radio access layer was still quite limited and had been the bottleneck, compared to the speed at which the wired domain of the network operate. Therefore the main area of focus was mostly limited to the radio access technology, and this narrow scope of improvement had relatively small impact on the overall architecture. Second, there were not many legacy systems already running in the production network, which required the full regression testing and certification for interworking. Back then when introducing a new technology to the network, the technology is either completely new (and therefore there is no need at all for checking the compatibility or interworking with the legacy technology) or there were not many legacy services already deployed and running in the network that must be considered for interworking. Third, the lifespan of a new technology, where the given technology is used and recognized as the major technology being used by the majority of users, was relatively longer in the past. In other words, although adopting a new technology incurred some costs and investments, the frequency of investments was not so often, and therefore, it was still considered manageable and the Return-Of-Investment (ROI) was still considered high enough.

The mobile network so far has evolved at a very fast pace. Yet recent studies report that, by 2020, the overall mobility traffic volume may be up to 1,000 times more than what we see today [5], and this trend of traffic-driven and fast-paced network evolutions does not seem to stop anytime soon.

2. Tight coupling between technology, network, and infrastructure

Although the mobile service providers have been diligently and aggressively catching up with the newest and greatest technologies and architectures successfully to satisfy the user requirements and demands, a number of symptoms were observed in the evolution process, which led to a realization that the network cannot simply continue evolving by simply repeating the traditional approach. Some of the symptoms are described as follows:

First, the Total Cost of Ownership (TCO) of the network has gradually become substantial over time, especially in terms of capital and operations expenditures as well as on the network management complexity. The main cause for this significant TCO is the price and heterogeneity of the specialized appliances that implement the technology. A mobile network today typically has to operate multiple layers of network service (e.g., 2G, 3G, WiFi, LTE, WiMax, etc.) simultaneously to support the legacy users. This is because, in general, a mobile service provider cannot simply retire a legacy service due to the fact that there are still users who use the legacy service. In this multi-layered network that runs numerous technologies all at once, a particular technology is oftentimes vendor-specific where a vendor uses its own specialized hardware appliances to implement the given technology. This leads to the network environment that consists of a wide variety of physical appliances from various vendors that implement the multiple layers of services. Since there are multiple network services and a wide variety of specialized hardware appliances that must simultaneously managed, it is natural that the overall cost of purchasing the appliances as well as managing the network will be exceedingly expensive and complex. In the near future, this TCO is likely to start increasing even at a faster rate with the new technologies to be implemented and will exacerbate the negative influence on the Average Revenue Per User (ARPU). Overall, the existing architecture has the diminishing return on the profit as ARPU gradually decreases and is too costly to maintain.

Second, the Time-To-Market (TTM) has increased to the level that is now no longer acceptable, as the network becomes increasingly sluggish over time after many heterogeneous network elements are introduced to the network. This is a serious concern for today's mobile service providers that must be addressed; a prolonged TTM of a particular service deployment may potentially mean a massive subscriber transition from one service provider to another service provider in today's highly competitive market. Another reason that the on-time TTM has recently become an important and differentiating factor for the quality of a given mobile service providers is that the specifications that define the recent technologies are very tight and leave only little room for the service provider to add its own specialized and differentiated services. This means that today the mobile service providers must not only have the ability to rollout a service with correctness and performance, but also have to configure (or program) the network flexibly in agile, safe, and secure manner.

Lastly, other deficiencies were started to get exposed as the network is pushed to its limits, such as limited flexibility and inefficient use of energy.

There is no doubt that the mobile network is a huge success. It has seen a rapid growth, and this explosive growth is not likely to stop anytime soon, driving continued evolutions in the future. However, it is becoming clearer that the existing architecture and evolution process do not well accommodate the rapidly changing technology advancements.

3. Decoupling technology and network from underlying infrastructure

Today's network structure does not support efficient evolutions towards new technologies and architectures. As briefly described above, an evolution towards new technologies can be too costly and will no longer be tractable in the future. This is because the entire infrastructure package such as the technology, network, and infrastructure must be tested, certified, and deployed altogether as a whole for every technology rollout.

We make an observation that one of the main causes for this intractable cost is the tight coupling between the technologies, the network configurations, and the underlying infrastructures. An ideal network structure that can easily accommodate new technologies would have well-modularized and separated layers of each component. This modularity allows the evolutions to happen at a minimal cost since each layer can independently be upgraded without having any impacts on the other layers.

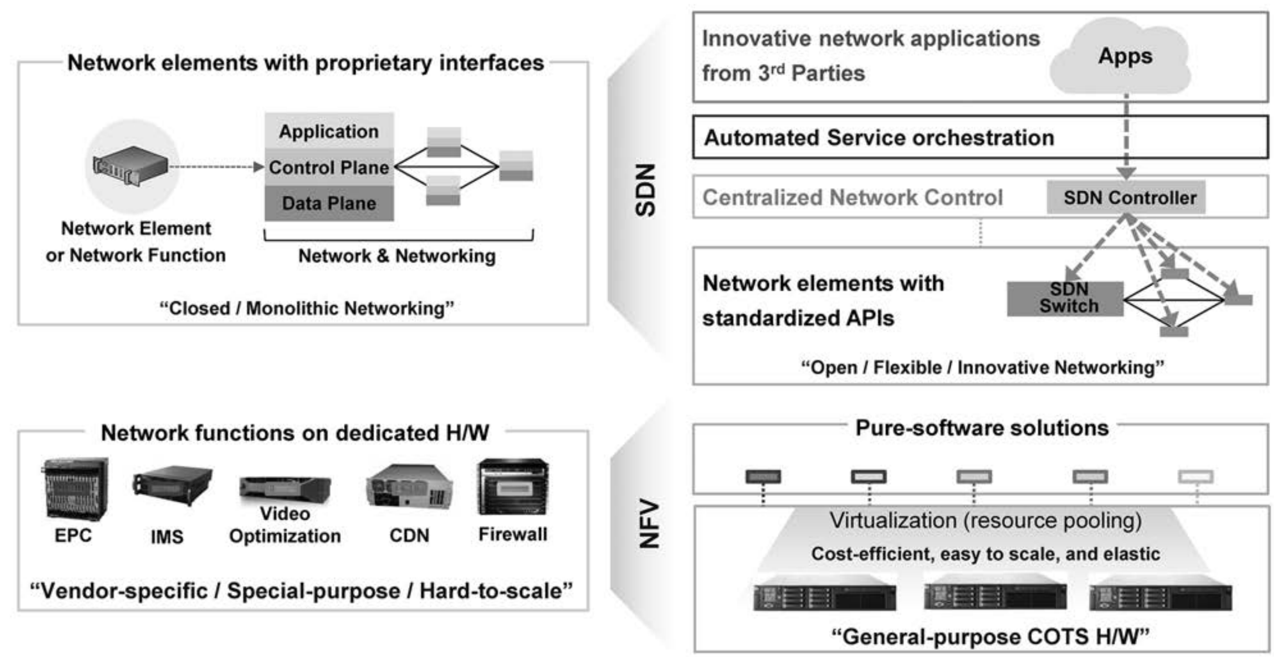

Figure 1. NFV/SDN and how they transform the existing network

For example, a mobile service provider should be able to deploy a new service (or corresponding software functions) on top of the existing infrastructure without any investments at all for the underlying infrastructure if the service simply re-uses the underlying network and infrastructure. Furthermore, this modularity makes the problems at each layer small enough and easy to solve, promoting additional innovations and enhancements to happen at each layer.

Given the observations that the traditional approaches for network evolutions are not amenable for accommodating the future network evolutions, we describe, in the following section, a disruptive yet promising shift in network management and operation paradigm, led by two key technologies named Network Functions Virtualization (NFV) and Software Defined Networking (SDN).

Ⅲ. Operator Vision: Harnessing the Transformative Nature of NFV and SDN

Our vision towards future mobile network exploits NFV, SDN, and cloud computing, which are widely adopted and already proven in the field of IT data network. In this section, we first explain the concept of NFV and SDN which we heavily utilize to draw our vision network. We then describe at a high level our vision network, and explain in more detail why we have drawn our vision in this particular manner.

1. Network Functions Virtualization (NFV)

The NFV concept is to virtualize all network appliances such as packet gateways or firewalls, and bring them on top of the cloud computing environment consists of off-the-shelf general hardware, realizing the cost-effective and highly efficient network infrastructure. Although the concept of virtualization may already be a common and obvious approach for IT functions such as web servers, the concept of virtualizing Telco functions initially was quite shocking and questionable because the Telco functions typically have very strict SLAs and also heavily use specialized hardware for a very high reliability (e.g., five 9's).

An example of how NFV can change the existing network is shown in Figure 1. As depicted, NFV decouples the network functions from the underlying physical appliances and run them on standardized, common, and cheap servers.

2. Software Defined Networking (SDN)

As a new networking paradigm, SDN decouples the control plane from the data plane and defines open APIs between them, creating open innovation platform for a more flexible and innovative networking. One of the key components in SDN is the centralized controller that is used to fully automate the network management with the global network view.

An example of how SDN can change the existing network is shown in Figure 1. With SDN, the application, control plane, and data plane are decoupled and now run inside separated layers.

3. Why change? A Motivating Example

We start with reviewing the real-world cases in our current mobility network operation that we feel are painstaking and not well handled. In the core network today, various types of core network service nodes are deployed between the packet core and the Internet (or provider edge router). Those middle-boxes are performing diverse network optimizations to deliver better quality-of-experience to customers. Some examples of these middle-boxes are mobile traffic classification, mobile video optimization, mobile web optimization, wireless TCP acceleration, or multimedia compression. These network services are mostly located at or around the 3GPP SGi or Gi interface. As the mobile network services have become more sophisticated, the complexity of network and operations also grow quickly with new middle-boxes. Those middle-boxes are implemented as a separate, dedicated hardware and placed in the network manually. In general, the traffic traverses every single middle-box sequentially or the same service chain, unless there is a pre-defined exceptional policy. The long and single service chain increases the end-to-end service latency, and more importantly, indispensable capital investment must be made and is infeasible in the face of the mobile data explosion; when the data traffic has doubled we need to upgrade doubly all middle-boxes along the path. The other problem is about the network manageability. We found that the network does not work as we expect as we add more middle-boxes with different functions and characteristics. For example, we may deploy packet throttling node for the context-aware traffic engineering and management. At the same time, we can also deploy WAN accelerator to optimize the wireless TCP performance. When we use both of two middle-boxes that are seemingly performing the opposite functions at the same time, we found the resulting performance is simply not predictable. In order to guarantee the network function correctness, it is critical to ensure the traffic passes through the correct sequence of middle-boxes. Some limited amount of traffic could be steered selectively according to the subscriber or service type currently. The traditional policy-based routing control is, however, static and inefficient, and furthermore, it suffers from manual and error-prone processes. The policy should be crafted carefully ensuring the correctness of the routing path; it entails significant manual efforts and operator expertise.

The example which we described in this section illustrates the complexity of today's mobile network and also clearly motivates the need for the transition to the future mobile network.

4. Renovating the Legacy

We find that both NFV and SDN complement one another to help us build the promising alternative mobile network infrastructure. This envisioned network infrastructure has the potential to increase operator agility, reduce the cost, and even disrupt the vendor landscape. We envision that the evolution to our vision network will proceed incrementally in small steps. The first step towards our vision network is what's commonly called as Carrier Cloud (or Telco Cloud), where the existing diverse mobile network functions such as the access and core network elements are fully virtualized and deployed remotely and automatically. (See Figure 2)

The virtualization and cloud computing technology offer a promising cost-effective and efficient alternative to

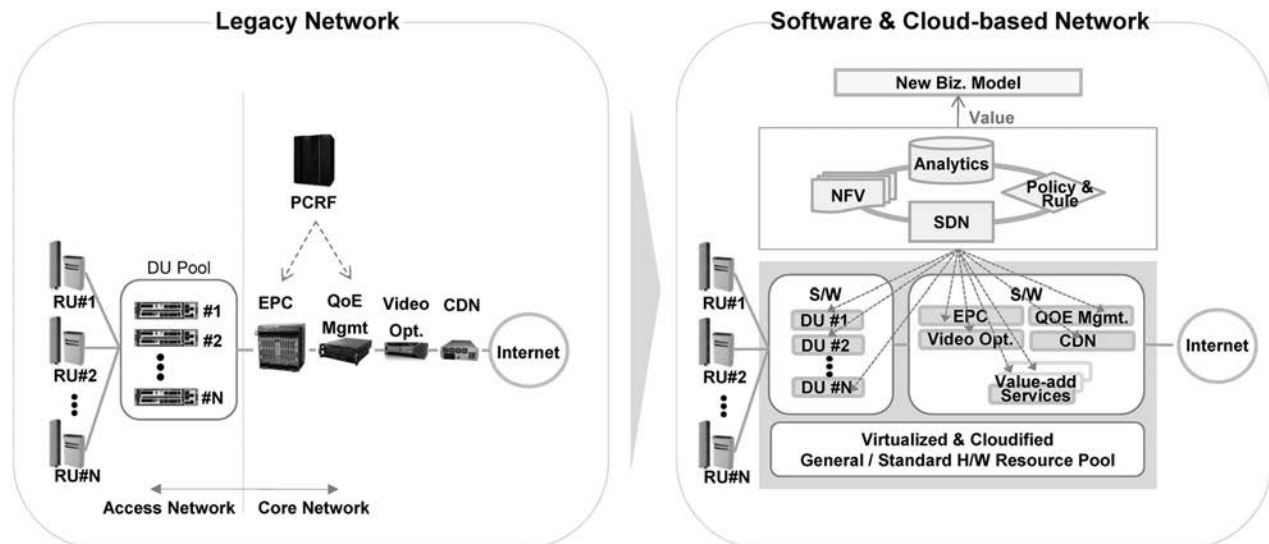

Figure 2. Migration to the cost-effective and highly efficient alternative to the legacy

the legacy infrastructure where the functions are tightly coupled with the underlying infrastructure consisting of vendor-specific and vendor-isolated physical appliances. We expect the Carrier Cloud would initially be implemented in the form of operator's distributed private cloud.

The next step is to introduce SDN to the Carrier Cloud. This is to inject dynamism into virtualized and cloud-based network infrastructure. In the Carrier Cloud, a virtualized network appliance scales itself dynamically in response to the elastic demand. If needed, the virtualized network functions may even move around in the cloud. In such case, the right network paths must be created and configured accordingly before a given function migrates, ensuring the packet streams are seamlessly steered and delivered to the right destination without disturbing the service. Rather than configuring the underlying physical network elements manually, SDN is used to automate the whole network configuring process. The benefit of SDN is not limited to automation. SDN also offers optimal network path assignment among network elements, which can best be achieved only with the centralized traffic engineering. Another benefit of SDN in the context of Carrier Cloud is that it serves as the networking infrastructure that enables new innovative network services.

One interesting example that may be implemented inside a Carrier Cloud with SDN is the dynamic service chaining [6]. Instead of providing one common sequential network path for all traffic flows, individual traffic flows can be treated differently based on the types of service or various other criteria. There may be many promising use cases of dynamic service chaining. Another use case makes expensive equipment like QoS optimization service and has it selectively enabled only in desired situations such as, when a given subscriber suddenly stays at a highly congested area and is predicted to suffer from the performance degradation. Another use case looks at the traffic characteristics as the differentiating criterion. That is, we may also be able to provide different types of data services according to the traffic characteristics. The 'elephant traffic' such as software-upgrades will bypass unnecessary core network services. The delay-insensitive high-volume traffic such as cloud synchronization traffic will be delivered over the delay-tolerant network service.

The final stage of our software-oriented network evolution is to add analytics-based intelligence policy and control on top of the Carrier Cloud (managed by SDN) infrastructure. There will be data collection or network intelligence extraction units (or agents) located at several strategic points in the end-to-end mobile network, such as user equipment, base stations, or core network nodes. Various types of statistics, user behaviors, and performance metrics are fed-back and collected from those agents. By analyzing the large volumes of data from various observation points, intelligent network policy for the network optimization or monetization may be deduced. These big-data analytics and network policy can then be used or shared by operator applications, customer experience management services, or even by other 3rd party applications through the open APIs. The policy may also be enforced to the entire end-to-end mobile network by 3GPP PCRF (Policy Control and Rule Function) or SDN/Cloud controllers.

5. Vision Network

Our strategic direction towards software-oriented network evolution is summarized in Figure 3. We envision a value network running on top of virtualized IT infrastructure intelligently managed by a single orchestrator. We find key architectural design principles as follows:

1) All IT Infrastructures: Use the industry standard, general-purpose, and COTS Hardware. Also, exclude any cellular-specific network nodes or protocols.

2) All Virtualization: Do not access the physical resources directly. Access only virtualized resources through resource abstraction layer that hides and encapsulates any physical details from the applications.

3) Pure Software-based Assets: Re-design or re-factoring existing network functions (or network assets) as a hardware-independent software. Life-cycle (on-boarding, deployment, auditing, scaling, healing, upgrading or termination) of network assets is remotely managed by orchestration software.

4) Centralized Orchestration: Decouple the control from the data plane. Implement fully automated analytics-based policy and control over entire mobility network.

5) Platformization: Provide network APIs and middleware in order to encourage innovation and competitiveness among 3rd parties, and realize the 'Network Platform'as a Service.

Every network asset is implemented as hardware-independent software, and run in virtualized general-purpose resources resident on cloud data centers. Along

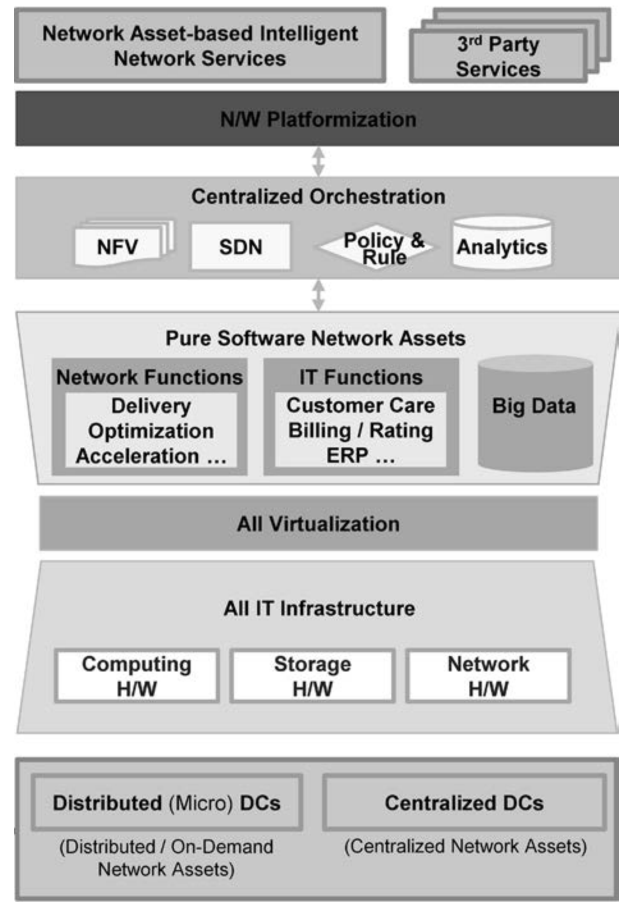

Figure 3. Future mobility network architecture

with the centralized data center, distributed or micro data centers also exist to bring distributed or on-demand network services such as edge caches or mobile multimedia optimizations. Those micro data centers might be located near base stations or resident on customer premises.

The resulting mobility network topology is simplified as data center network inter-connecting between the user equipment and Internet.

6. Mind the Gap

So far we have discussed our vision on future mobile network infrastructure. We have demonstrated that SDN, virtualization, and cloud computing are powerful and versatile tool to build cost-effective and efficient alternative to the legacy infrastructure.

The study, however, also reveals that those technologies still has some hurdles to overcome in order to be fully adopted in the Telco world. They are inherently constrained by the idiosyncrasies that stem from its origin. In this section we list those remaining challenges.

3GPP Compliancy

In SDN, the controller directs the packet forwarding devices (or data planes) how to route the flow to the destination. Before the controller makes proper decisions, it needs extra information about the mobile data flows. Few examples are the profile (e.g., data plan or location) of subscriber who owns that flow, layer-7 application characteristics, or the on-the-spot wireless network condition. Current SDN standard such as OpenFlow only recognizes L2-L4 packet matching rules which are restricted in the specification. The additional intelligences are, therefore, needed to be acquired from other key network elements in the mobile network: PCRF (Policy Control and Rule Function), DPI (Deep Packet Inspection), and OSS/BSS.

For the seamless integration with the mobile network, 3GPP-compliancy, subscriber-awareness, service-awareness and controllability should be devised. It is, however, still ambiguous that how SDN provides the interfaces (i.e., 3GPP-compatible interfaces) with other network elements (see Figure 9). For those inter-working features, it seems that the standard has gotten behind, and vendor-specific or vendor-isolated solutions are emerging, making it far from the open innovation platform that the SDN is supposed to be.

Telecom Requirements: Performance and Reliability

We also have challenges to virtualize the mobile network elements with diverse characteristics. In compassion with normal IT workloads (or network elements) such as web servers, workloads in the Telecom world have their own stringent requirements, such as five-nine (≈100%) reliability and availability, high-performance data plane, and extremely low latency. We have roughly categorized the workloads and their requirements in the Telecom world into three types as in Figure 10: (1) Pipe, (2) Control, and (3) Service.

- Type 1. Pipe: eNodeB(DU), GGSN, S/P-GW or DPI
- Type 2. Control: SON, MME, PCRF or CSCF, that controls Pipe workloads
- Type 3. Service: SMS/MMS, IMS AS or CDN that provides Telecom's value add services (Similar to IT workloads)

The Service workloads seem to have similar requirements to the normal IT workloads. For the other two types of workloads the requirements should be carefully analysed and reflected in the virtualization technology in order to guarantee proper levels of performance and reliability.

In Figure 3, we illustrate example reference architecture of the Carrier Cloud that probably gratifies Telecom Requirements. From the physical resources (i.e., compute, storage, and networking devices) to the hypervisors, virtual switches, virtual machines, and cloud management stacks, we need the right design decisions in order to guarantee the requirements.

For example, the Carrier Cloud should include the life cycle management of virtualized network elements (vNE). Deploying and monitoring virtualized network elements in the cloud should be remotely and automatically done. vNE should scale itself according to the fluctuating nature of the demand. Additionally, on-demand scaling of vNE can be triggered by OSS/BSS or by authorized users. (i.e., vNE and cloud administrator) The Carrier Cloud should provide high-availability (HA) and fault-tolerance (FT) facilities in order to deal with software and hardware failures. vNE should be allowed to be recreated after a failure through N+1 or 2N redundancy. The key functionalities for those redundancies are VM mirroring and live migration. In cases where vNE consists of sub-modules with internal hierarchy, HA and FT schemes should be provided to each sub-module, instead of recreating the entire vNE. The switch-over delay should be less than a second.

For the latency-sensitive vNE, the real-time OS and real-time hypervisors are mandatory. In order to optimize the data plan and guarantee near native hardware performance, virtual switches should be patched with a direct communication path between vNEs that detours unnecessary memory addressing or interrupt handlings.

Network Renovation Strategy and Backward-Compatibility

No one can virtualize the whole telecom network at once, but the process should be done step-by-step. For example, The SDN will be firstly adopted in the self-contained or isolated IP network such as SGi/Gi service network in the packet core. The virtualization would be started with service nodes that have relatively loose requirements on the performance and reliability, and then extended to the control and pipe nodes. The proof-of-concepts should be preceded thoroughly before the extension. The operators and vendors should work together and come up with that smooth network renovation strategy.

Ⅴ. Conclusion

In service provider perspective, the networking paradigm shift is just around the corner. The mission is to renovate our network infrastructure in a cost-effective way against the ever-growing network complexity and user demands. We see that both NFV and SDN will be at the core of future mobility network that is characterized by flexible resource use and fast implementation of innovative services.

In this work we have reviewed our Carrier Cloud as a promising alternative. For the success of SDN and virtualization in the wild, however, the viable use cases should still be developed. The operators and vendors should work together to make the technology be fully qualified according to the performance and reliability requirements in the real world.

References

[1] Network Functions Virtualization, ISG web portal, June 20, 2013.

[2] Network Functions Virtualization – Introductory White Paper, ETSI. October 22, 2012, http://portal.etsi.org/NFV/NFV_White_Paper.pdf

[3] Open Networking Foundation's definition of SDN, Open Networking Foundation. https://www.open-networking.org/sdn-resources/sdn-definition

[4] LTE-A Commercialization, SK Telecom, September 2013. http://www.skt-lte.co.kr/contents/ltea/intro.jsp

[5] Scenarios, requirements and KPIs for 5G mobile and wireless system, METIS (Mobile and wireless communications Enablers for the Twenty-twenty Information Society), April 29, 2013. https://www.metis2020.com/documents/deliverables/

[6] Zafar Ayyub Qazi, Cheng-Chun Tu, Luis Chiang, Rui Miao, Vyas Sekar, Minlan Yu, "SIMPLE-fying Middlebox Policy Enforcement Using SDN," ACM SIGCOMM, August 12-16, 2013, Hong Kong, China

[7] N. McKeown, Software Defined Networking, INFO-COM keynote talk, 2009.

약 력

이 동 기

2004년 아주대학교 컴퓨터공학과 공학사
2011년 KAIST 전산학과 공학박사
2011년~현재: SK Telecom, Manager
관심분야: Large-Scale Networked Systems, SDN, NFV

박 종 한

2000년 University of California, San Diego 공학사
2006년 University of California, Irvine 공학석사
2011년 University of California, Los Angeles 공학박사
2011년~2014년 AT&T Labs, Senior Member of Technical Staff
2014년~현재 SK Telecom, Manager
관심분야: Future Mobile Network, SDN, NFV

박 진 효

1994년 고려대학교 수학교육과 이학사
2003년 고려대학교 정보통신공학과 공학석사
2009년~2013년 SK Telecom, Access Network Lab장
2013년~현재 SK Telecom, Network기술원장
관심분야: LTE-A 진화 기술 및 차세대 이동통신 기술(B4G, SON, SDN/NFV)