Transformation and Opportunity at the Service Provider Edge: Using Virtualized System Infrastructures in High-Value Use Cases in VCOs & at the Edge

Moving toward a Fully-Connected, Software-Driven World

Service providers are driving toward higher levels of innovation and achievement by the lure of a fully-connected, software-driven world. Widespread, continuous innovations in endpoint intelligence, capacity and reach of networks, and software capabilities harnessing the functionality of intelligent virtual infrastructures enable new applications and enhance existing services. Many of these are being delivered as a service by a mosaic of interconnected, interoperating service provider companies: application providers, cloud computing service providers, and telecommunications service providers most prominent among them. Across the landscape of these innovations, the contexts for interconnecting between providers and federating offerings among them are continuously expanding. This is creating a special and growing set of circumstances in which telecommunications service providers are enabling new service offerings by creating a new tier of capabilities in their service delivery infrastructures using cloud-based, software-driven technologies at the edge of their network deployments.

In this paper we explore the reasons why telecommunications service providers are pursuing these infrastructure enhancements. We consider both the economic and the technical factors underpinning their initiatives; we examine the top use cases in which the new agile infrastructures will be deployed; and we highlight the merits of the cloud-based, virtual system infrastructures being used in those deployments for achieving the service providers' goals.

The Special Role of Telecommunications Service Providers in a Fully-Connected, Cloud-Enabled, Software-Driven Environment

Telecommunications service providers have both a special role and a compelling set of opportunities in this evolving landscape. They are suppliers of both highly capable network connections for their users and of many innovative, value-adding services delivering applications on their own and with their ecosystem partners. Several years ago, SPs embraced a profound transformation in their operating methods by adopting the technological frameworks of the cloud for their network and application delivery infrastructures, using cloud-native paradigms and anchoring them to the reference architectures of network functions virtualization (NFV) in their most important use cases.

Significant progress has been achieved since the beginning of that journey. Service providers are evolving the design of their most important functions in residential, enterprise, and mobile network infrastructures to embrace NFV. There has been enhancement of the virtual system infrastructures used to support NFV in areas such as network throughput, use of processor cores and memories by virtual infrastructure software, and automating deployment of model-driven solutions to make operations increasingly efficient. Yet, developments in parallel continue, and additional innovations are being created as network sizes and capacities expand and new functional models for application deployment are brought into play.

What exactly is inspiring service providers to expand their use of virtual system infrastructures and extend their implementations to the edge of their network deployments? There are fundamental motivations in both economic and technical terms that are underpinning their focus. Those two dimensions are converging to give their initiatives their direction and rationale. In the following sections of this paper we highlight the key aspects of each.

Economic Incentives for Creating the New Virtual Edge

Service providers' motivations to invest in edge deployments are clearly linked to their opportunities to make new revenues and profits. This said, however, where will the opportunities for earning those revenues and profits come from, and how will they materialize?

The following four dimensions of service providers' evolving markets are creating the majority of their justification for pursuing the new edge plans. It is from these areas that new services and applications will emerge and from which new revenues and profits will be earned:

- Ongoing growth of mobile traffic in the foreseeable future led by video and IoT.
- Anticipated adoption rates for 5G network services in the coming five years and for multiple years thereafter.
- Continuing adoption of virtualized, cloud-enabled enterprise network and cloud computing services, including SD-WAN and multi-cloud XaaS.
- Superior total cost of ownership (TCO) and return on investment (RoI) for service providers using cloud-native, software-driven infrastructures.

Brief highlights in each of these areas help put the service providers' motivations in perspective.

First, focusing on the sources of growth in revenues in their mobile network services, traffic classified as data traffic in mobile is projected to grow at a compound rate of 43% per year over the coming five years through 2023.[1] This includes streaming video, gaming, social media, ecommerce, email and internet search applications, which are at the heart of users' adoption of smart mobile devices (from smartphones to tablets to wearable technology and others). At the end of that period, over 70% of traffic in mobile networks will exist as one form of video or another (live broadcast, recorded content streaming, user-generated video calls, safety monitoring videos, and more). Although users' budgets for applications are not infinite, the experience of the past decade in mobile has shown a steadily increasing willingness to allocate spending in reasonable increments to consume time-saving and life-enhancing applications, which has produced the growth rates we have cited).

During the same timeframe, mobile network-based IoT connections are expected to grow at a compound rate of 30% worldwide with more than 4 billion devices connected to cellular IoT applications by 2023. And in parallel, while the development of 5G will be in its initial phases during this period, by 2023 projections indicate that 20% of mobile data traffic worldwide will be carried on 5G. As we will discuss further (in other sections of the paper) many of the applications that will be generating the new revenues and many of the network functions that will be needed to support the expanded services will not, in fact, be possible without service providers enhancing their infrastructures with additional deployments at the edge. This simply creates additional focus for why and how the new edge investments will be made.

[1] Ericsson Mobility Report, June 2018, https://www.ericsson.com/assets/local/mobility-report/documents/2018/ericsson-mobility-report-june-2018.pdf.

In parallel with the mobile network evolutions, enterprise ICT will be gravitating to one form of cloud deployment or another (ultimately converging on a model of hybrid, multi-cloud ICT) and the ongoing growth of workloads in cloud-based environments is projected to grow at a compound rate of 32%.[2] This creates a powerful incentive for deploying flexible SD-WAN, vCPE and multi-cloud XaaS solutions. Finally, with respect to the TCOs and RoIs of the new virtualized deployments, over the course of numerous analyses, the economic advantages to service providers of pursuing implementations using cloud-native, openly-architected, software-driven platforms have regularly shown reductions in operating expenses for the deployments in excess of 50% and TCO reductions (which include both operating and capital expenses) consistently in the 30% and higher range.[3]

Driven by all of these factors, the motivations for service providers to create new offerings and expand their infrastructures to support the demand for these offerings is clear. As we have discussed, the most effective means for them to do so is to employ the most agile and efficient platform they can to extend their service delivery infrastructures to the edge.

Technology Considerations to Bring Cloud-Enabled, Software-Driven Platforms to the Network Edge

Having looked at the top economic drivers for expanding their virtualized infrastructures to the network edge, let's look at key technical considerations that combine with the business goals to support this set of enhancements.

Taking Virtual Infrastructures Further. Service providers' initial implementations of NFV have focused on decoupling the purpose-built platform and enabling functionality as software modules in virtual machine-based deployments. Most of the effort has been applied to enabling these functions in core SPs' network sites (where traffic is aggregated and forwarded to and from other network domains) and in more versatile implementation of platforms for customer premises deployments (in offerings such as SD-WAN and virtual CPE). Adoption of cloud-native technologies in these environments will continue over the next five years and beyond as their performance and flexibility improve.

While these advances have been ensuing, developments in other areas have been occurring that have a direct effect on the journey to the edge. Innovations in endpoint systems of many types (consumer devices, industrial and commercial equipment, municipal infrastructure elements, and more); in specialty processors used to support complex application functions; in networking technology to support higher capacities and greater flexibility; and in software supporting greater flexibility and agility in application development have been achieved. These innovations are also creating incentives for SPs to extend the reach of their virtual infrastructures further to the edge.

What Are the Attributes of Emerging Network and Application Functions that Are Stimulating Edge Deployments?

There are four primary factors motivating service providers to expand the capabilities of their infrastructures at the network edge:

- **New applications** emerging that have the potential to create new revenues.
- Creating the **elasticity and agility** to support the new applications.
- Supporting **new architectural frameworks** that expand capabilities in 5G, IoT and cloud-native designs.
- Maximizing the **efficiency and scale** required in supporting the new applications and frameworks.

Let's look at the influence of each on service providers' journey to the edge.

---

[2] Cisco Global Cloud Index Forecast and Methodology, 2016-2021, February 2018, https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf.

[3] *Creating Agility & Efficiency at Scale: The Economic Advantages of Open Architecture Platforms in NFV Deployments*, ACG Research, February, 2017, https://www.acgcc.com/creating-agility-efficiency-at-scale-the-economic-advantages-of-open-architecture-platforms-in-nfv-deployments/.

Addressing New Application Requirements. Because of innovations in endpoint capabilities in industrial and consumer applications, new functions are being built into them that create opportunities to add value and new requirements to be met. The new requirements are grouped into three major categories:

- Supporting **significantly lower latencies** in functions located closer to the endpoint than is possible using platforms located further away.
- Using **location-based information** within the applications.
- **Maximizing the economic value** of processing in infrastructure at the edge versus relying on functionality in cloud facilities that are further away.

Let's look at how these requirements materialize by examining one type of applications whose uses are growing rapidly: streaming video (and supporting analytics). As we mentioned previously, video streaming will be used in a wide variety of offerings that service providers support (from entertainment content streaming to social networking to safety and surveillance monitoring). The example in Figure 1 is in a safety and surveillance application. In this application, the value of low-latency processing, locally-resident analytics, integration with large-scale network environments, and integration into unified application solutions running in both local and remote sites are all highlighted. Because much of the functionality of the application cannot be realized without the presence of application functions at the network edge, it is a clear example of how this type of infrastructure expansion will create value for both service providers and their customers.

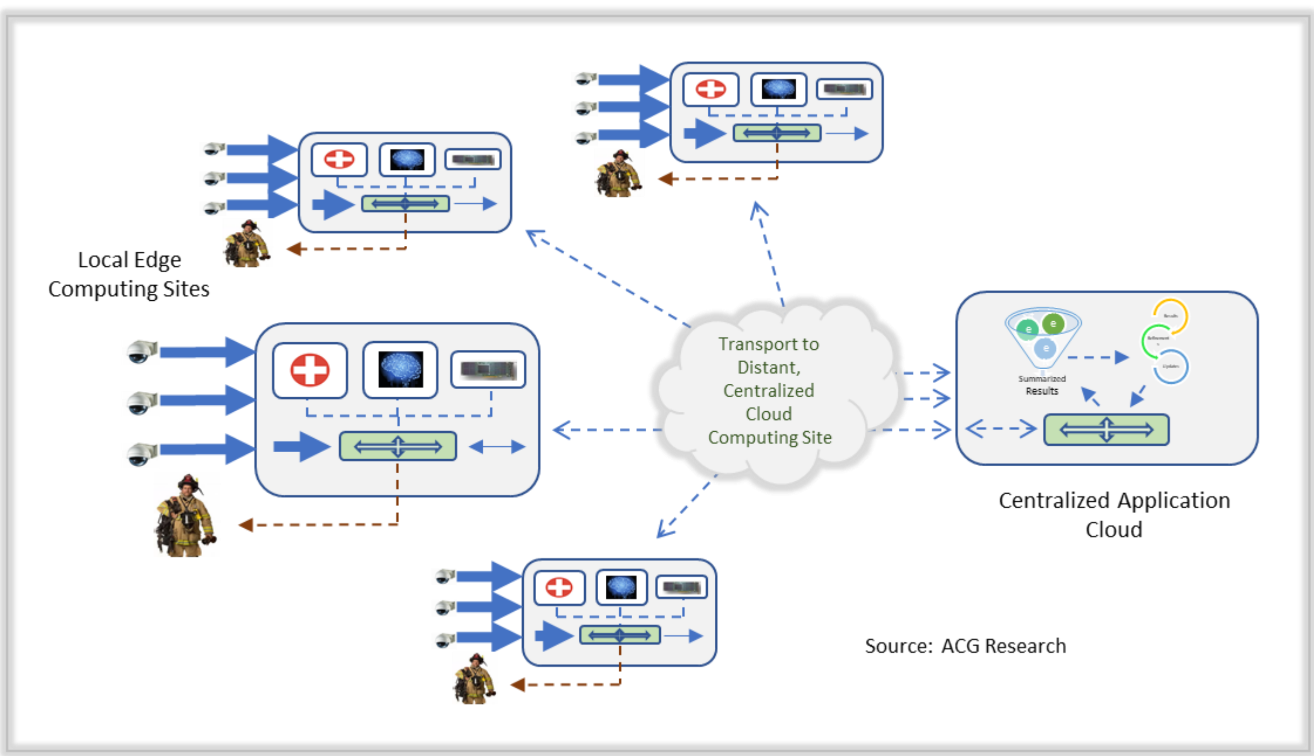

Figure 1. Example of an Emerging, Edge-Optimized Application: Streaming Video Analytics

Creating the Elasticity and Agility to Support the New Applications. Beyond low latency operation, new applications will require an agility and elasticity in how they use their infrastructure resources that is similar to the way large-scaled cloud-based applications run. They will vary in when they each create peak traffic demands (video surveillance during rush hours, for example, versus optimized CDN during early morning and later evening hours). The resources consumed by each will vary, differing in how much bandwidth, how many processor cycles, and how much memory and storage they need. Operators will use a virtual system infrastructure to meet these requirements most efficiently.

Applications will be implemented on a shared virtual infrastructure, and the unique requirements of each will be addressed by the software-driven infrastructure supporting it. Without this form of implementation, deploying the services at the edge would be infeasible.

<u>Supporting New Architectural Frameworks in 5G, IoT and Cloud-Native computing.</u>

5G Networks

From an architectural point of view, 5G will create a wide range of impacts on service deliveries, including new functionalities at the network edge. The extent of these impacts is well-documented.[4] For our purposes, we cite three aspects of 5G that will contribute to service providers using cloud-based, software-driven infrastructures in edge deployments: the introduction of virtualized RANs (vRANs) to the network; deploying networks using the architectural framework known as CUPS (control and user plane separation); and supporting partitioned services to end-users using network slicing.

- 5G will introduce several new radio access technologies, expanding performance, capacity and versatility.[5] Because these will be so diverse, the notion of virtualizing control for multiple radio infrastructures to simplify network operation in 5G has been created. Radio controls will be supported in pools of cloud-based, software-driven resources at locations in 5G network edges. These will be vRAN processing pools. Because they will use cloud computing implementations, vRAN pools may also support edge applications, as requirements and demand emerge.
- Simultaneously, 5G is employing the CUPS architectural framework that separates network control and user plane functions into different modules and can run them at varying points in the network topology. CUPS takes advantage of the agility and elasticity of virtualization by allowing operators to optimize platform resources according to the requirements of their services.
- Finally, with respect to 5G and clouds at the edge, 5G includes the ability to logically partition network resources so they are allocated to a particular service or end-customer, creating resource allocations that tune service offerings to the distinct requirements of each slice.

In these ways, 5G is stimulating service providers to expand on their use of cloud-based, software-driven infrastructures by deploying them broadly at the network edges.

IoT Applications

In parallel and in their own way, applications in IoT are creating requirements for cloud-based infrastructures at the edge. The video surveillance and traffic monitoring applications we cited previously are among these cases, combined with others, including emergency response, industrial quality control and safety monitoring, and AR enhanced consumer applications. In each of these cases, requirements for minimizing delay, using location-based information in the application, and minimizing the cost of the deployment are a part of the motivation.

Cloud-Native Computing in New Software Implementations

The third architectural transition is the use of microservices and cloud-native software design as the framework for how virtualized network functions (VNFs) and edge applications will be deployed. A prime advantage of cloud-native designs is efficiency in using underlying platform resources. A second advantage is relative simplicity of orchestration when deploying applications at scale (as proven out in hyper-scale cloud provider implementations). Although there

[4] *View on 5G Architecture (Version 2.0)*, 5GPP Working Group, 07/18/2017, https://5g-ppp.eu/wp-content/uploads/2017/07/5G-PPP-5G-Architecture-White-Paper-2-Summer-2017_For-Public-Consultation.pdf.

[5] Summary of new radio (NR) specifications for 3GPP Release 15, July 2017, http://www.3gpp.org/release-15.

are areas in which cloud-native designs will need to be strengthened to achieve their promise, the prospect of including them in edge offerings is appealing to both service providers and developers. There is little doubt that as edge deployments proceed there will be increased use of cloud-native computing designs within them.

Maximizing Efficiency and Scale in the New Applications and Architectural Frameworks. The final trigger motivating providers to embrace new virtualized designs in their edge deployments is to maximize the return on their investments by achieving the greatest possible efficiency that they can. A deeply held tenet in service provider transitions using NFV is, in as many cases as possible, to use general-purpose hardware versus purpose-built hardware platforms; to use openly architected, virtualized platform software; and to use openly architected and maintained data models and application programming interfaces that allow them to maximize flexibility in deciding what to deploy. Cloud-based, software-driven solutions based on open architectural designs provide them with the path to achieving these efficiencies.

Where Will the Service Provider Network Edge Be Located?

Having explored **why** service providers are investing in a new tier of infrastructure at the edge of the network deployments, an important question is, exactly **where** will the network edge be, and how will it be deployed? An inclusive, end-to-end view of the answer is shown in Figure 2.

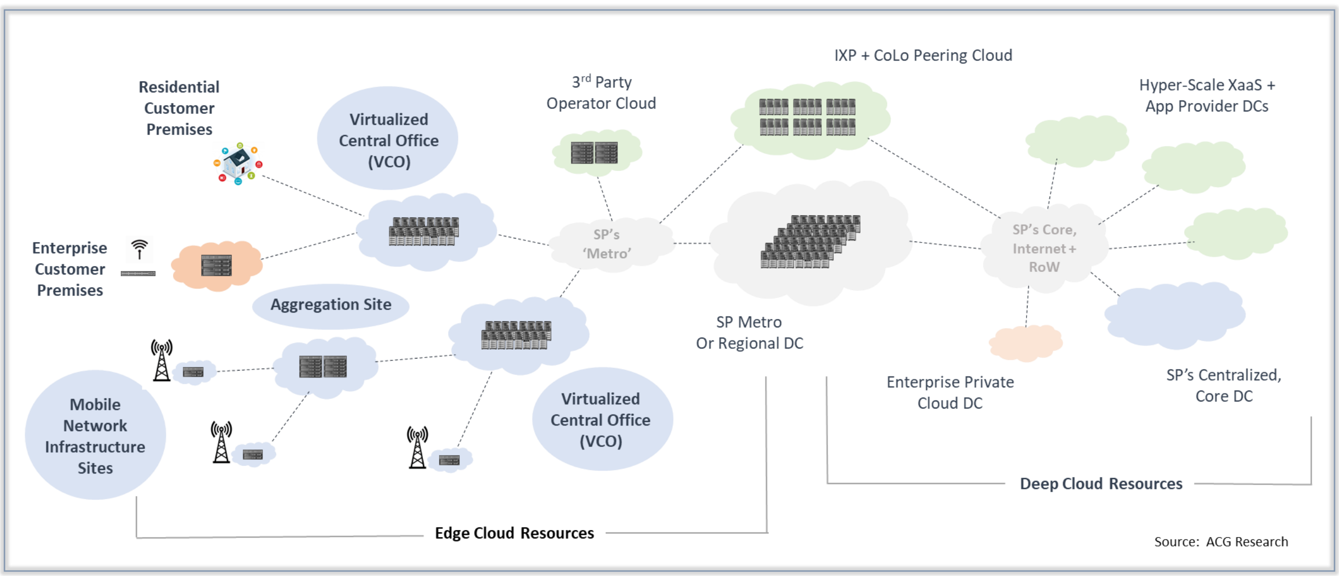

Figure 2. Network Edge Deployment Sites

In the long run there will be many forms of edge deployment, including service provider, application provider, and cloud provider deployments. They will each have their own topology and in many instances will be closely interconnected and integrated for a seamless experience. Because we are concentrating on SPs' deployments in this paper, we will focus on the locations illustrated in blue in the edge cloud resources section of the diagram.

The service provider's central office (CO) is among the key sites that will be used in network edge deployments. For edge deployments, these sites have the virtue of being a small number of kilometers away from the majority of their end-customers. They are well-suited to delay-sensitive and location-aware functions SPs want to support. Because central offices were originally built for a different generation of services (principally the original voice and data services that service providers offered over many decades) not all central offices will be converted for use in the new virtual edge. Of the approximately 150,000 central offices that have been deployed across the globe, we expect on the order of 50% to be employed in one form of new deployment or another as the evolution of edge proceeds.

Figure 2 illustrates that COs are traffic aggregation sites for additional locations that are even closer to end-users (such as mobile network radio towers), as well as to end-users' locations themselves (in the case of enterprise and residential services). This mix of downstream sites is important to consider with respect to how edge deployments will be made. Each is associated with use cases and services that service providers will be offering. Though the exact functionality of the components in each deployment will be specific to the users being supported, all offerings will be running in, and in many cases sharing, the virtual system infrastructure the SP has deployed in the CO. Because of this, CO deployments in the new network edge will often be referred to as virtualized central offices or VCOs.

Let's look at the three major end-user environments VCOs will support and how they will shape edge network deployments: residential, enterprise/business, and mobile services. Since there is ample material available in parallel to this paper for understanding the technical details of how these deployments will be done[6], our concentration is to highlight the role of virtualized infrastructures in creating VCOs and the importance of the model for service providers as they pursue their edge initiatives. We address them in the top to bottom order in which they display in Figure 2.

In the upper left of the diagram is a broadband connected home. Services in this case exist in multiple forms (based on DSL, PON and coax cable). They support over a billion endpoints throughout the world.[7] Multiple components in these cases operate in VCOs in parallel with elements deployed at the residential site (such as gateways and application devices), along with elements such as content libraries and provider interconnections deeper in the network. In many instances a virtualized Broadband Network Gateway solution or in other cases a virtualized Head End implementation is deployed. In each of these, control plane, user plane, and value-adding functions (such as malware detection) run in the VCO supporting users according to their entitlements and related policies. As developments progress, an increasing range of delay-sensitive (security surveillance and content delivery optimization, for example), location-aware (WiFi calling and in-home appliance controls), and user-aware functionality will be run for residential customers in the VCO.

In parallel (and just below the residential deployments in the diagram) are virtualized network and cloud-based services to enterprises in many industry sectors and geographies. As the diagram shows, in some cases, resources may be located at the customer's premise; while other resources reside in the VCO (and in deeper cloud computing sites). The paradigm being used overall is grounded in the underlying goals of NFV: to move services off of purpose-built appliances onto software-based virtual network functions that can be dynamically composed and allocated in the virtualized environment.

Significant work has been done in deploying virtualized services to enterprises using virtualized CPE (or vCPE) and software-defined WAN (or SD-WAN). The role of VCOs in these services to date has been to supply virtualized functions such as firewalling and malware protection on a dynamically allocated basis from the cloud-based infrastructure. As markets and applications progress, it is likely VCOs will support new latency-optimized, and location-aware IoT and sector-focused applications (for example, for healthcare and retail customers). These will benefit from distributing intelligence into the VCO while economizing on delivery costs by using the shared virtual infrastructure in the VCO, which will also participate in the orchestration required in multi-cloud deployments for application suppliers to support their offerings using the VCO.

---

[6] See, for example, a description of the VCO 1.0 and VCO 2.0 solutions demonstrated by the OPNFV ecosystem at the OPNFV Summit in Beijing, China, in June, 2017: https://wiki.opnfv.org/display/OSDD/VCO+Demo+1.0+Home, and at the LFN Open Networking Summit in Amsterdam, The Netherlands, in September, 2018: https://wiki.opnfv.org/display/OSDD/VCO+Demo+2.0+Home (respectively).

[7] Robin Mersh, CEO, Broadband Forum, *Broadband Access Abstraction*, Presentation at the SDN + NFV World Congress, The Hague, Netherlands, October 11, 2018.

The final set of locations where edge deployments will occur is in the distributed sites of mobile network infrastructures (lower left of the diagram). As 5G specifications describe, important components will continue to be installed in deeper core sites where policy decision-making and network interconnection with other network infrastructures will be made. However, substantial portions of 5G infrastructures will also be deployed in distributed sites to support latency-sensitive functions, achieve flexibility in resource allocations using CUPS, and introduce edge applications in parallel with VNFs. We believe a continuum of virtualized capacity will be created in service providers' mobile networks, from core to VCO to aggregation hubs and ultimately to radio towers to support new applications.

As we can see from these descriptions, new edge deployments in service providers' networks will frequently be based in VCOs and will also include resources in smaller sites closer to customer endpoints when installing new resources in those smaller sites as well will meet service delivery requirements best.

How and When the New Virtual Edge Will Materialize

Many efforts have been underway in the past two to three years to examine use cases, investigate dimensioning and technical requirements for VCO and edge deployments, and to begin aligning service providers, application providers, and infrastructure solution providers on what will define success in the various environments. It is fair to say that the requirements scoping and solution prototyping phases in this work are nearing conclusion, and the process of developing and delivering solutions at each layer of the service delivery stack has begun. Efforts within the OPNFV's VCO 1.0 and 2.0 demonstration initiatives and other endeavors such as the Linux Foundation's Akraino edge computing project and the OpenStack Foundation's edge cloud working group are all contributing materially to what solutions will contain. In parallel, service providers are conducting their own evaluations and developments to demonstrate and deploy trial implementations in various cases.

We believe deployable offerings at each layer of the stack, from infrastructure blueprints, to VNFs for the functions critical in each environment, to management and orchestration that begin to ease the burden of deploying widely distributed solutions will begin to arrive in market steadily between today (2018−2019) and 2020−2021. VCOs will be among the early targets for deployments, accompanied by sites further toward the edge, such as radio towers and aggregation hubs, as 5G comes online. As solutions mature, we expect uptake for network edge solutions to be substantial, on the order of a 50+% CAGR in the period through 2025. That date will not be the endpoint in deployments of this model. It simply defines the period in which, like many other transformations in ICT, the foundations of the cloud-based edge will have been put in place. Uptake following that will be driven by the pace at which applications taking advantage of the new software-driven environments can materialize.

How the New Virtual SP Will Operate in the Fully-Connected, Software-Driven World

Strategically, we see an end-game in which SPs are supplying dynamically provisioned, value-adding services to the end-users they reach best. They will, of course, use physical infrastructures with advanced functional capabilities (new 5G radio deployments, for example). Beyond that, the critical service delivery components will be automatically enabled in elastic, cloud-native infrastructures using open software and solutions integrated via well-known APIs and data models to supply on-demand services from a mosaic of operators collaborating based on which brings which functionality to the customer best. Today's service providers will progress toward being seamless enablers of advanced applications, whether the resulting service is completely provided by themselves or is a composite put together by multiple providers. This will be the case in both mobile and wireline deployments (see Figure 2). It will be the case for both human and machine interactions (as IoT solutions will be software-driven as well).

Service providers will follow an iterative path of deploying tightly scoped installations for a limited set of use cases in the 2019−2020 timeframe. This period will see a significant number of technical evaluations, proof of concept engagements, trial and early, initial stage deployments in well-targeted use cases. However, based on the designs of the solutions deployed during this initial cycle, they will keep working to reduce the barriers to full-scale automation in deployments that remain in the early stages and proceed toward a model in which applications can be delivered with a minimum of delay, with a maximum of agility and scale. The barriers of latency and distance will be significantly overcome in the new edge deployments. The agility with which services and applications can be deployed (and consumed) will be accelerated by an order of magnitude by the capabilities of the new software-driven solutions and their operation at the network edge. By the 2023−2025 timeframe, we expect the model to be in widespread adoption across the range of use cases we have highlighted.

This paper has been sponsored by Red Hat.