Strategies for Implementing Edge Services in the 10G Cable Network

A Technical Paper prepared for SCTE•ISBE by

Eric Heaton
Platform Solutions Architect
Intel Corporation – Network Platforms Group
2200 Mission College Blvd. Santa Clara, CA 95054
408-765-3447
eric.d.heaton@intel.com

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

1. Introduction

Cable network bandwidth demands are growing exponentially as video becomes ubiquitous, Internet of Things (IoT) devices proliferate, and new wired and wireless access technologies come online. Gartner estimates 90 percent of the data generated by the massive number of Internet-connected devices is sent to regional data centers for processing,[1] further stressing network infrastructure and increasing average response times for everyone.

With edge computing, there is an incredible opportunity for broadband connectivity providers and those offering over-the-top (OTT) applications and services to make sense of and take action on the data coming from cars, cameras, factories, enterprises, and homes. It enables entirely new categories of services requiring ultra-low latency (i.e., augmented/virtual reality (AR/VR)), enhanced data privacy (i.e., medical records), or bandwidth optimization (i.e., video surveillance).

Figure 1 shows a range of industries and application segments that will benefit from these edge network and compute capabilities.

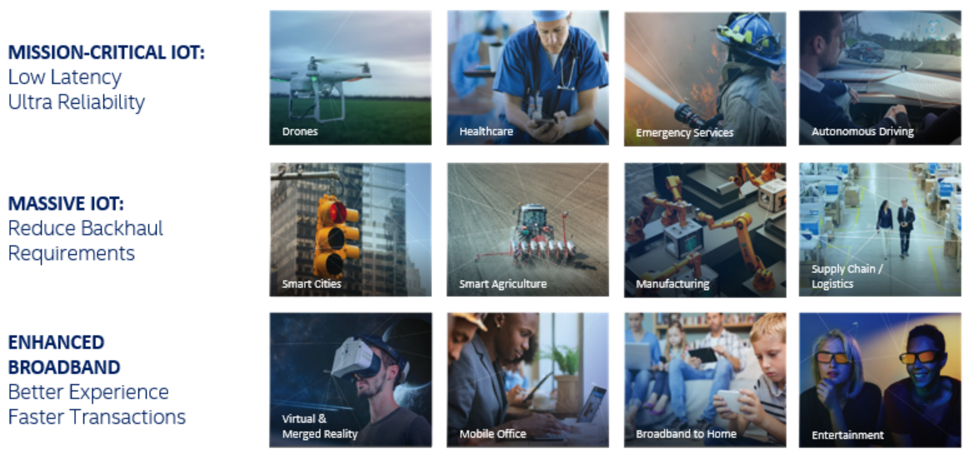

Figure 1. Edge Technology Network & Services Impact

For many networks, the latencies and other key performance indicators (KPIs) for specific services within these segments, as shown in Figure 2, will require upgrades across the network infrastructure to:

- Reduce end-to-end latency by an order of magnitude
- Allow data to be processed closer to where it is generated, and
- Provide a coordinated deployment and management system to keep costs in line with revenue

[1] Gartner, "Edge computing promises near real-time insights and facilitates localized actions.";
https://www.gartner.com/smarterwithgartner/what-Edge-computing-means-for-infrastructure-and-operations-leaders

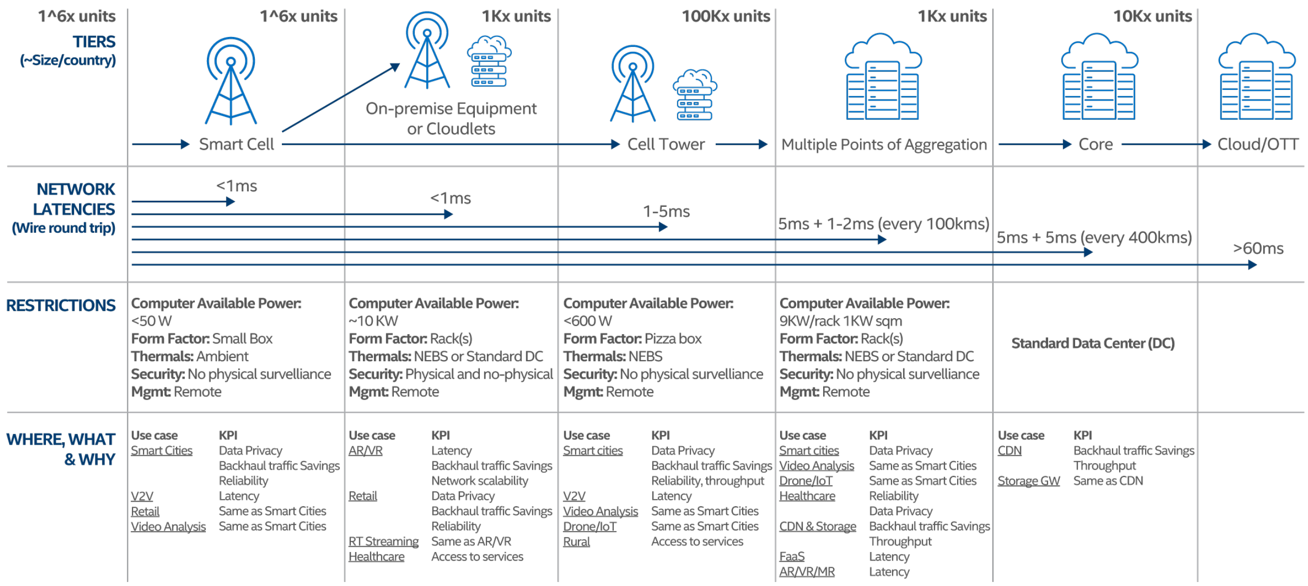

Figure 2: Use Cases and Associated KPIs

Looking at the topology of the network, there are multiple places across network infrastructure where computing resources can be placed . "The Edge" makes for an interesting and obvious place to manifest key infrastructure because it is physically closer to end users. Figure 3 shows that there are multiple places that could be considered the Edge, and they all can exist in the same network. Figure 3.Section 2 will explore both deployment models, what types of services are best provided by one model compared to the other, and other considerations to deliver maximum flexibility and return on investment (ROI).

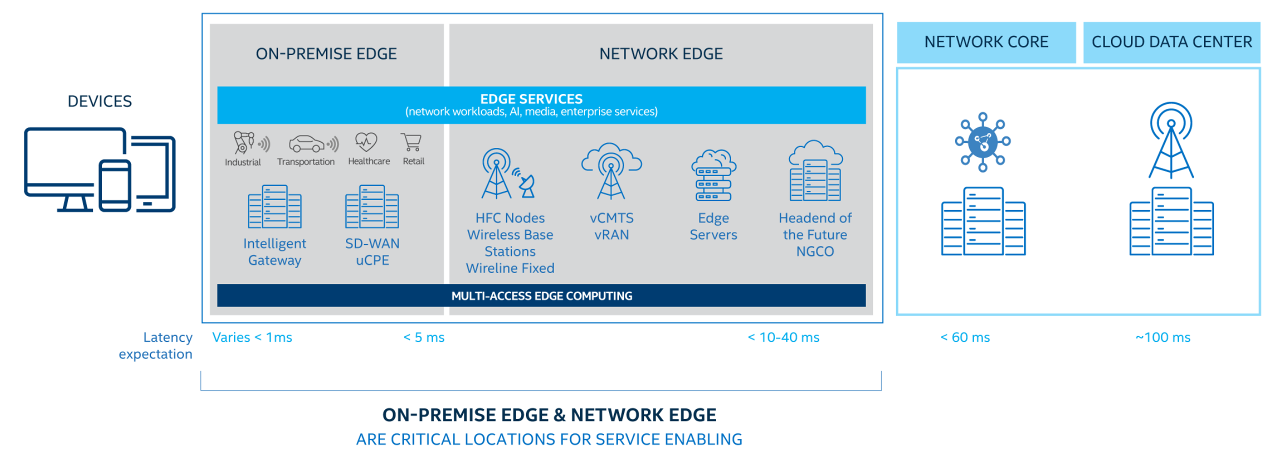

Figure 3. Logical Parts of the Communication Service Provider Network

Figuring out where to host network infrastructure is only part of the challenge. This paper discusses considerations for designing and buying edge hardware and software that will run the network, how functions are split across different equipment, and who owns the functions, users, and traffic.

In a single-access or single-service world, it is easy to line up bespoke solutions that include appliances and custom management interfaces. Finances may work for two or three solutions set up in parallel. To truly scale edge infrastructure and maximize resources and operational efforts, a common platform (at minimum common building blocks) that addresses all access technologies and service deployments makes the most sense. This paper discusses the benefits of converging network, IoT, and other edge workloads onto a single set of standard hardware with open software and interfaces. In doing so, complexity and time-to-market are reduced, and orchestration is streamlined across any edge location regardless of functional needs and environmental constraints.

There are many reasons that compute resources are moving to the edge, including network optimization through virtualization and data locality, cost savings through white box platforms and automation, and new monetization opportunities through the introduction of new services and business relationships. This paper primarily focuses on the latter—opportunities and considerations around implementing services. Topics like virtualization and automation will be referenced as they underlie modern network architecture and implementation.

In fact, recent headlines have shown that some communications service providers (CoSPs) and cloud service providers (CSPs) are already implementing services at the edge and selling them commercially[2]. This paper should serve as further encouragement and also as a guide to show there are many considerations to creating an effective and scalable edge that supports multiple access technologies and the latest services, hosted by the CoSPs, CSPs, and companies with OTT offerings. It is the right time for Multiple System Operators (MSOs) to plan their new edge offerings, as it can be done in conjunction with the ongoing march towards a distributed access architecture and the additions of virtualized environments. Understanding the key architectural options discussed in this paper will enable network operators, MSOs in particular, to realize the value of existing infrastructure to improve the customer experience at the edge.

2. Edge Deployment Models

Though there is no single definition of "edge," we can generally think of either an "enterprise/on-premise edge" or "network edge." Within an enterprise environment, this can include smart sensors, intelligent gateways, edge servers, and hyperconverged "local" data center infrastructure, with generally increasing complexity, power, and capability as one moves left to right. Similarly, the "network edge" contains platforms and at a variety of locations, as shown in Figure 4.

Some networks will not manifest all locations because the physical real estate to host equipment does not exist, locations and functions may have been consolidated onto fewer platforms, or some functions may not be implemented for one reason or another. Conversely, the network operator does not have to necessarily choose on-prem or network edge models exclusively, as each has pros and cons dpending on technical and business goals.

[2] Fierce Wireless, "Verizon, AWS bring 5G MEC to Boston, Bay Area"; https://www.fiercewireless.com/operators/verizon-aws-bring-5g-mec-to-boston-bay-area

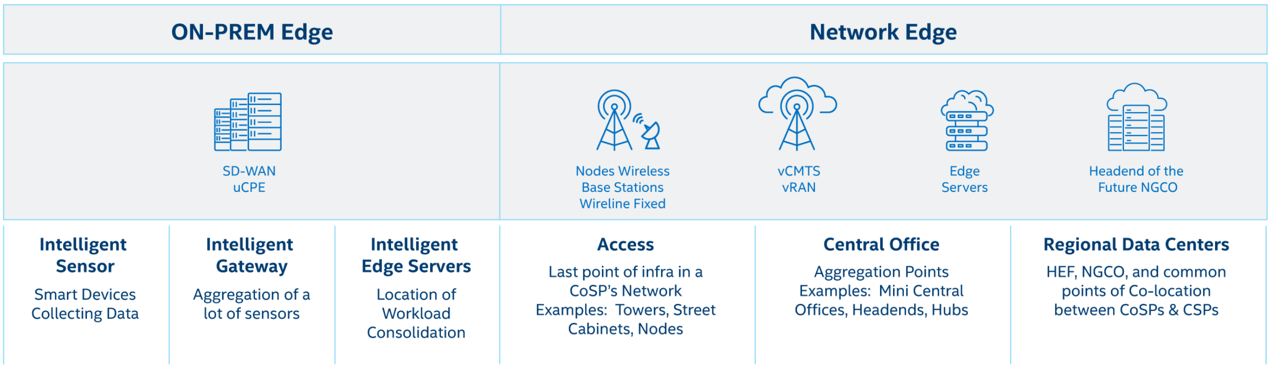

Figure 4. Edge Locations and Terminology

The following sections cover these deployment models in turn.

2.1. On-Premise Edge

Figure 5 shows an on-prem edge deployment model. Here, there is a controller for services that is located at some centralized place in the network and manages functions and services running on an edge platform, like universal customer-premises equipment (uCPE). Non-real-time functions such as controllers and service management will reside in the most cost-effective place, typically deep in the network or even in a public cloud. Alternatively, the controller could be deployed at a network edge location to comply with legal requirements or satisfy operator or customer requirements for full data locality.

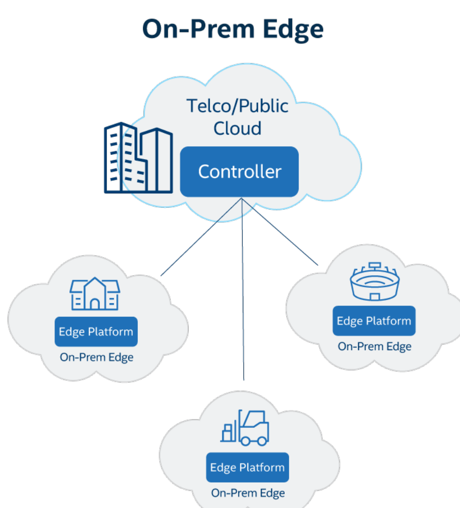

Figure 5. View of an On-Premise Edge Deployment

Growing at a triple-digit rate in recent years,[3] SD-WAN solutions are generally offered through an on-prem model, serving various applications for stadiums, farms, industrial IoT, and the like. True to its name, the main distinguishing factor for the on-prem model is a flexible platform at the customer premises that can run dynamic, real-time workloads locally. These workloads can be virtual network functions (VNFs) or services.

While this paper focuses more about the requirements and options in the Network Edge, the platform choices for on-prem models are comprehended in the convergence framework discussed in Section 5 .

2.2. Network Edge

Figure 6 shows a network edge deployment model. In this case, there is also a controller—or more likely, a set of controllers—to control network function virtualization (NFV) infrastructure, multiple VNFs, and/or services hosted directly by the operator or its partners. These controllers will orchestrate and manage such functions on edge platforms at a variety of locations. As in the on-prem model, controllers will likely run from a regional data center, private cloud, or at a CSP.

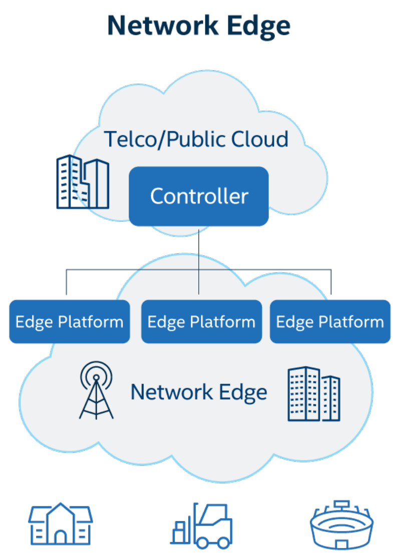

Figure 6. View of a Network Edge Deployment

This infrastructure will support broadband access, perhaps more than one type, as well as host services that can be delivered by operators themselves, CSPs, or third parties. Given these possibilities, edge platforms are not necessarily a single server or appliance but can manifest as a collection of equipment at

[3] Mann, "Coronavirus Cleaves SD-WAN Revenue Growth"; https://www.sdxcentral.com/articles/news/coronavirus-cleaves-sd-wan-revenue-growth/2020/06/

a location, providing a set of APIs upstream to controllers and other functions, and downstream to users. In short, edge platforms may involve more than one piece of hardware or more than one physical location.

To keep costs down as the number of access types and services increases, network edge deployments need common ground between the different hardware elements and a homogeneous software layer for management.

Many of the platform-as-a-service products that have been announced[4] by various CoSPs, CSPs, edge compute specialists and real estate management companies (i.e., towers), are based on a network edge deployment model. There are a variety of ways these parties organize resources to offer edge platforms. They must consider what network functions and services to offer (and who should own them), where equipment should exist, who should own the customers, how revenue should be divided, and so forth. Answers to these questions have both business and technical repercussions, and for the latter, affect how the edge platforms/locations are architected. The next section describes emerging network edge platform architectures and their associated advantages and disadvantages from the network operator's perspective.

3. Network Edge Deployment Models

Generally, on-prem networks run network functions and services on a uCPE at the customer location, with a software controller running deeper in the network on standard servers to manage edge services.

However, the network edge model has many options for splitting functions across various locations and deciding who will own those functions and equipment (i.e., some may be owned by a CoSP, CSP, or a third party). The CoSP edge architect will have to consider which services the network operator would like to host, where equipment (and its capabilities) can be deployed, how to split network functions (i.e., controllers versus data plane), and who is going to own which parts of the solution, in order to come up with a comprehensive edge platform deployment model.

There are several divergent approaches emerging in the market. Each approach allows for different types of business arrangements (i.e., how revenue is paid and split up) and will require different technical arrangements to be made between the partners involved.

1. CoSP + CSP Co-location
2. CoSP Led
3. CSP Led
4. CoSP Aggregator

The following sub-sections will review what it takes to implement each of these business arrangements, along with the driving forces for a CoSP to pursue one over the over.

3.1. CoSP + CSP Co-location

A popular emerging model for network edge deployments is for a CSP to co-locate equipment to deliver edge services at locations owned by a CoSP. Figure 7 shows how this would work. Presumably, the

[4] RTTNews.com, "Microsoft To Use Telefonica Infrastructure for Datacenter Region in Spain"; https://www.nasdaq.com/articles/microsoft-to-use-telefonica-infrastructure-for-datacenter-region-in-spain-2020-02-26

locations being "shared" by the CoSP and CSP would have latency advantages over what the CSP could promise on their own, with the CoSP having physical real estate very close to end users.

A CSP could then charge a premium when offering a content delivery network (CDN), a video analytic engine, or a generic platform-as-a-service with tighter and better guarantees around latency or data locality than the same services deployed from a centralized data center.

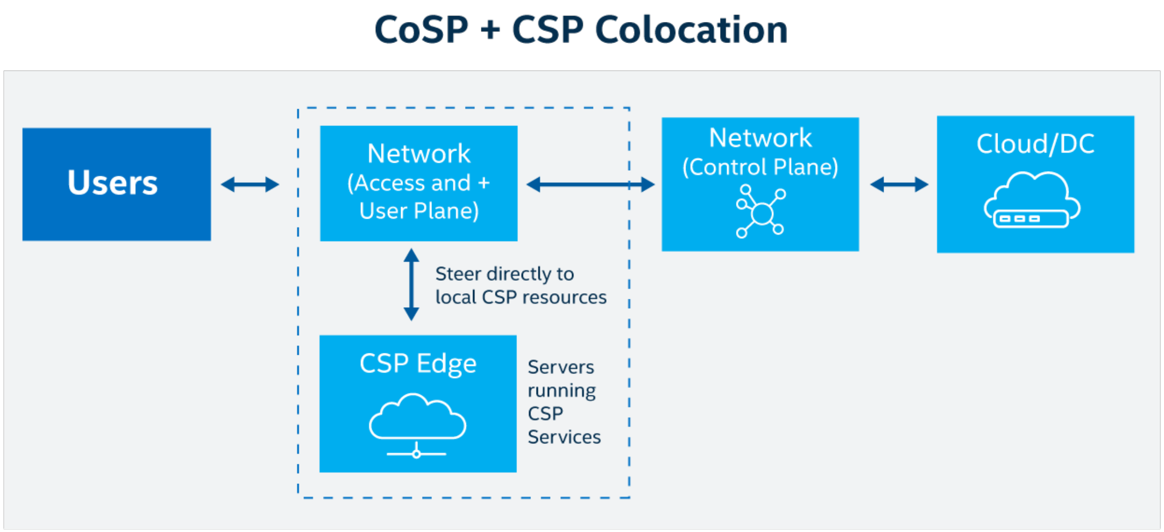

Figure 7. CoSP and CSP Co-location

The most straightforward benefit for a CoSP is "rent" or revenue share from the CSP. More interesting for business development could be a "better together" story—marketing the unique combination of CSP services and CoSP access medium that is not available through other providers. There may also be practical reasons to enter into this type of agreement, where the CoSP lacks expertise or desire to develop and manage the services being provided by the CSP.

The downside for a CoSP is lack of control on how co-located services are monetized, and lack of access to user or traffic telemetry data for further monetization. If the CoSP location is not set up to host and secure multiple parties, there may also be inadvertent CoSP/CSP lock-in or constraint to working with one CSP at a time.

For this model, there are two sets of equipment hosted at the CoSP location: one owned by the CoSP providing network connectivity and the other owned by the CSP providing services. However, this arrangement does not preclude the CoSP offering services, as the CoSP may utilize the CSP to host services on their behalf.

This arrangement can allow knowledge transfer of cloud technologies to the CoSP and be a stepping stone to some of the following models.

3.2. CoSP Led + CoSP/CSP Services

The second most common model is for the CoSP to own the edge platform altogether, hosting CSP services. In addition to owning the access equipment, the CoSP develops a services API to allow CSPs to exist at advantageous edge locations. Here, the CoSP owns the execution and delivery of the services to the customer. Figure 8 shows how this would work.

For the CoSP there is more complexity and internal expertise needed to develop and maintain a commercial platform-as-a-service. CoSPs do not have ownership of the services themselves but could monetize the management of edge locations and equipment.

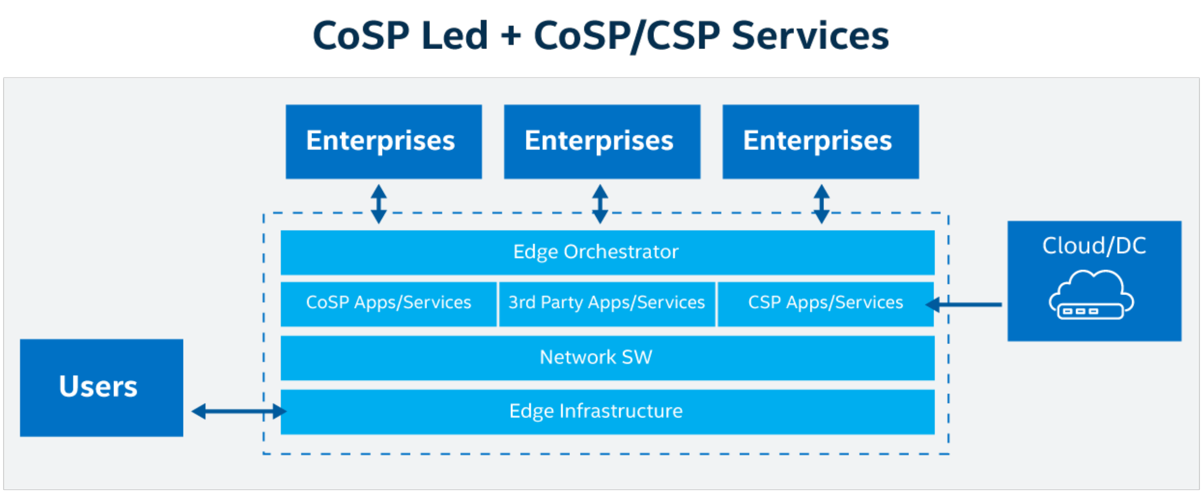

Figure 8. CoSP-Led Edge Deployment

Additional benefits for the CoSP include avoiding the logistics of hosting another company's equipment, and having a unified interface to sell its edge platform to any number of CSPs or third parties. Developing this model also implies a CoSP investment in understanding cloud technologies, which benefits overall network virtualization.

In a CoSP-led model, the edge platform is wholly specified and managed by the CoSP. There may be multiple types of equipment involved, but a single owner is likely to host all the functions and services on a common set of hardware for better economies of scale in both CapEx and OpEx. High-performance VNFs running on standard servers and new options for programmable switches are making it easier to reduce the number of specialized appliances in the network, as described in more detail in the next section.

3.3. CSP Led

The last two models are less prevalent, but do fit certain niches. In the CSP-led model shown in Figure 9, the CSP owns the edge locations and platforms. This assumes the CSP has real estate close enough to users to distinguish their service offerings from those running at their central data centers. In this case, the

CSP uses a CoSP for last-mile connectively, but otherwise owns all aspects of the service delivery and management.

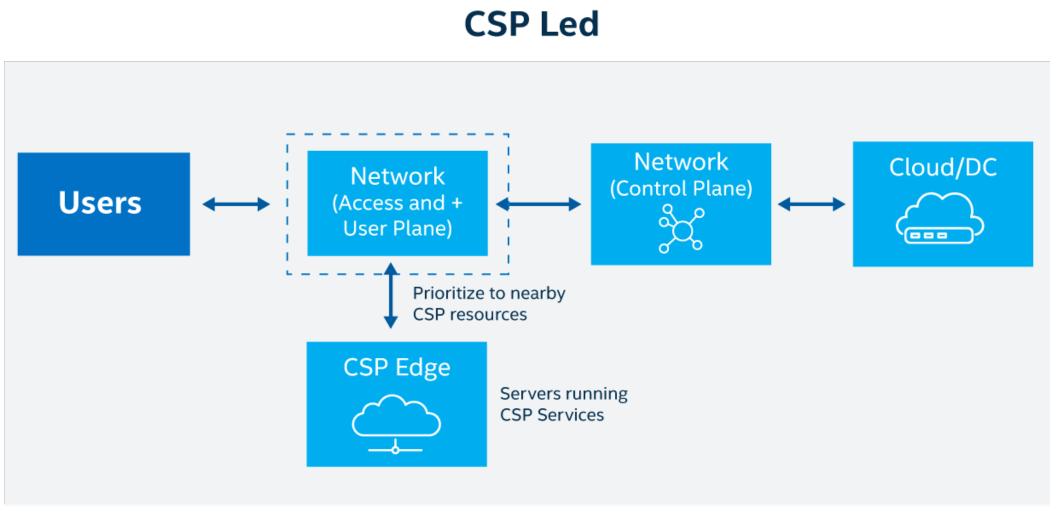

Figure 9 – CSP-Led Edge Deployment

For a CoSP, this model may appear as a high-end "business as usual" arrangement, where a CSP or any other customer is paying for a broadband offering, albeit one with low or ultra-low latency guarantees. That said, CoSPs are finding that this partnership is a convenient alternative to the CoSP + CSP Co-location option insofar that a similar "better together" story can emerge without the added logistics of sharing the same physical space[5]. In this model, the CoSP and CSP platforms can be developed and deployed relatively independently if the connection between the two networks can minimize latency to an absolute minimum.

Like the CoSP + CSP Co-location model, the CoSP could also become a customer of the CSP, hosting new services on their behalf.

3.4. CoSP/CSP Aggregator

The last model is one in which a third party aggregates connectivity and service options from a variety of CoSPs and/or CSPs and offers a common API to other service providers. In this case, the aggregator owns the real estate, buys connectivity from one or more CoSPs, contracts with CSPs for services, and may even offer its own edge services. The result is called an aggregated edge offering.[6] Figure 10 shows how such an aggregator can develop its own platform or service API that then plugs into the offerings of partner companies for execution and delivery.

---

[5] Robuck, "Cox targets the Edge for the next evolution of network performance and security", https://www.fiercetelecom.com/telecom/cox-targets-Edge-for-next-evolution-network-performance-and-security
[6] Dano, "SBA, American Tower double down on Edge computing opportunity"; https://www.lightreading.com/the-Edge/sba-american-tower-double-down-on-Edge-computing-opportunity/d/d-id/762941

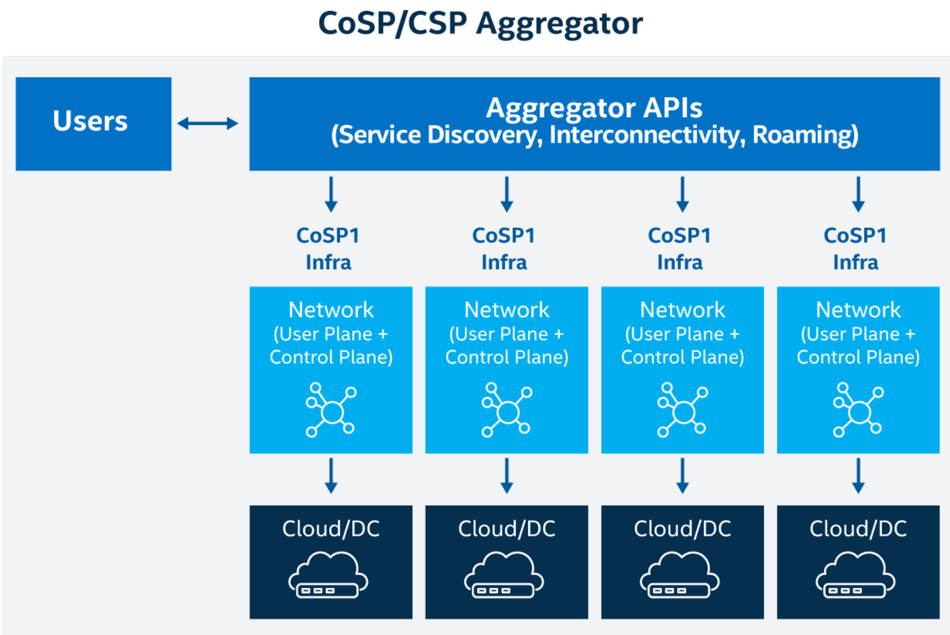

Figure 10 – Aggregator-Led Edge Deployment

CoSPs could monetize this arrangement in multiple ways. If they are just selling broadband connectivity, it is basically the same as the CSP model. Alternatively, they could offer their own platform-as-a-service, as described in the CoSP-led approach.

For the aggregator, the access and broadband connectivity from CoSPs, and therefore the services coming from CSPs, is ideally via an Ethernet/IP network. In this case, the edge platform architects do not have to worry about the disparate requirements of physical access technologies, making it is easier to choose standard servers and switches.

4. Mapping to Real Estate

Putting edge deployment models and platform architectures into the real world requires equipment, software, and operations to live in physical locations in the MSO network. This section maps the theoretical to the empirical world and discusses the competitive advantages cable networks may have compared to networks with other access technologies.

CableLabs recently described typical MSO network locations and how their characteristics could apply to edge deployments.[7] Figure 11 shows how various locations are connected and how far they are from end users.

[7] Levensalor, Stuart, "The Modular, Virtualized Edge for the Cable Access Network", https://community.cablelabs.com/wiki/plugins/servlet/cablelabs/alfresco/download?id=2c46cef2-af44-47be-bdd4-98a948cbc60d

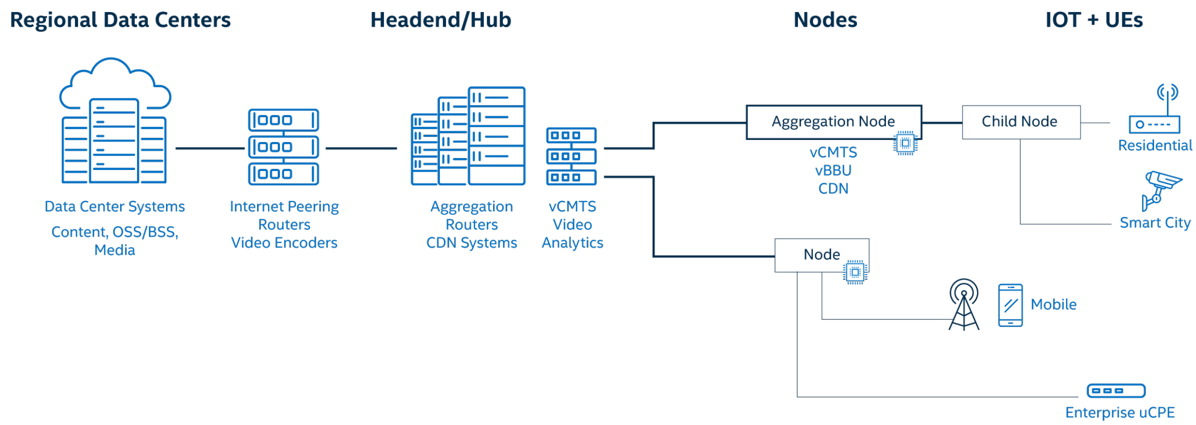

Figure 11. Type of Locations in the MSO Network

Both on-prem and network edge locations have benefits and constraints. On the benefits side, the closer one gets to devices and users, the lower the latency, the higher the data locality, and the lower bandwidth required upstream in the network. On the constraints side, the closer one gets to devices and users, the less power, the less environmental control, and the more costly it is to deploy and service equipment.

The left side of Figure 12 shows a large, multi-story, regional data center (e.g., central office or CO) serving up to hundreds of thousands of users. Generally, the set up uses a typical data center approach. Moving to the right along this continuum toward users, the environmental constraints increase, the compute capacity goes down, and the deployment and management of equipment becomes more costly.

Latency	Further from users and higher aggregation density << ------ >> Closer to users and serves more specific area
Environment	Full power and environmental controls << ------ >> Limited power delivery; space and thermal constraints
Management	Can use typical data center techniques << ------ >> Equipment is distributed and harder to manage/service

Figure 12. Network Edge Location Characteristics

CoSPs must decide if they want all the above locations to be "available" for their edge. That is, will a given location and the equipment therein have the right level of physical connectivity in both directions, and an ability to run network functions or enterprise software? Further, how much of this hardware and software can be brought into a centralized management domain?

In the ideal cloud-extended-to-the-edge vision, every location is part of a large pool of flexible and distributed compute, storage, and networking resources and functions. Services can be set to run where they are needed to satisfy technical and business needs at the lowest possible cost. This is easiest when all hardware is standard, and the software has similar resource needs, as in data center and CSP locations. However, the further out one gets from regional data centers, the more likely software applications have higher data throughput requirements, lower latency requirements, and have traditionally been served by specialized appliances that raise the TCO and limit the flexibility of the solution. Moving to a distributed access architecture, virtualization, the power of general-purpose servers, and the growing market for programmable switches will reduce the need for such legacy solutions.

In the past, it may not have been possible from a technical or economic standpoint to have flexible compute resources at the Nodes or smaller huts and cabinets, but this is changing. The expanded availability of NEBS-compliant servers and industry innovations like the SCTE Generic Access Platform (GAP) will bring down the cost of putting small-form-factor servers at space and power-constrained locations.

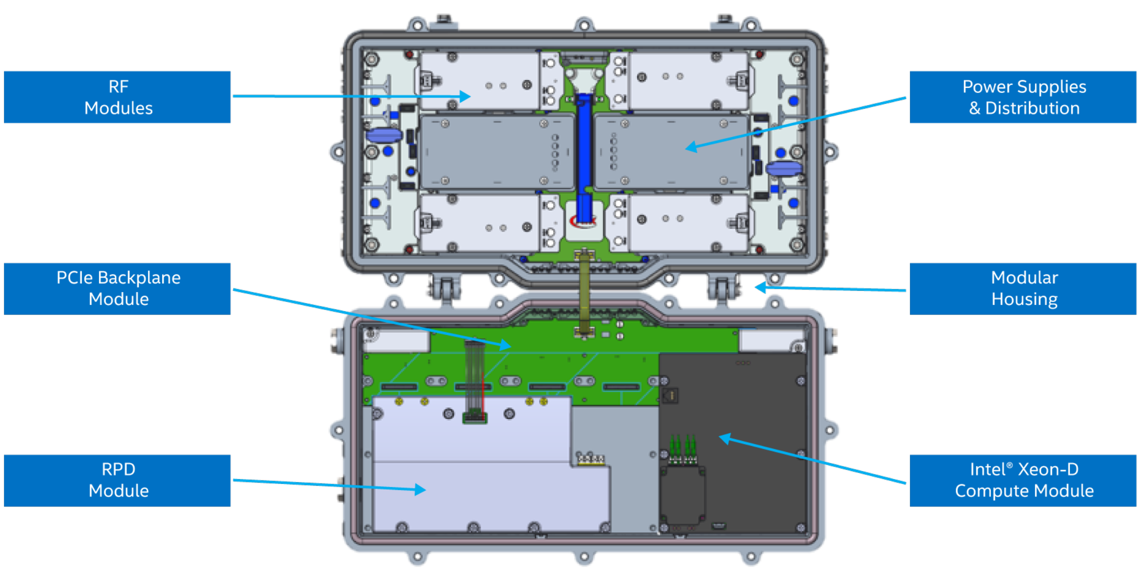

Figure 13. Generic Access Platform with Compute Module

Figure 13 shows a prototype of a GAP-compliant node in which the form factor, electrical and logical connectivity, and module management are standardized such that a vendor can provide the same types of compute, storage, and network capabilities found in a data center. The scale might be different—in the hybrid fiber-coaxial (HFC) case, the node may serve only a couple of service groups—but the architecture and the way its resources can appear to the larger management infrastructure of the network operator are very much the same.

What developments like GAP mean is that it is now technically possible and economically feasible to distribute flexible resources to all parts of the network. For the cable MSO with a lot of unique real estate and right-of-way investments, this is a powerful competitive advantage.

5. Convergence in the Network

Once an edge architect has determined its optimal locations, there is still the choice of which equipment and software to deploy and the subsequent business arrangements discussed earlier. If a CoSP seeks to cost-effectively develop and scale more than one access technology and service infrastructure for the edge, converging workloads onto a common platform can be beneficial.

CableLabs identified a Converged Network Architecture Framework that defines the different types of convergence that apply to the 10G network:

- Access Convergence
- Transport Convergence
- Platform Convergence
- Core Convergence
- Operations Convergence

Edge platforms can be implemented across all these domains. The key challenge is designing infrastructure to enable a seamless user experience across all access types, having a limited set of common hardware, utilizing the efficiency of cloud technologies, consolidating services management, and easing operations with telemetry and automation. On top of that, the edge platform needs to be a carrier-class solution, computationally lightweight and efficient, high performance, and have facilities for optimized life cycle management.

In the past, it was not possible to set up a common hardware and software infrastructure to meet the needs of all areas of a network, support multiple access technologies, and offer multiple services in more than one domain. Figure 14 shows how technical solutions associated with various business owners were set up in silos, and thus were developed, deployed, and managed independently. In this scenario, each solution is bespoke, and in a world where only one or two solutions was needed at a time, it was enough and still cost-effective to develop such independent systems and institutional expertise.

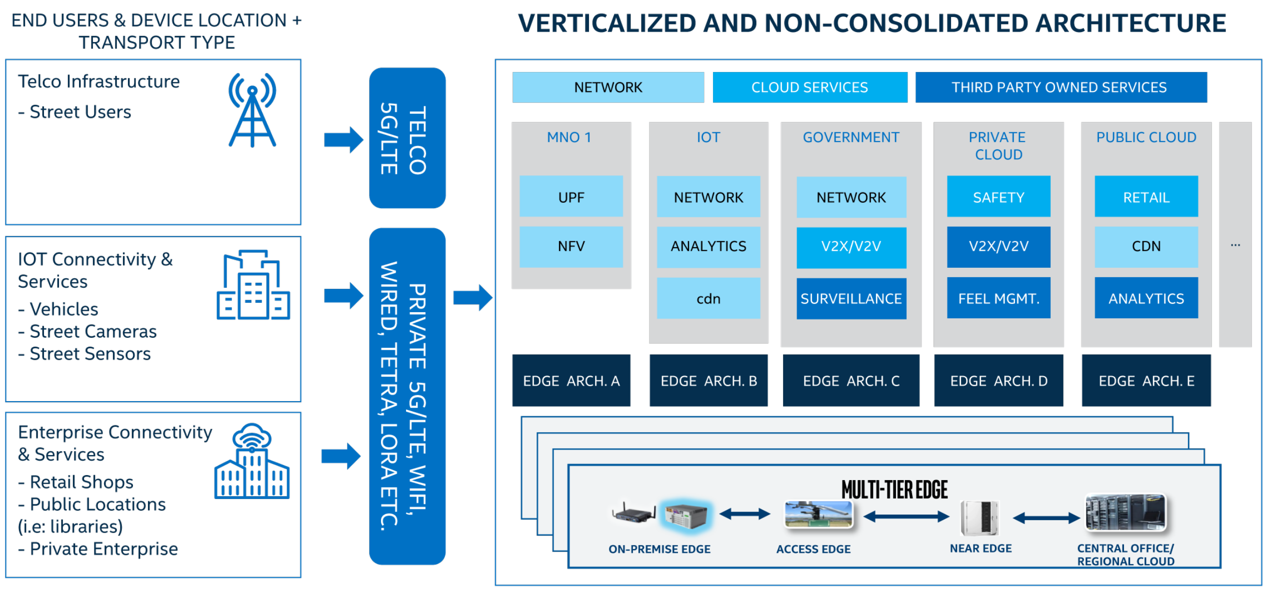

Figure 14. Access and Service Infrastructure in Silos

However, the future 10G network needs to accommodate many wired and wireless access technologies, and a host of new services. Consequently, the previous way of architecting networks will not meet modern economic or operational needs. What ameliorates this challenge is the rise of virtualization, containers and cloudification, new standards for power management, telemetry, slicing the network for different service tiers, and the resulting ability to converge a multitude of workloads on the same standards-based servers and switches. Yet there still needs to be a framework or set of building blocks to put everything together in a way that scales across all the locations in the network, from large data centers to the edge.

Several industry efforts aim to accelerate the development of edge platforms, including OpenNESS, CNTT Edge, Project Adrenaline, OpenVINO, and Open Visual Cloud. Some of these efforts have a relatively broad scope, for example, to move software infrastructure for the cloud to what is presumed to be a scaled-down platform for the edge. Other efforts address the needs of specific domains (i.e., visual processing). However, to deliver on all aspects of network and edge application convergence on a single platform, a framework that aids both hardware and software design using scalable building blocks across any network edge location is required.

With that in mind, Intel has been working on ways to unify and converge access, IoT, and other edge appliations on standards-based hardware and software for any location and for any set of functions.

The approach assumes the hardware can be constructed using components that provide common features important for edge deployments across the full range of performance needs and power constraints found in an operator's network. For an edge platform, these base features need to include easy and performant virtualization, large software support across many vertical domains, extendibility though accelerators, strong security capabilities, and functions for real-time machine learning algorithms. Silicon supporting x86 architecture satisfy all these criteria today, with headroom as other options emerge.

Possible hardware configurations include:

- Headend: a rack of x86-based 1RU or 2RU servers with a programmable top of rack switch
- Node: a small form factor x86 server for a standardized GAP enclosure
- Outdoor uCPE: a small form factor x86 server in a custom ruggedized enclosure

In all these cases, a common software management infrastructure identifies the compute, storage, and networking capabilities and connectivity of each location and orchestrates access functions and services according to defined service level agreements. Figure 15 shows the high-level design to converge multiple workloads onto a single software infrastructure. It consists of a sub-infrastructure to host the data plane functions for access technologies, sub-infrastructure to host services, a transport/switching infrastructure to move data from hardware (e.g., NICs) to the dataplane or services or between the dataplane and any of the services, and an infrastructure for coordinated deployment, orchestration, and management of all elements therein.

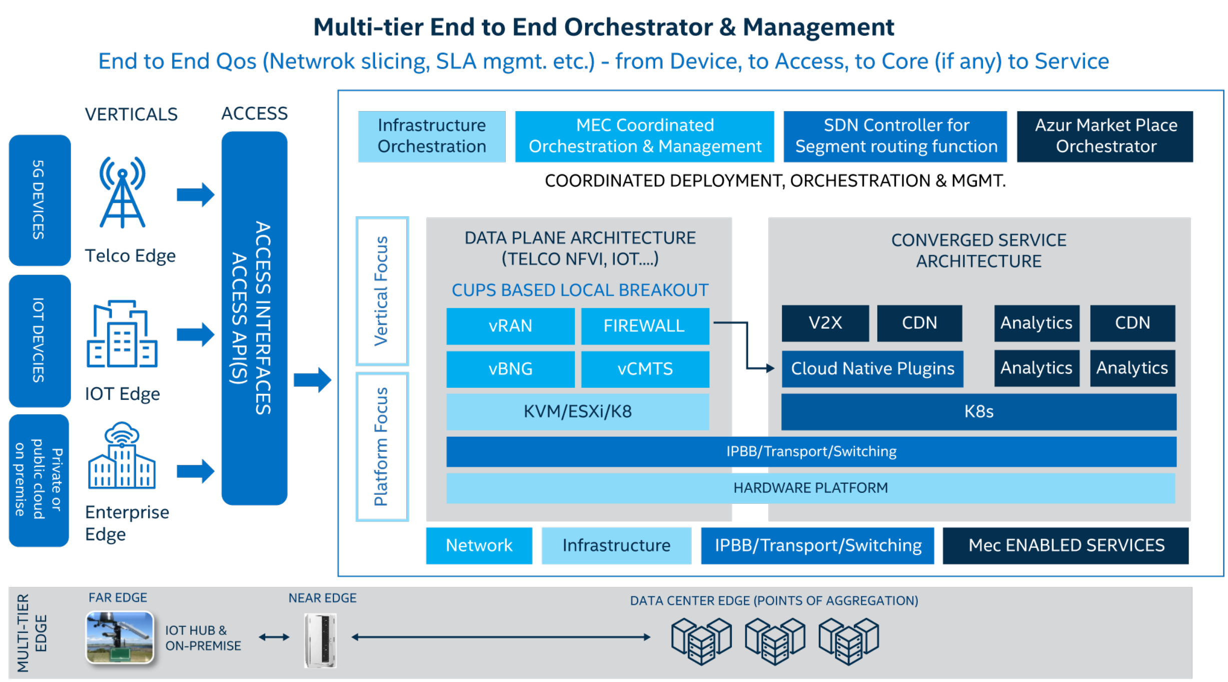

Figure 15. Software Architecture for Converged Workloads

Figure 16 shows the next level of granularity for types of software components that can be implemented within the software architecture. A solution would include at least one option from each row, though it would be common to include multiple elements. For example, a smart cities platform may have both Wi-Fi and HFC access technologies, supported along with two or more frameworks to perform local video analytics and execute action plans.

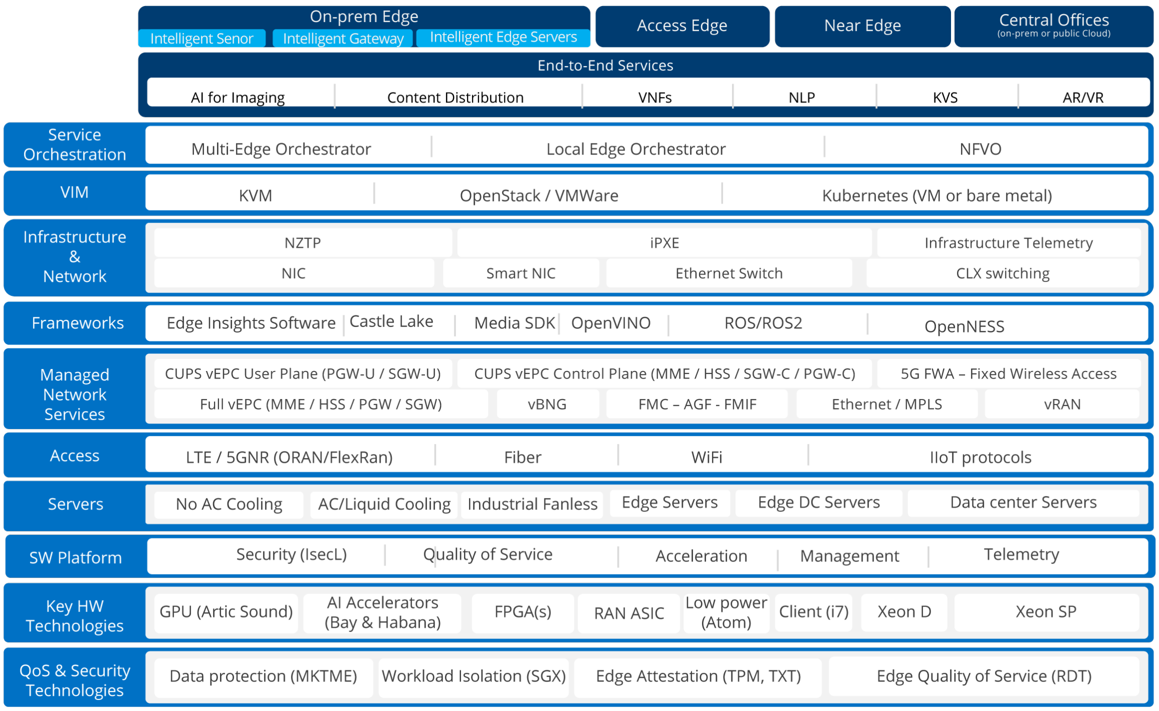

Figure 16. Mix and Match for an Edge Platform

Architectures should be designed and adapted to meet the specific requirements for on-prem and network edge solutions and to use the latest capabilities of new hardware and software, with a goal to develop scalable, flexible, platforms for wherever they may be needed across the network.

6. Considerations for Designing an Edge

The MSO network of the 10G era is multi-access, allows next-gen services in a wide range of performance tiers, and has the flexibility to deploy capabilities where they are needed in the network to deliver on business goals. Some of these new services require lower latencies or tighter controls around data sovereignty, which often means considering capabilities of facilities closest to the end users.

6.1. A Summary of the Options

The edge is not necessarily constrained to one location, platform, or business arrangement. At the top level, an operator can plan for an on-prem edge, network edge, or both. The on-prem edge implies that equipment at the customer site, like a uCPE, will run one or more services locally while being managed and controlled centrally in the network. The network edge is based on hosting access and services from equipment owned and operated by the CoSP, partner CSPs, or other third parties in various arrangements of a network edge platform architecture:

- CoSP + CSP Co-location
- CoSP Led + CoSP/CSP Services
- CSP Led
- CoSP/CSP Aggregator

The platform architecture for each of these is made of a limited set of hardware (i.e. if virtualization and programmable components are involved) and software infrastructure to manage different elements of the edge solution as well as the interface(s) to outside network elements. The details are in an earlier section, but the difference between the architecture types boils down to several questions including:

- Who owns the real estate?
- Who owns the physical equipment?
- Who owns the software infrastructure(s)?
- Who owns the data/customer?

Generally speaking, ownership gives more opportunity for monetization, but requires more institutional knowledge in the domain at hand to make sure the technology delivers the desired results.

The MSO network, with years of developing and deploying an HFC plant, is in a unique position to "own the real estate" where the edge platform equipment is hosted, to provide the best latency and data locality profiles for new services. The MSO/CoSP has the expertise to design and deploy whatever access elements are required for the edge solutions, although partnerships may be involved when new access technologies, like a 5G wireless service, are added to the network.

Where it gets more interesting—and where a lot of innovation and experimentation is happening—is in answering the rest of the questions, as they relate to providing services. Default behavior might be to bring in a CSP or third party to host and manage their services over the last mile broadband connection provided by a CoSP. This is generally because cloud technologies used in this case might be outside of the core competencies of the CoSP, so partnering with a CSP is the most straightforward way to monetize the aforementioned real estate advantages.

In the CoSP + CSP Co-location model, the CSP physically houses their equipment in the same location as the edge location of the CoSP, be it a Headend, Hub, or Node. The CSP-Led model is similar, except that CSP equipment resides in a point of presence near-to, but outside of, the CoSP Edge location. In these cases, the CSP and/or third parties making use of CSP resources "own the physical equipment, software infrastructure, and data" and therefore the customers for the services being sold. They could even host services that the CoSP wants to provide like localized SD-WAN offerings for small and medium businesses.

CoSPs wanting to follow the CoSP Led + CoSP/CSP services model and offer their own edge infrastructure to host services (i.e. to own the equipment and software infrastructure) will have to develop or hire their own expertise in cloud technologies. This might seem like a difficult and far-out proposition to those accustomed to single function appliances in their network, but as access workloads get virtualized (ex. vCCAP, vBNG, vRAN, etc.) the technologies to manage and deploy both access and services are starting to converge. The final alternative is to have an aggregator with a local point of presence host the equipment and software infrastructure and sell broadband and potentially services (along with perhaps similar offerings from competitors).

A key competitive advantage for the MSO is that it has invested in and has rights of way for edge-friendly locations to host edge platforms for either itself or for a CSP. Even though these locations range in physical, power, and other environmental constraints, the use of virtualization, telemetric capabilities, facilities for remote security, and software infrastructure for managing distributed computing elements make it possible to consolidate the number of platforms into the minimum possible. Less disparate architectures and technological domains provide a reduction in the total cost of ownership across the network.

The ability to mix and match common hardware and open source software elements through a common framework allows the industry to construct scalable and flexible platforms that match the performance, lifecycle, and form factor requirements for any edge location. Commercial examples of industrial on-prem, 5G vRAN, and other solutions prove the advantage of converging edge applications on standard hardware and using reference software architectures to lower costs and speed time to market.

6.2. Asking the Right Questions

Admittedly, it will take more than a checklist to consider all the options above and architect an edge location in a given network. But a few key decisions that will drive the planning and architecture are:

- What type of services do you want to offer and what requirements do they have on the network?
- What business models / partnerships do you want to support – who owns what?
- Where are you willing to deploy equipment / functions / infrastructure?
- What equipment and software infrastructure can be consolidated across the network?
- Who is going to own the various parts of the Edge solution in the organization?

The last question about organizational ownership may be the hardest as “the edge” crosses what were typically separate domains—multiple access technologies, enterprise services, and data center and cloud resources. Yet it is because of this breadth that edge is important for MSOs to grow their businesses, whatever options are chosen. Thinking about a grand edge strategy may be the chance for an operator to disrupt existing silos, align on the latest technological innovations, and be a distinguishing factor against the competition.

Practical realities to leverage existing systems, skillsets, and business relationships will make this an iterative process. It is advisable to develop a vision for the ultimate end-state for the network and the organizations supporting it, but the transformation may be accomplished in stages. For example, would implementing a CSP-Led model first allow a CoSP to get into the market and show the value of its edge locations while simultaneously developing internal cloud expertise and driving the ecosystem to consolidate access and services onto the same servers and switches?

Regardless of the path, the time is now to get started on this journey. MSOs can leverage their unique infrastructure and real-estate investments to provide improved network visibility, performance, control, flexibility, and agility with a distributed compute architecture all the way to the edge—wherever it may be.

Abbreviations

API	application programming interface
CERA	Converged Edge Reference Architecture
CPE	customer premise equipment
CSP	cloud service provider
CoSP	communications service provider
CNF	cloud native function
DAA	Distributed Access Architecture
GAP	Generic Access Platform
ISV	independent software vendor
KPI	key performance indicator
MSO	multi-service operator
NFVI	network functions virtualization infrastructure
ODM	original design manufacturer
OEM	original equipment manufacturer
OTT	over-the-top
PAAS	platform-as-a-service
SD-WAN	Software-defined wide area network
SCTE	Society of Cable Telecommunications Engineers
TEM	telecommunications equipment manufacturer
uCPE	universal customer premise equipment
VNF	virtual network function

Bibliography & References

Gartner, "Edge computing promises near real-time insights and facilitates localized actions."; https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders

Xu et. al.,"Edge Intelligence: Architectures, Challenges, and Applications", https://arxiv.org/abs/2003.12172

FCC, "2016 Measuring Broadband America Fixed Broadband Report"; https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-fixed-broadband-report-2016

Fierce Wireless, "Verizon, AWS bring 5G MEC to Boston, Bay Area"; https://www.fiercewireless.com/operators/verizon-aws-bring-5g-mec-to-boston-bay-area

Mann, "Coronavirus Cleaves SD-WAN Revenue Growth"; https://www.sdxcentral.com/articles/news/coronavirus-cleaves-sd-wan-revenue-growth/2020/06/

RTTNews.com, "Microsoft To Use Telefonica Infrastructure For Datacenter Region In Spain" ; https://www.nasdaq.com/articles/microsoft-to-use-telefonica-infrastructure-for-datacenter-region-in-spain-2020-02-26

Robuck, "Cox targets the edge for the next evolution of network performance and security", https://www.fiercetelecom.com/telecom/cox-targets-edge-for-next-evolution-network-performance-and-security

Dano, "SBA, American Tower double down on edge computing opportunity"; https://www.lightreading.com/the-Edge/sba-american-tower-double-down-on-Edge-computing-opportunity/d/d-id/762941

Levensalor, Stuart, "The Modular, Virtualized Edge for the Cable Access Network"; https://community.cablelabs.com/wiki/plugins/servlet/cablelabs/alfresco/download?id=2c46cef2-af44-47be-bdd4-98a948cbc60d

SCTE Generic Access Platform; https://www.scte.org/generic-access-platform/

Open Network Edge Services Software; https://www.openness.org/

CableLabs Project Adrenaline; https://www.cablelabs.com/edge-adrenaline-boost-kubernetes-orchestrate-fpgas-gpu

OpenVINO toolkit; https://docs.openvinotoolkit.org/

Open Visual Cloud; https://01.org/openvisualcloud

"Converged Edge Reference Architecture (CERA) for On-Premise/Outdoor", https://builders.intel.com/docs/networkbuilders/converged-edge-reference-architecture-cera-for-on-premise-outdoor.pdf

