Why the Adoption of NFV Mandates a Big Data Management Approach

Introduction

Communications Service Providers (CSPs) are under intense pressure. One source of this pressure is the dramatically increasing demand for some of the services they offer, such as mobile data, combined with their limited ability to increase the price they charge for these services. A second source of this pressure is the growth of Over The Top (OTT) players that leverage the networks provided by CSPs in order to compete with CSPs in offering a wide range of services such as unified communications. A third source of this pressure is the broad movement on the part of customers to shift away from CSPs' high margin services such as MPLS and adopt low margin services such as Internet access.

One of the ways that CSPs are adapting to this changing environment is by adopting NFV. NFV holds the promise of enabling CSPs to quickly and easily deploy and operate services and functions in a more cost effective manner than is currently possible. One existing function that CSPs are aggressively working to apply the key concepts of NFV to is the Evolved Packet Core (EPC), which is a framework for converged voice and data transport on a 4G LTE network. One of the specific reasons why CSPs are so interested in the cost reduction that is associated with implementing virtualize functions such as a virtualized EPC (vEPC) was discussed in a recent report. Per that report, although mobile operators around the world have spent over $800 Billion in infrastructure investment over the past ten years, revenue growth has been almost flat during the time-period.

However, the disaggregation, virtualization and automation that is associated with NFV, combined with other factors such as the distributed nature of NFV deployments and the dramatic growth in bandwidth, is creating a situation in which the traditional centralized approach to management breaks down. This white paper will demonstrate that to effectively adopt NFV, that CSPs must evolve away from their traditional approach to management to one that focuses on both pervasive data collection and a sophisticated big data approach to storing and analyzing management data.

The Promise of NFV

One of the central concepts that drives NFV is that some or all the network functions that CSPs deploy must be available as Virtualized Network Functions (VNFs) that can be easily provisioned, integrated and managed. The transition from the current hardware-centric environment to a fully virtualized environment is depicted in Figure 1.

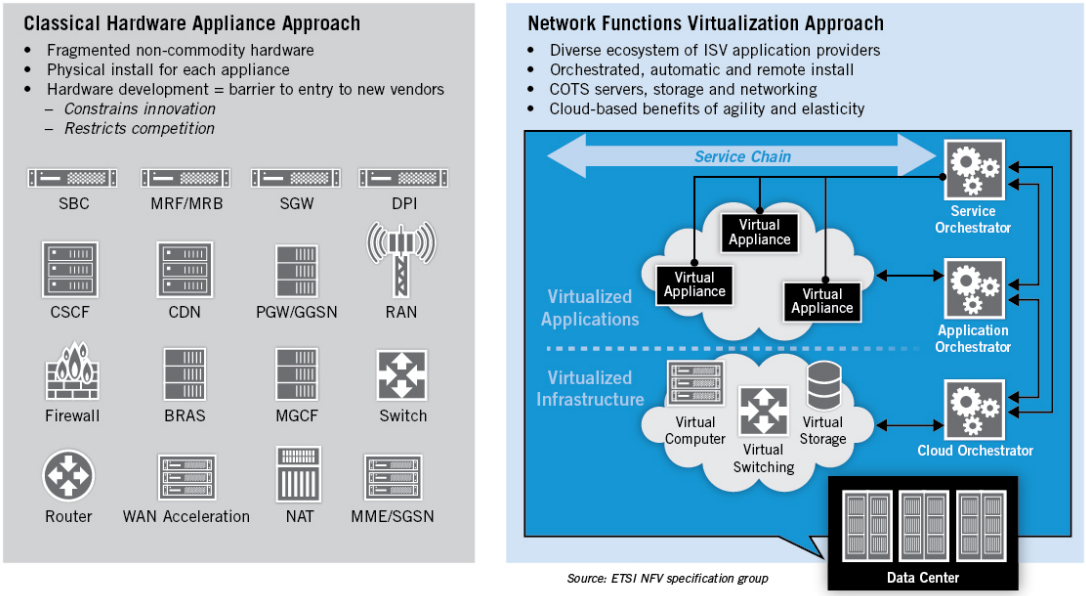

Figure 1 - The virtualization of network functionality

However, implementing NFV involves more than just virtualizing existing functionality. Some of the key characteristics of the European Telecommunications Standards Institute (ETSI) vision for NFV include:

- Achieving high performance and minimum cost. As described below, this will not be achieved by merely porting existing solutions to commodity hardware in a monolithic fashion.
- Implementing automation to enable the scalability of the solutions.
- Managing and orchestrating multiple VNFs while ensuring security from attack and misconfiguration.

Initially it was thought that VNFs should be implemented exclusively in centralized data centers. In the current environment, it is widely accepted that VNFs should be located where they are the most effective and least expensive. That means a CSP may locate VNFs in a variety of locations, including data centers, network nodes as well as on the customer premises. This approach is sometimes referred to as distributed NFV.

The interest in distributed NFV is in alignment with the broader industry movement towards edge computing. Edge computing pushes applications and data to the logical extreme of the network. One of its advantages is that it enables analytics and knowledge generation to occur at the source of the data.

In order to leverage the emergence of edge computing in a way that enables mobile operators to fully realize the promise of NFV, in late 2014 ETSI announced the creation of an Industry Specification Group (ISG) for Mobile-Edge Computing. Per that announcement "Mobile-Edge Computing provides IT and cloud-computing capabilities within the Radio Access Network (RAN) in close proximity to mobile subscribers. Located at the base station or at the Radio Network Controller, it also provides access to real-time radio and network information such as subscriber location or cell load that can be exploited by applications and services to offer context-related services. For application developers and content providers, the RAN edge offers a service environment characterized by

proximity, ultra-low latency, high-bandwidth, as well as real-time access to radio network information and location awareness. Mobile-Edge Computing allows content, services and applications to be accelerated, maintaining a customer's experience across different radio and network conditions."

In a separate announcement, ETSI stated that Mobile-Edge Computing has several use cases including:

- Video analytics
- Location services
- Internet-of-Things (IoT)
- Augmented reality
- Optimized local content distribution and
- Data caching

The Virtual Evolved Packet Core (vEPC)

As noted, the EPC is a framework for converged voice and data transport on a 4G LTE network. It provides the control, management, security and intelligence to connect many mobile edge elements, such as base stations and base station controllers. The standards for EPC operation were specified by the Third Generation Partnership Project (3GPP). Per the 3GPP the primary components of the EPC are:

HSS
The HSS (Home Subscriber Server) is a database that contains user-related and subscriber-related information. It also provides support functions in mobility management, call and session setup, user authentication and access authorization.

Serving GW
The gateways (Serving GW and PDN GW) deal with the user plane. They transport the IP data traffic between the User Equipment (UE) and the external networks. The Serving GW is the point of interconnect between the radio-side and the EPC.

PDN GW
The PDN GW is the point of interconnect between the EPC and the external IP networks, called PDNs (Packet Data Networks). The PDN GW routes packets to and from the PDNs and performs various functions such as IP address / IP prefix allocation or policy control and charging.

MME
The MME (Mobility Management Entity) deals with the control plane. It handles the signaling related to mobility and security for E-UTRAN access.

One example of how CSPs are adopting NFV is by taking a service such as the EPC, virtualizing it in a manner that also disaggregates the functionality into several VNFs and then running the VNFs on commercial off-the-shelf (COTS) hardware. This concept is demonstrated in Figure 2. In Figure 2, the MME has been virtualized (vMME) as has a combined Service GW and PDN GW (vS/PGW). In this example, each of these VNFs runs on KVM and each VNF is comprised of several sub-functions each of which runs in a Virtual Machine (VM). In the broader world of IT, when a piece of software such as a vMME is disaggregated into sub-functions, those sub-functions are often referred to as microservices. However, when focused just on NFV these sub-functions are increasingly

being referred to as being Virtual Network Function Components (VNFCs).

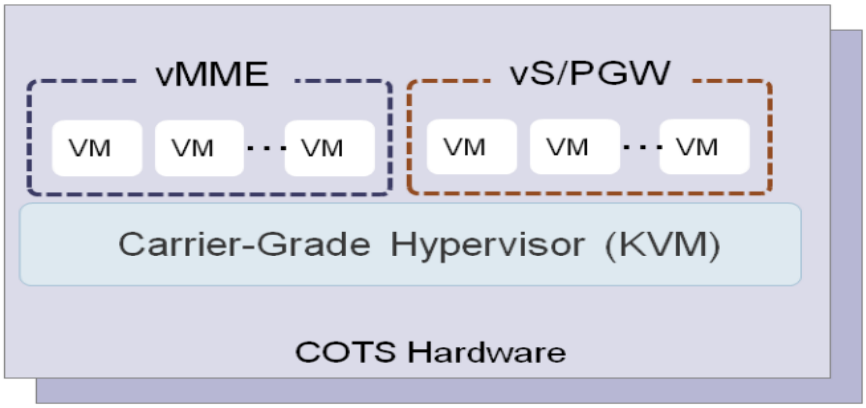

Figure 2: Representative vEPC

As previously mentioned, one of the ways that NFV provides value is that it reduces cost and it makes CSPs more responsive to shifting demand. For example, if a CSP had deployed a traditional, hardware-based, monolithic EPC, then if one of the EPC's primary components, such as the MME, runs out of resources, the CSP would have to go through a lengthy process to acquire and deploy a new instance of the entire EPC. With a vEPC, the CSP can automatically scale up the capacity of just the component that ran out of resources.

Another way that NFV provides value is that it reduces risk because it enables CSPs to establish a checkpoint on the deployment of a new version of software and roll back the upgrade if there is a problem. In addition to mitigating risk, being able to test new functionality in a virtualized lab environment reduces the time and the cost of testing due to the simplicity of downloading a new version of software versus acquiring and implementing new hardware. For example, if the EPC is virtualized, it is easy to prepare new instances of an update to one of the EPC's components in a CSP's lab, test those components and if the tests are successful, move them into production. This approach results in less down time, less testing and an increased likelihood of interoperability than would be the case if a hardware-based, monolithic EPC were implemented. In addition to saving time and money, being able to test new functionality in a virtualized lab enables CSPs to roll out those new capabilities more quickly.

The Management Challenges Associated with NFV

The concepts that are described in this section will be exemplified by referring to the vEPC. While using the vEPC as an example is helpful, the infrastructure changes that CSPs are making and the management challenges they face are the same independent of whether the principles of NFV are being applied to the EPC or to some other function, such as the IP Multimedia Subsystem (IMS).

The promise of NFV is very appealing. However, to fully realize the promise of NFV CSPs must make fundamental changes in their infrastructure. Part of the change that CSPs must make to their infrastructure focuses on moving from solutions that are hardware-centric and monolithic to solutions that are software-centric and highly modular. Once that change is made, driven in part by the goal of reducing the impact of network latency and enabled by the work of groups such as ETSI's Mobile-Edge Computing ISG, CSPs distribute much of the modularized functionality out close to the users. For example, a CSP may decide to

improve the user's experience by moving functionality such as the Serving GW or the MME out as close as possible to the user. A key factor driving CSPs to implement a variety of alternatives for where functionality is deployed is driven by the fact that a vEPC allows CSPs to effectively support new use cases such as the IoT. However, these new use cases have very different network requirements and price points than the smartphone and tablets that predominate in the current networks.

In the traditional way that a CSP deploys a service such as an EPC, the control function is entirely centralized. Given that these services were centralized, it is not surprising that the associated management solutions were also centralized and were based on implementing devices such as a packet capture probe at a relatively small number of central sites. However, that centralized approach to management becomes extremely difficult, if not impossible, as CSPs move away from services that are centralized and adopt services that are based on highly-distributed sub-functions. Because these sub-functions communicate with each other as well as with other centralized resources in a highly dynamic way, this creates complex traffic patterns that are comprised of a combination of east-west and north-south traffic. Further complicating the task of gathering management data is that fact that the cost efficiencies of NFV are predicated on being able to automatically provision and de-provision resources in a variety of locations. As a result, one critical characteristic of the management model that CSPs must adopt is that it must facilitate capturing management data directly from all of the elements of a service or function.

When trying to identify performance degradation before it impacts users, CSPs typically look at network performance indicators such as latency and packet retransmissions. However, to add the context that is necessary to identify an emerging performance-clogging hotspot, flow data is required and this generates huge volumes of management data. Netflow, for example, consumes between 1% and 2% of the capacity of link on which it is running. To quantify the volume of flow data that is currently being generated and which will be generated in the near term, it is helpful to recognize that many mobile devices that are currently shipping support the 4G LTC Cat 6 standard which enables download speeds up to 300 Mbps. In addition, some vendors have announced devices that support more advanced standards, such as 4G LTE Cat 16 which enables download speeds up to 1 Gbps. Assuming that Netflow generates management data at a rate of 1.5% the speed of the link, then at the low end (4G LTC Cat 6) this is generating 4.5 Mbps of management data and at the high end (4G LTE Cat 16) this is generating 15 Mbps of management data. Even if the CSP uses sampling techniques, this still creates huge volumes of management data.

The dramatic growth in management data that is generated by each successive iteration of cellular services bumps up against another fundamental limitation of the current centralized management model. That limitation is that in the traditional centralized model only relatively small amounts of detailed management data are stored for more than a brief period. CSPs will not be able to troubleshoot a highly dynamic, highly distributed VNF such as a vEPC by using small amounts of summarized management data. Hence, another characteristic of the management model that CSPs must adopt is that it must facilitate gathering and storing massive amounts of management data.

However, just implementing a big data monitoring solution that gathers and stores massive amounts of management data isn't sufficient. Another fundamental limitation of the traditional approach to management is that management tools typically focus on individual technology domains. To overcome the limitations of a siloed management model, an

effective big data solution must enable the management data that is gathered to be unified with Netflow data, location data as well as other data sources. To support the real-time requirements of a CSP's operations teams, the solution must also have a cloud architecture that supports multi-tenancy to ensure that a user or a group of users doesn't negatively impact other users. These types of sophisticated big data solutions are being driven in part by the growth of open source tools and communities that are focused on big data.

Summary

Most CSPs have already deployed NFV and the penetration of NFV-based solutions is expected to continue to increase. The reason for this is that NFV can help CSPs reduce cost, simplify their operations and reduce the amount of time and resources it takes to deploy new or enhanced services.

Implementing NFV-based solutions requires CSPs make fundamental changes to service delivery. This includes decomposing services into their primary components and running those components in a software-based, highly distributed manner. However, as CSPs adopt new, NFV-based service delivery models, their traditional approach to service management breaks down along three primary dimensions.

One dimension along which the model breaks down is that a centralized approach to management makes sense in those instances in which the functions and services being managed are also centralized. However, it breaks down when those functions and services have been dis-aggregated into a set of highly distributed components which communicate with each other in a decidedly dynamic fashion. To satisfy the requirements of managing a highly distributed system, CSPs must implement a management solution that gathers management data directly from all the centralized and decentralized elements of a service.

The second dimension along which the model breaks down is that in the traditional centralized management model only relatively small amounts of detailed management data are stored for more than a brief period. The increasingly large amounts of management data that is generated by each successive iteration of cellular services means that traditional management solutions will store an increasingly smaller percentage of the management data that is created. Managing NFV-based services and functions and services requires CSPs to adopt a big data approach to management so that they can store granular management data at massive scale.

The third dimension along which the traditional management model breaks down is that it focuses on individual technology domains. To overcome the limitations of a siloed management model, an effective big data solution must enable the management data that is gathered to be unified with a broad set of data sources. To support the real-time requirements of a CSP's operations teams, the solution must also have a cloud architecture that supports multi-tenancy to ensure that a user or a group of users doesn't negatively impact other users.