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**

1. Introduction 3
2. Edge Deployment Models 5
   2.1. On-Premise Edge 6
   2.2. Network Edge 7
3. Network Edge Platform Architectures 9
   3.1. CoSP + CSP Co-location 9
   3.2. CoSP Led + CoSP/CSP Services 11
   3.3. CSP Led 12
   3.4. CoSP/CSP Aggregator 13
4. Mapping to Real Estate 14
5. A Converged Edge Architecture 16
6. Considerations for Designing an Edge 20
   6.1. A Summary of the Options 20
   6.2. Asking the Right Questions 22

Abbreviations 23

Bibliography & References 24

List of Figures

Title **Page Number**

Figure 1 - Requirements on the Network for New Services 3
Figure 2 - Use Cases and Associated KPIs 4
Figure 3 - Logical Parts of the Communication Service Provider Network 4
Figure 4 - Edge Locations and Terminology 6
Figure 5 - View of an On-Premise Edge Deployment 7
Figure 6 - View of a Network Edge Deployment 8
Figure 7 - Communications Service Provider and Cloud Service Provider Co-location 10
Figure 8 - Communications Service Provider-Led Edge Deployment 11
Figure 9 - Communications Service Provider-Led Edge Deployment 12
Figure 10 - Communications Service Provider Aggregator-Led Edge Deployment 13
Figure 11 Type of locations in the MSO Network 14
Figure 12 - Network Edge Location Characteristics 15
Figure 13 - Generic Access Platform with Compute Module 16
Figure 14 - Access and Service Infrastructure in Silos 17
Figure 15 - Converged Edge Architecture 19
Figure 16 - Mix and Match for an Edge Platform 20

1. Introduction

Cable network bandwidth demands are growing exponentially as video becomes ubiquitous, Internet of Thing (IoT) devices proliferate, and new high bandwidth 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.

That said, there is an incredible opportunity for broadband connectivity providers and those offering over-the-top (OTT) applications and Services to help make sense of and take action on the data coming from cars, cameras, factories, enterprises, and homes, and to do so in a timely manner. In fact, whole new categories of Services have been dreamed up, 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 new network and compute capabilities if they can be delivered in a cost-effective manner.

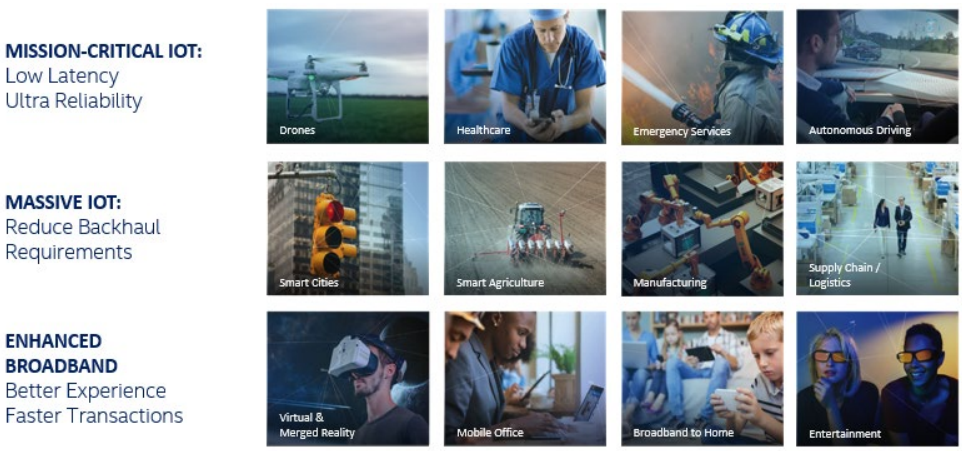

Figure 1 - Requirements on the Network for New Services

For many networks, the latencies and other key performance indicators (KPIs) for specific Services within these segments, as shown in Figure 2, will require smart 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

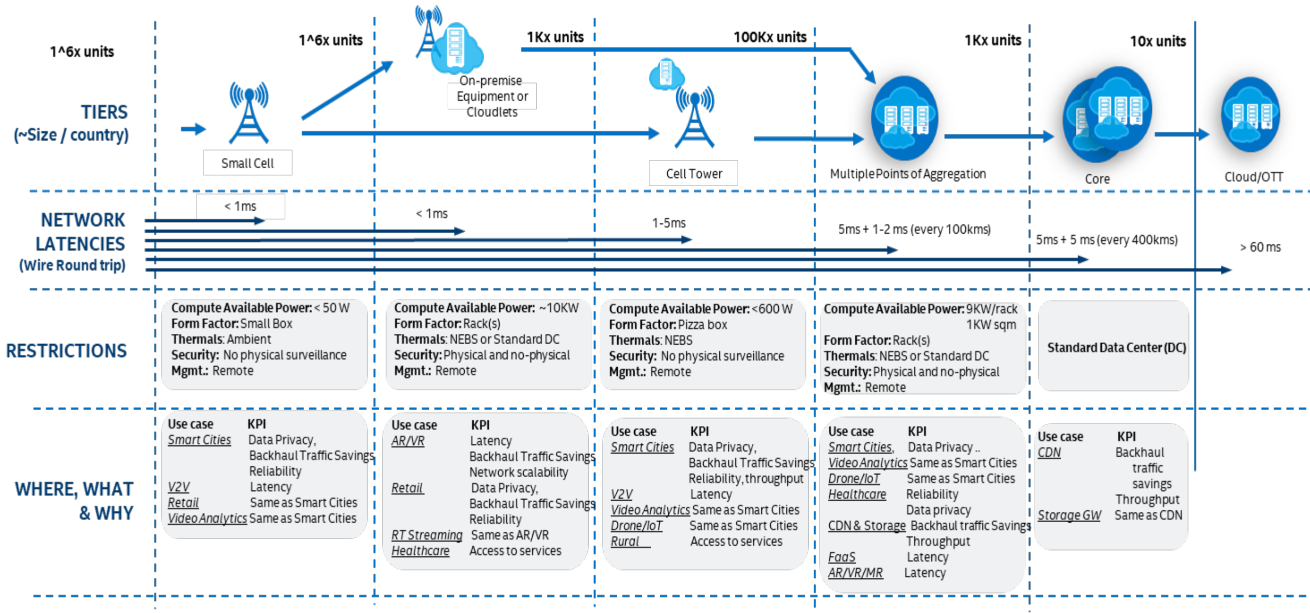

Figure 2 - Use Cases and Associated KPIs

Looking at the topology of the network, "The Edge" makes for an interesting and obvious place to manifest key infrastructure because it is physically closer to end users. The reality is there are multiple places that could be considered the Edge, and they all can exist in the same network. Figure 3 shows that at the top level, the Edge can be split into an On-Premises Edge and a Network Edge. The next section will go deeper into defining 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).

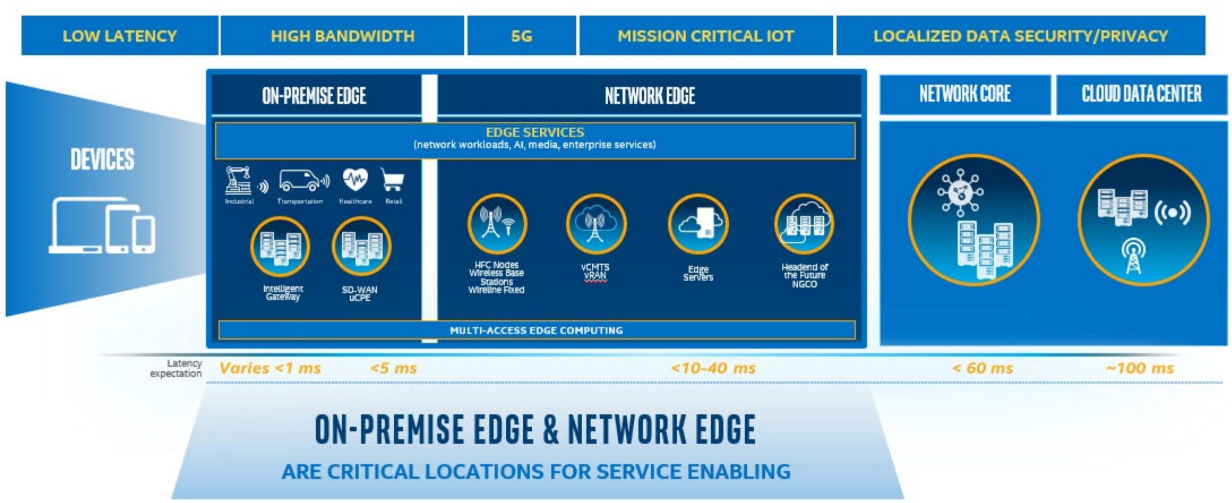

Figure 3 - Logical Parts of the Communication Service Provider Network

Figuring out where network infrastructure may be hosted is only part of the battle. This paper will continue on to discuss considerations for designing and buying Edge platforms, as in the actual 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. Perhaps the finances work out for two or three solutions set up in parallel. But to scale Edge infrastructure and maximize resources and operational efforts, a common platform (or at least common building blocks) that addresses all Access technologies and Service deployments will make the most sense. To this end, this paper introduces the idea of a Converged Edge Architecture, comprised of commercial off-the-shelf (COTS) equipment and standard software frameworks and interfaces that can be used to develop any number of specific solutions. This approach reduces complexity and time-to-market and allows for common orchestration across any Edge location. Converged Edge Architecture does not define a single platform, but rather is a common framework that will deliver the right Edge platform for the functional needs and environmental constraints of a given location in the network.

There are many reasons the cloud is 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 last category – opportunities and considerations around implementing Services – though topics like virtualization and automation will be referenced as they underlie modern network architecture and implementation.

In fact, recent headlines[2] have shown that some communications service providers (CoSPs) and cloud service providers (CSPs) are already implementing Services at the Edge and selling them commercially. 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 both 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(s), as it can be done in conjunction with the ongoing march towards a Distributed Access Architecture and a virtualization environment in its many forms.

By asking the right questions and knowing some of the key architectural options discussed in this paper, network operators – and MSOs in particular – will be in a better position to realize the value of their existing infrastructure and improve customer experience in an increasingly Edge-focused world.

2. Edge Deployment Models

Figure 4 breaks down the taxonomy of Edge platforms and locations beyond the top-level of an On-Premises or a Network Edge. The On-Premises Edge, or "On-Prem Edge," can be broken down into platforms/locations for smart sensors, intelligent gateways, and intelligent Edge servers, with generally increasing complexity, power, and capability as one moves left to right. Similarly, the Network Edge contains platforms and locations at the Access Edge, near Edge, and data center Edge.

Note that some networks may not manifest all these categories because the physical real estate to host certain equipment might not exist, locations and functions may have been consolidated onto fewer

[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

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 Edge types or Network Edge types exclusively, as they each have their own pros and cons with respect to achieving different types of technical and business goals.

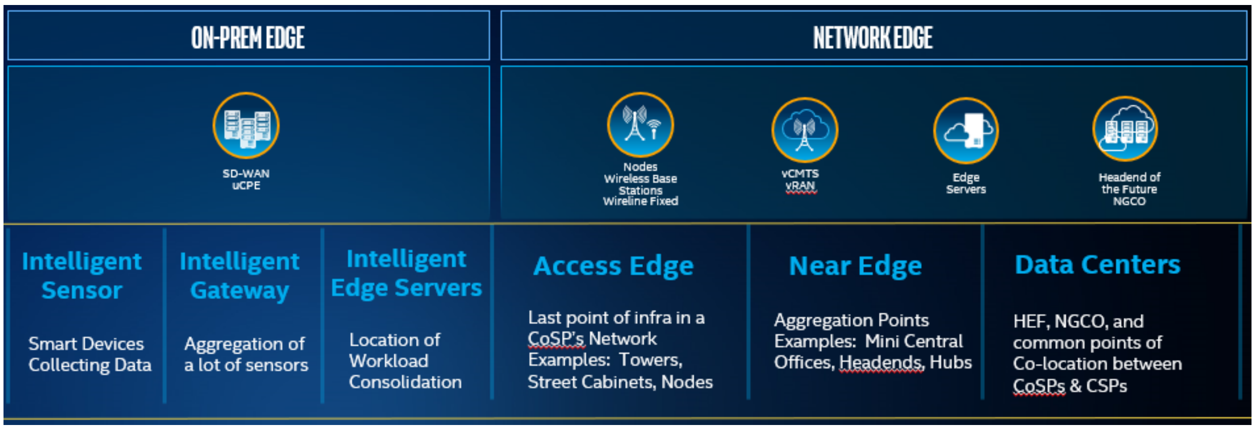

Figure 4 - Edge Locations and Terminology

The following sections cover these deployment models in turn.

2.1. On-Premise Edge

Figure 5 shows an On-Prem Edge-type deployment model. Here, there is a controller for Services that is located at some centralized place in the network and manages functions and Services that ultimately run on an Edge platform, like a universal customer-premises equipment (uCPE). Non-real-time functions such as controllers and Service management will usually reside in the most cost-effective place, which is typically deep in the network at a regional data center or even in a public cloud. That said, the controller may absolutely be deployed at another Network Edge location, like a Headend or Hub, to comply with legal requirements, or perhaps to satisfy operator or customer requirements for full data locality.

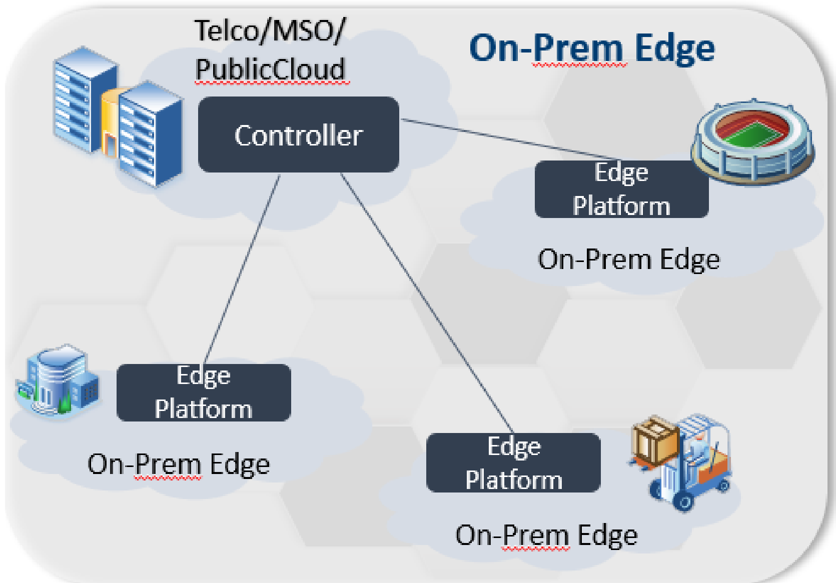

Figure 5 - View of an On-Premise Edge Deployment

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 it includes 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 are comprehended in the Converged Edge Architecture framework introduced in Section 5.

2.2. Network Edge

Figure 6 shows a Network Edge-type deployment model. In this case, there is also a controller – or more likely, a set of controllers – within the CoSP hierarchy to control network function virtualization infrastructure, multiple VNFs, and/or Services hosted directly by the operator or its partners. These controllers will orchestrate and manage such functions to run on Edge platforms located generally at Nodes, Hubs, Headends, point-of-presence server locations, or central offices. As in the On-Prem model, the controllers could exist anywhere “upstream” from Edge platforms in the network; it is likely that they will run from a regional data center, private cloud, or at a CSP.

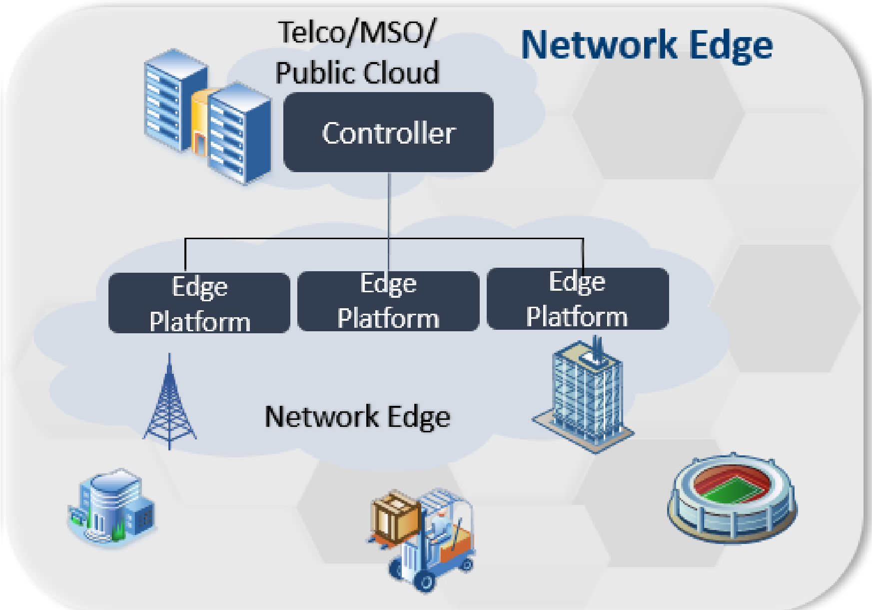

Figure 6 - View of a Network Edge Deployment

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 all these possibilities, the Edge platforms in Figure 5 are not necessarily a single server or appliance but can manifest as a collection of equipment at 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 make the most of a Network Edge deployment and to keep costs down as the number of Access types and Services increases, it will be key to find as much common ground as possible between the different hardware elements and to provide a homogeneous software layer for management. These are some of the main goals of the Converged Edge Architecture effort covered later.

Many of the platform-as-a-service products that have been announced[3] by various CoSPs, CSPs, Edge compute specialists, and real estate management companies (i.e., towers) are based on a Network Edge deployment model. Behind the scenes, there are a variety of ways these parties can organize themselves and their resources to make their own Edge products. They will ask themselves – what network functions and Services should we offer (and who should own these products), where should the equipment exist, who should own the customers, how should revenue be divided, and so forth. The answers to these questions have both business and technical repercussions, and for the latter, will affect how the Edge platforms/locations can be architected. The next section describes several Network Edge platform architectures emerging in the marketplace and the associated advantages and disadvantages from the network operator's perspective.

---

[3] 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

3. Network Edge Platform Architectures

The On-Prem network deployment model has its own implementation challenges, but at a high level, network functions and Services will run on a uCPE at the customer premise with a software controller running deeper in the network on COTS servers and managing those Services.

As discussed earlier, the Network Edge case has many options for splitting up functions across different network 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 architecture.

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

1. CoSP + CSP Colocation
2. CoSP Led
3. CSP Led
4. CoSP Aggregator

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

3.1. CoSP + CSP Co-location

Currently, the most popular, or at least the most talked about, architecture for a Network Edge deployment is for a CSP to co-locate equipment to deliver their Services at Edge locations owned by the CSP. Figure 7 shows the logical view of how this would work. Presumably, the 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 being deployed at a regional data center or at their own site.

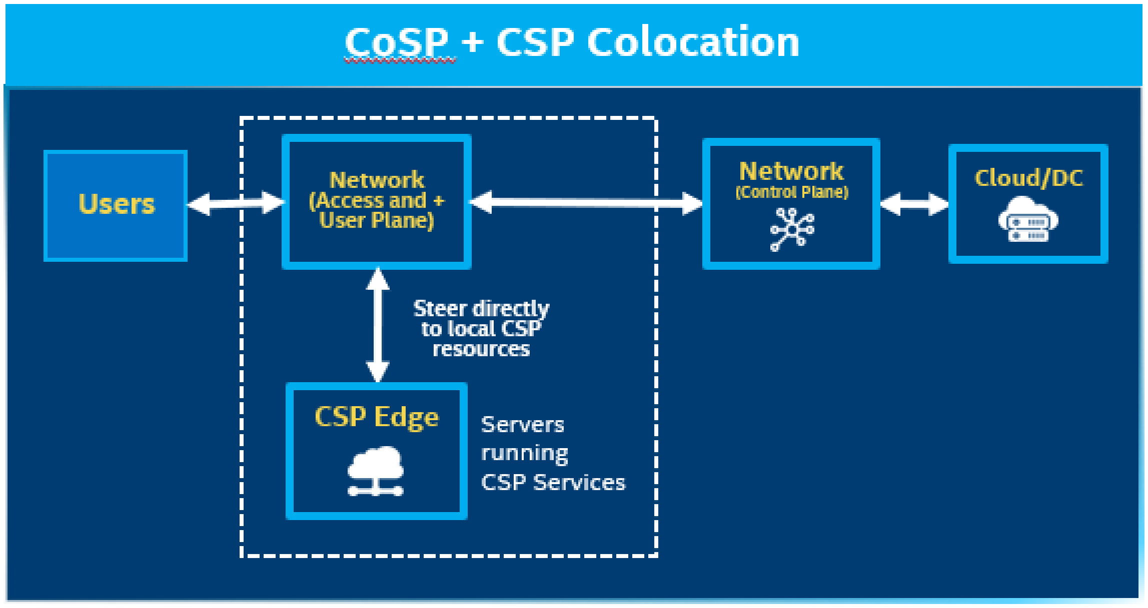

Figure 7 - Communications Service Provider and Cloud Service Provider Co-location

The most straightforward benefit for the CoSP is it gets some sort of "rent" or revenue share from the CSP. More interesting for business development is that these partners can engage in joint marketing to advertise the CSP is offering Services over the Access medium owned by the CoSP that would not available through other means – a "better together" story. There may also be practical reasons to enter into this type of agreement as the CoSP may not have the expertise or desire to develop and manage whatever Services are being provided by the CSP.

From the CoSP perspective, the downside is that it does not have direct control on how co-located Services are monetized, and, further, would not have access to the user information or traffic telemetry for secondary opportunities for data monetization. In addition, there may be some inadvertent CoSP/CSP lock-, or at least constrain the relationship to one CSP at a time if the CoSP location is not set up to easily host and secure multiple parties.

For this model, the "Edge platform" is made up of two subsets of equipment hosted at the same Headend, Hub, or Node: one owned by the CoSP providing the network connectivity and the other owned by the CSP providing the Services. The CoSP and the CSP would select their individual equipment based on their own experiences and needs.

Note that the CoSP could also become a customer of the CSP and have them host new Services on their behalf, so this arrangement does not preclude the CoSP trying their hand at offering Services. It's just that the Cloud/Service infrastructure is being provided by CSP and getting paid for it. In this case, though, perhaps the co-location aspect of this arrangement would even allow knowledge transfer of Cloud technologies to the CoSP?

3.2. CoSP Led + CoSP/CSP Services

The second most common architecture is for the CoSP to own the Edge platform altogether, hosting CSP Services on it on behalf of the CSP. In this case, the CoSP must develop an API, allowing CSPs to access resources at advantageous Edge locations. With this architecture, the CoSP owns the execution and delivery of the Services to the customer. Figure 8 shows the logical diagram of the platform.

From the CoSP perspective, there is more complexity and internal expertise needed, as they need to develop and maintain their own commercial platform-as-a-service. CoSPs do not have ownership of the Services themselves but would be able to monetize the management of the Edge locations and equipment.

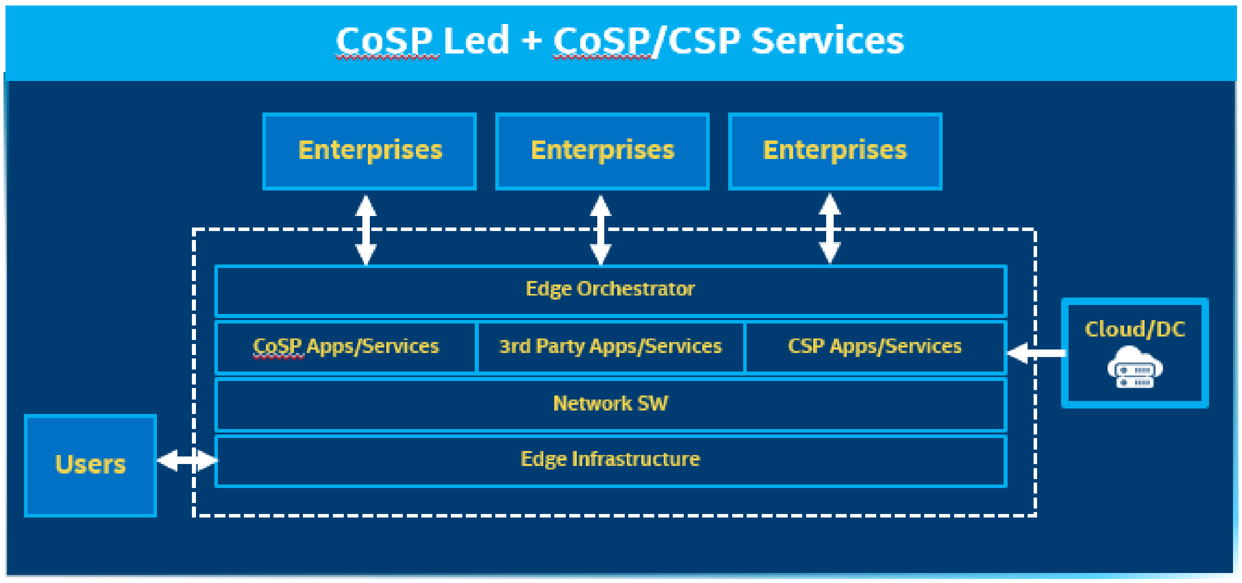

Figure 8 - Communications Service Provider-Led Edge Deployment

The additional benefits for the CoSP is it does not have to deal with the logistics of hosting another company's equipment in their buildings, and it has a unified interface to sell its Edge platform to any number of CSPs or third parties. Developing this model also implies the CoSP invested in internal resources that understand cloud technologies, which is generally a good thing as more elements of the network are virtualized.

For this model, the Edge platform is wholly specified and managed by the CoSP. There can still be multiple types of equipment involved in the final implementation, but a single owner is more likely to host all the different functions and Services on a common set of hardware in order to get better economies of scale for both the capital costs and for the ongoing costs towards the workforce. High performance VNFs running on COTS 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 to be presented may be less prevalent, but nonetheless 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 coming from the general cloud. In this case, the CSP uses a CoSP for last mile connectively, but otherwise owns all aspects of the Service delivery and management.

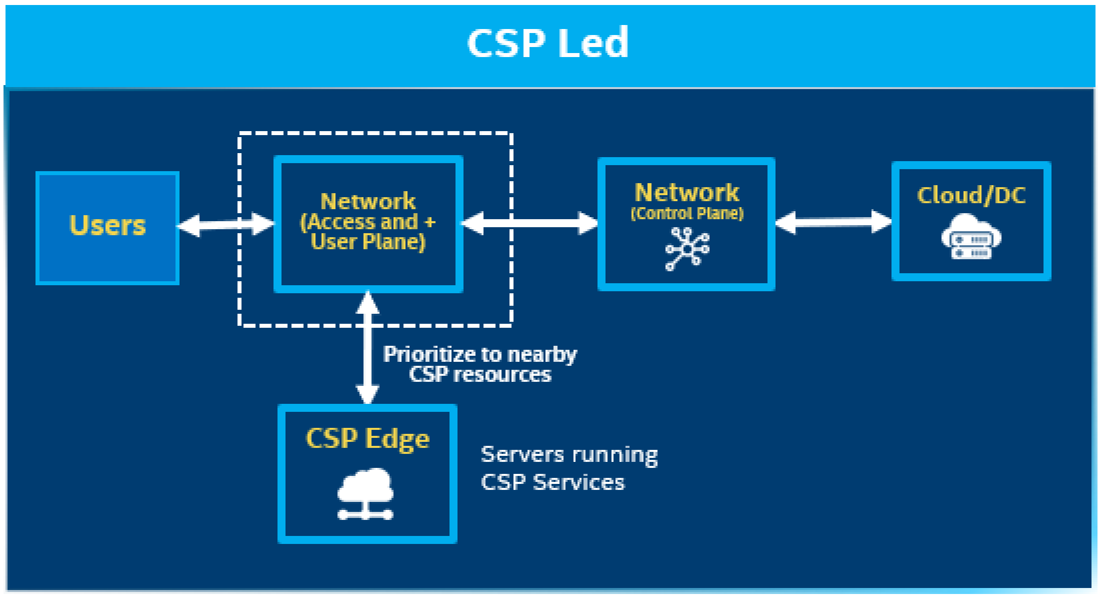

Figure 9 - Communications Service Provider-Led Edge Deployment

For the CoSP, this model may just be construed 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 partnerships with CSPs focused on Edge-specific infrastructure (i.e. that have developed, or can develop, many points of presence near last-mile CoSP locations) may actually make for a convenient alternative to the CoSP + CSP Co-location option insofar that a similar "better together" story can be created but without the added logistics and coordination of sharing the same physical space and Headend or Hub management[4].

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.

So, the CoSP's Network Edge platform is whatever the CoSP wants to use for its most advanced Access and broadband offerings and should follow the best practices of the industry (i.e. generally moving to DAA and virtualization-based solutions). Similarly, the Network Edge platform of the CSP will be optimized for the products and Services they are offering and generally will be a common software management infrastructure hosted on COTS servers and switches – typical for any CSP.

[4] 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

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

This 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 its own real estate, buys connectivity from one or more CoSPs, contracts with CSPs for Services, and perhaps even offers its own Edge Services. The result is an aggregated Edge offering.[5] 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.

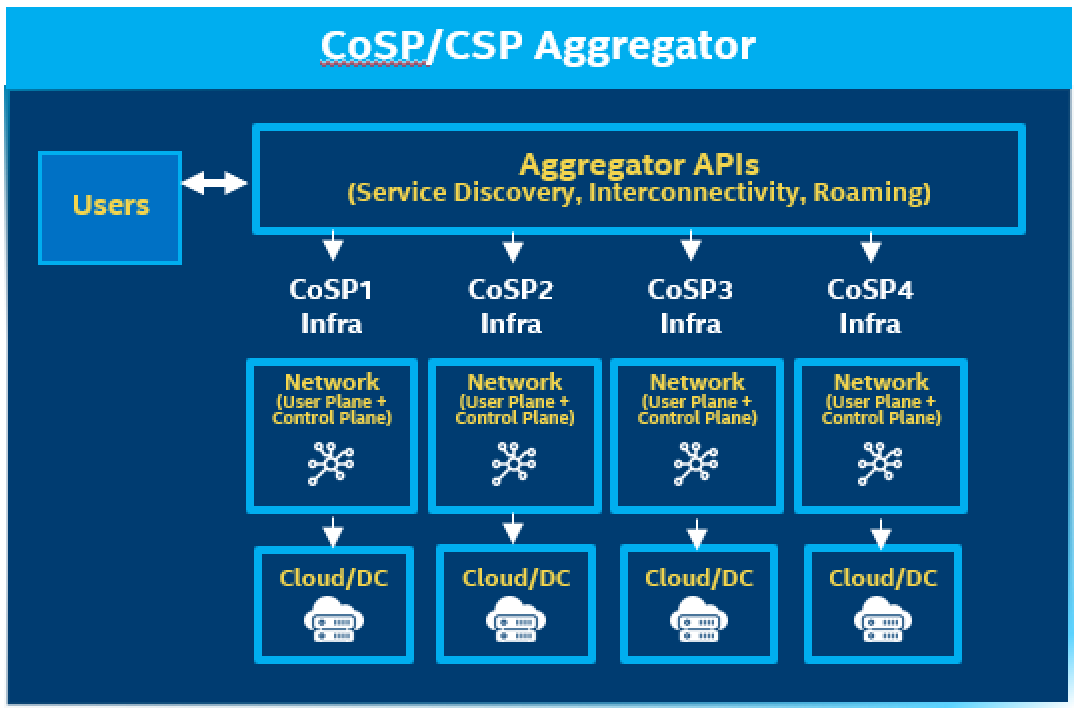

Figure 10 - Communications Service Provider Aggregator-Led Edge Deployment

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

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 (and sometimes baggage) of physical Access technologies and therefore it is easier to choose COTS hardware – servers and switches – for the reasons mentioned in previous sections.

[5] 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

4. Mapping to Real Estate

This paper has discussed both Edge deployment models and Edge platform architectures as logical entities, but in real life, the equipment, software, and operations to manifest these things need 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 Access technologies.

CableLabs recently described typical MSO network locations and how their characteristics could apply to Edge deployments.[6] Figure 11 shows how the following locations are connected and how far they are from end users:

- Cloud
- Regional Data Centers
- Headend/Hub
- Aggregation Nodes
- Child Nodes
- Customer Premises
- Cell Sites

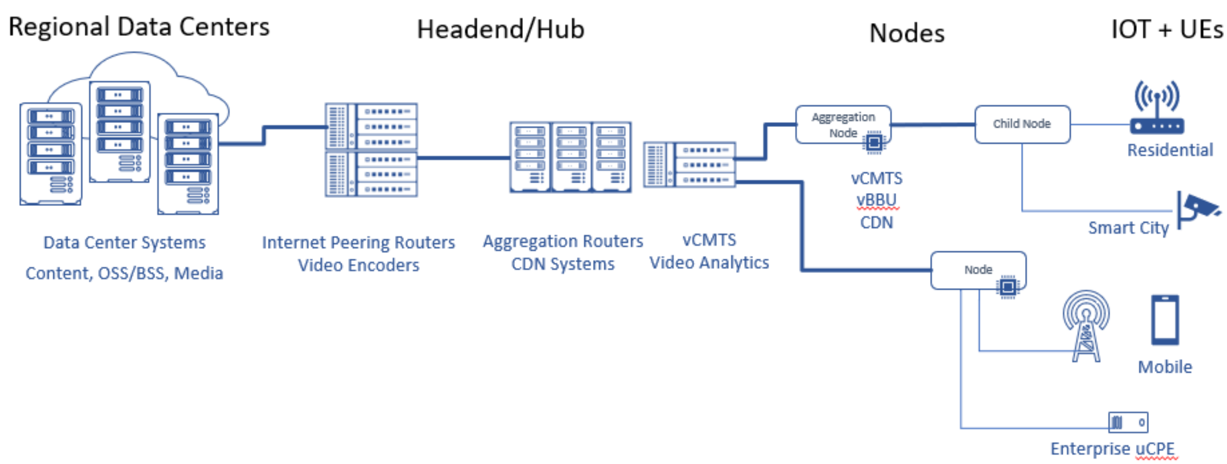

Figure 11 - Type of locations in the MSO Network

On-prem or Network Edge locations have associated benefits and constraints. On the benefits side, the closer one gets to the IoT 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 the IoT devices and users, the less power, the less environmental control, and the more costly it is to deploy and service equipment.

The benefits for On-Prem or Network Edge are straightforward to understand – “closer is better” for latency and data locality. The following discusses the constraints for the different types of host equipment that can be hosted.

[6] 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

The left side of Figure 12 shows a large, multi-story, regional data center (e.g., central office or CO) or large Headend, serving tens or hundreds of thousands of users, and for the most part, the set up can use typical data center approaches. Moving to the right along this continuum towards the users, the environmental constraints increase, the compute capacity goes down, and the deployment and management of equipment is more costly.

Figure 12 - Network Edge Location Characteristics

CoSPs need to decide if they want all the above locations to be "available" for their Edge. That is, will a given location and the equipment therein not only have the right level of physical connectivity in both directions, but will they also have 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, 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 the hardware is common, and the software has similar resource needs – as in data center and CSP locations. However, the further out one gets from the regional data centers, the more likely software workloads have higher data plane needs, lower latency requirements, and traditionally, have been served by specialized appliances. These types of devices will be part of an Edge plan for many operators, noting that fixed-function devices raise the TCO, and by definition, limit the flexibility of the solution.

As discussed earlier, the move to Distributed Access Architecture (DAA), virtualization, the power of general-purpose compute via COTS servers, and the growing market for programmable switches will reduce the need for such legacy solutions. Certainly, many Edge platforms residing in regional data centers, Headends, and Hubs will be based on COTS servers and switches.

Regarding the locations in the outside plant, 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 in the past, but this is changing. Operators still need to do their homework to understand what is possible based on their existing infrastructure and what types of businesses they want to pursue in the future, but 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 these types of locations.

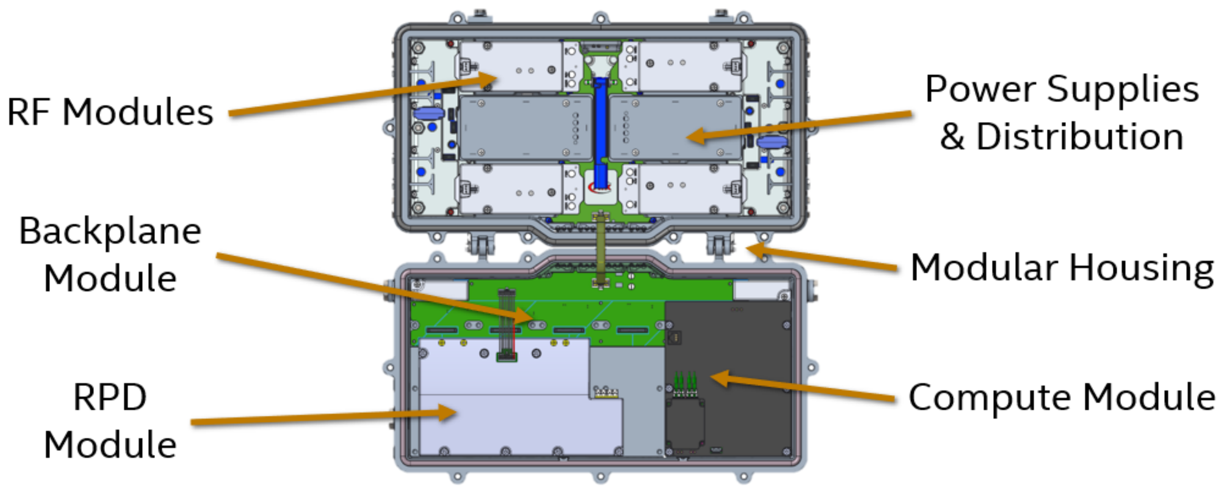

Figure 13 - Generic Access Platform with Compute Module

Figure 13 shows an example of a GAP-compliant Node in which the form factor, the 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 possible, from both technical and business perspectives, 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. A Converged Edge Architecture

Once an Edge architect has determined which locations can support Edge platforms, there is still the choice of which equipment and software to deploy to execute the desired On-Prem or Network Edge models 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, there must be some common platform to run everything.

The idea of minimizing the number of platforms comes under the banner of "convergence." 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

The "platform" or Edge platform that has been discussed in this paper is a concept that, in the ideal case, spans across all these domains. That is, how does one design Network Edge equipment and software infrastructure to enable a seamless user experience across all Access types, have a limited set of common hardware, take the best elements of the cloud world, consolidate management of Services, and ease operations with telemetry and automation? On top of that, the 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 they 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.

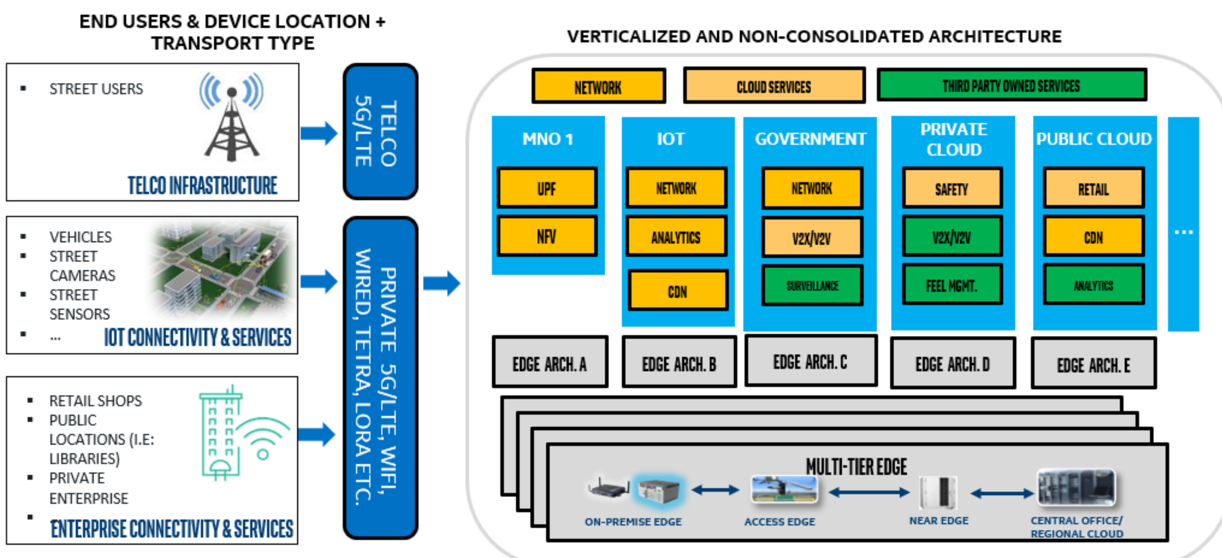

Figure 14 - Access and Service Infrastructure in Silos

However, the future 10G network needs to accommodate many wired and wireless Access technologies, and a host of new Services; and consequently, the old way of architecting networks is not going to work from economic or operational perspectives. What ameliorates this challenge is the rise of virtualization, containerization, and cloudification, new standards for power management, telemetry, and slicing the network for different Service tiers, and the resulting ability to converge a multitude of workloads on the same COTS equipment (i.e., standards-based servers and switches). But there still needs to be a framework or a set of building blocks to put everything together in a way that scales across all the locations in the network, from the large data centers to the Nodes.

There are already several industry efforts towards software platforms at the Edge, 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 focus only on addressing the needs of specific domains (i.e., visual processing). However, to really touch on all aspects of convergence, there needs to be a view and a framework that can aid both hardware and software design using scalable building blocks so solutions can be deployed in any part of the Edge network.

With that in mind, researchers at Intel, led by Francesc Guim, Principal Engineer, and Timothy Verrall, Senior Principal Engineer, for the Intel Edge Architecture Group,[7] have been contributing to the development of a framework called the Converged Edge Architecture. The objective of this effort is to unify and converge Access, IoT, and other workloads on standards-based hardware and software. This Converged Edge Architecture is not a specific piece of hardware or software, but rather it is a set of building blocks to create a "Plug and Play" Edge platform that is relevant to the specific requirements of CoSPs for any location and for any set of functions.

The base of the Converged Edge Architecture framework 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 moving forward, functions for real-time machine learning algorithms. Silicon supporting x86 architecture satisfy all these criteria today, but in the future, other options may emerge.

Possible hardware configurations for a Converged Edge Architecture-based solution are:

- 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 for Converged Edge Architecture 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.

[7] Francesc Guim, Principal Engineer, Timothy Verrall, Senior Principal Engineer Intel Edge Architecture Group, 2020.

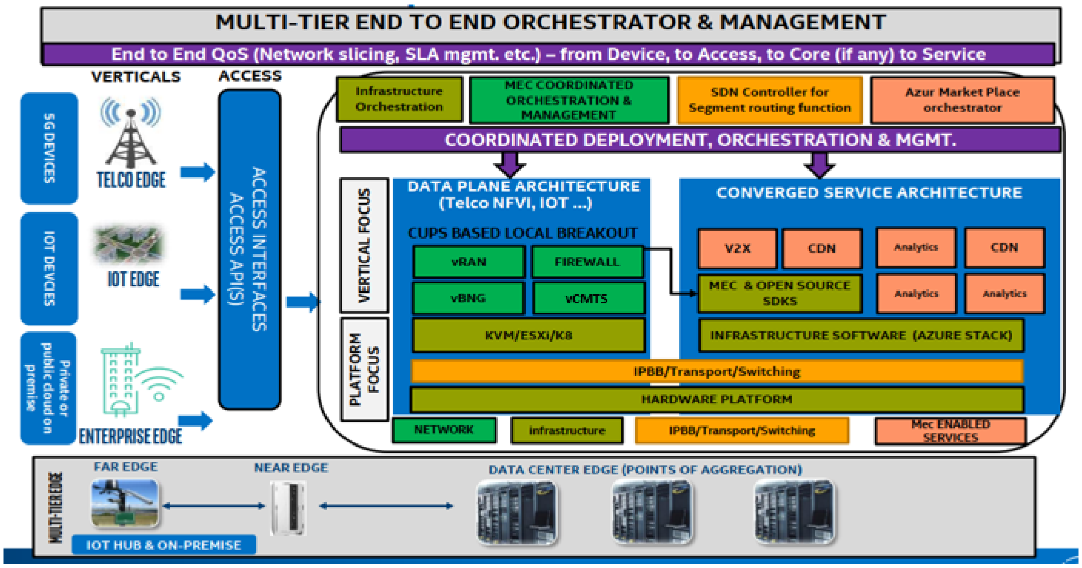

Figure 15 - Converged Edge Architecture

Figure 16 shows the next level down of granularity for types of software components the Converged Edge Architecture framework considers within the previously mentioned sub-infrastructures. Also shown is a sampling of candidate options from the market, including some of the familiar names in the industry. 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.

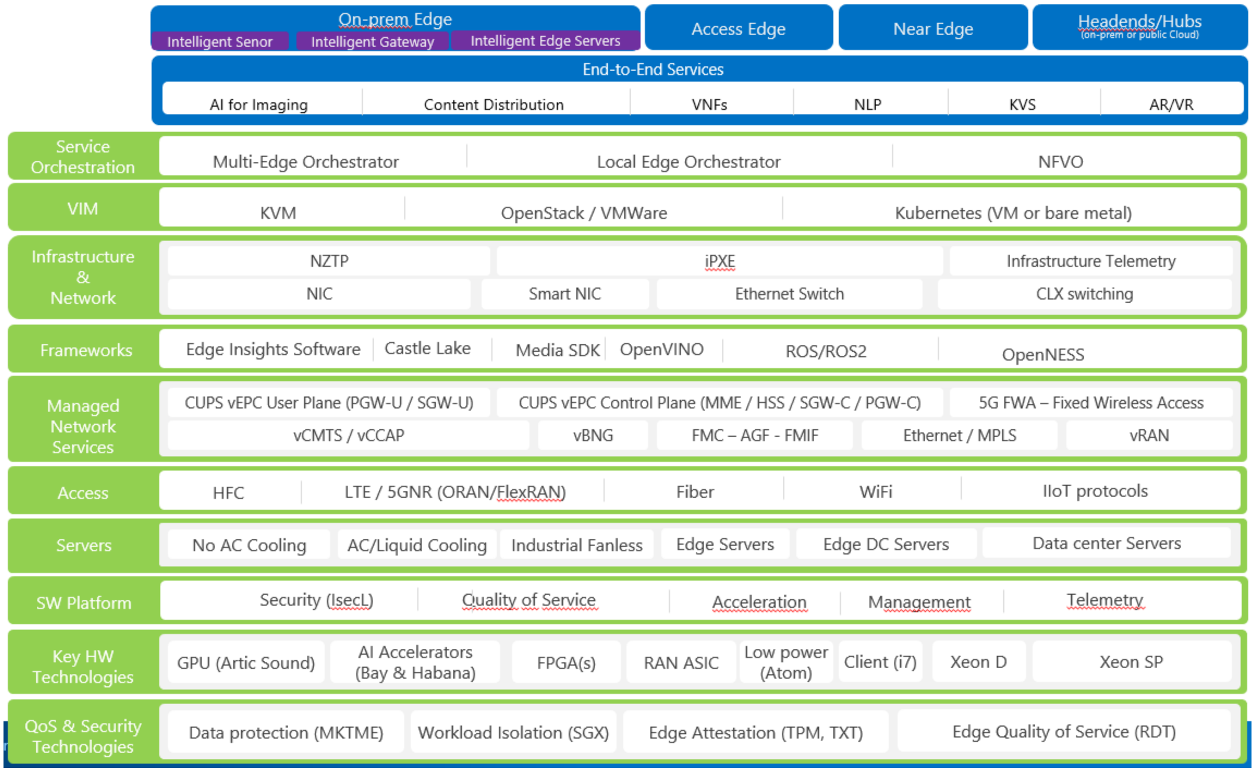

Figure 16 - Mix and Match for an Edge Platform

Work on Converged Edge Architecture continues as real-life solutions for the IoT and Network Edge solutions are developed, new hardware and software elements are deployed, and lessons are learned in real-world conditions.[8] The goal is to provide a framework or template that can be used by OEMs, ODMs, TEMs, and ISVs to develop scalable, flexible, Edge platform solutions for wherever they may be needed across the Edge.

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 leads us to look at the capabilities of facilities closest to the end users – infrastructure collectively known as "the Edge".

6.1. A Summary of the Options

This Edge, of course, is not necessarily constrained to being one location, one platform, or one type of business arrangement. At the top level, an operator can plan for an On Prem Edge or 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

[8] "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

based on hosting Access and Services from equipment owned and operated by the CoSP, partner CSPs, or other 3rd 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 really boils down to the questions of:

- 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 it also means the owning organization needs 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", that is, where to host Edge Platform equipment to provide the best latency and data locality profiles for new Services. In fact, intelligence and compute capabilities are coming to all places in the network – even to Nodes and other locations in the outside plant. Along these lines, the MSO/CoSP also has the expertise to design and deploy whatever Access elements 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 there is a lot of innovation/experimentation happening – is in answering the rest of the questions as they relate to providing Services. It seems the default behavior would be to bring in a CSP or 3rd party to host and manage their own Services over the last mile broadband connection being provided by the CoSP. This is simply because the 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 the CSP equipment resides in a point of presence near-to, but outside, of the CoSP Edge location. In these cases, the CSP and/or 3rd parties making use of CSP resources "own the physical equipment… software infrastructure… 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.

CoSP's 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 used 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 leave it to an aggregator with a local point of presence host the equipment and software infrastructure and sell broadband and, potentially, Services through them (along with perhaps similar offerings from competitors).

Again, 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 space available, the amount of power that can be delivered, and other environmental constraints, with virtualization, telemetric capabilities, facilities for remote security, and software infrastructure for managing such distributed computing elements, it is possible to consolidate the number of platforms into the minimum possible. Less disparate architectures and technological domains means a better total cost of ownership across the network.

To this end, the Converged Edge Architecture project is an effort to help OEMs, ODMs, TEMs, and ISVs mix and match common COTS hardware and open source software elements through a common framework to construct a scalable and flexible platform that matches the performance, lifecycle, and form factor requirements for any Edge location. Commercial examples of industrial On-Prem, 5G vRAN, and other solutions that made use of the Converged Edge Architecture framework are starting to emerge and providing the industry with proof points of the benefit leveraging standards, COTS hardware, and reference software architectures for faster time to market and lower total costs of ownership.

6.2. Asking the Right Questions

Admittedly, it will take more than a checklist to consider all the options above and architect an Edge or Edges 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, Data Center and Cloud resources, etc. But it is because of this breadth that it is clear the Edge is important for MSOs to grow their businesses; whatever options are chosen. If fact, while difficult, thinking about a grand Edge strategy may be the chance for an operator to re-think existing silos, align on the latest technological innovations, and be a distinguishing factor against the competition.

All that said, practical realities to leverage existing systems, skillsets, and business relationships will make this an iterative process. That is, it is good to develop the ultimate end-state for your network and the organizations supporting it, but you may have to accomplish the transformation in stages. For example, would implementing a CSP Led model first allow you to get into the market and show the value of your CoSP Edge locations while you start developing Cloud expertise internally and drive the Ecosystem to be able to consolidate Access and Services later 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

