SCTE•ISBE CABLE-TEC EXPO® VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall Technical Forum SCTE·ISBE • NCTA • CABLELABS®

Framework for Convergence of Services on The MSO Network

Using the Principles of Network Slicing

A Technical Paper prepared for SCTE•ISBE by

Fernando X. Villarruel
Chief Architect
Ciena
1185 Sanctuary Pkwy, Alpharetta GA, 30009
fvillarr@ciena.com

David Reale
Network Architect
Ciena
1185 Sanctuary Pkwy, Alpharetta GA, 30009dreale@ciena.com

Table of Contents

List of Figures

1. Introduction

The digitization of the cable access network enables the promise of end-to-end network convergence for different lines of service. The promise, however, does not come with a framework and this leads to trepidation on how to proceed. A framework for service convergence must recognize three key principles. The first is that lines of services can have unique prioritization, throughput and latency requirements that need to be met. The second is that within a service there will be distinctions of endpoint types and applications that will eventually need their own unique treatment or policy through the network. The third is that the principles mentioned above are transitional over the lifecycle of the service and automation mechanisms to adapt and coexist are necessary to maintain viability over the long term.

In this paper we propose a framework for service convergence using network slicing for MSO networks. We review the network slicing principles for 5G and point to possible analogies that aid in developing a framework for MSO slices including residential services, business services, and mobile services. We cover the concept of network slicing functions which organize and partition network resources available to each service. We describe hard and soft slicing mechanisms, the implementation for slice-aware logical networks and the functionality necessary to maintain end-to-end slice visibility and usability over the lifecycle. We also provide several industry recommendations useful for a converged service environment utilizing network slicing, such as open interfaces for core functions, QoS implementation in packet networks, listing of slice expectations, usability of hard and soft slicing.

2. Opportunity Statement

The state of the cable industry has top-of-mind investments and revenues from mobile services, the growth of enterprise business services, and maintaining a robust residential high-speed internet and video service. There are multiple MSOs worldwide that already own and manage all three services, and for the last decade in the United States there have been coalitions formed with wireless providers to collaborate in transmission and management of wireline and wireless services. What has been noted, however, is that this is a financially complicated situation when done piecemeal, and that profitability has better odds coming from infrastructure-based network convergence. (BAUMGARTNER, 2019)

Enterprise business services on the other hand has been a solid growth engine for MSOs in recent years, becoming a solid primary connectivity option for enterprises. As a metric, in the past several years cable providers continue to make their way up in the Vertical Systems Group Carrier Managed SD-WAN Services Leader Board with a placement recently in the top seven (Vertical Systems Group, 2020). But one of the drawbacks to even more growth has been time to deployment and SLA enforcement partly due to a dependency on existing methods of non-automated bring up of circuits and separate networks including operation and engineering. Thus, the enterprise network depends on evolution to automation, for itself, in the context of automation being also useful for mobile and residential services.

The residential network continues to have solid growth, with revenues for high speed internet growing at 5-10% year-over-year (Comcast , 2020), (Charter , 2020). The status of the residential network exploring convergence with mobile services, and no convergence with business services is understandable due to historical precedents, but there are several dynamics at work in the evolution of the residential network that allows us to take a renewed look at converging multiple services on common network infrastructure. One is the nature of Distributed Access Architectures (DAA), where the legacy residential cable plant, both fiber and coaxial plants have used analog technologies for distribution, but are now transitioning the fiber portion to digital Ethernet and IP, thus all networks will now approximate the same format for transmission. Second is the evolution of DOCSIS and Video to cloud native cores. Now DOCSIS, Video,

Mobile and general broadband network gateways that handle subscriber management will have a deployment path to use containerized computation schemes with generic hardware support. And third is the large expansion of endpoints that need connectivity for 5G and DAA and enterprise 1/10GbE fiber services which could benefit directly from the very broad coverage of HFC and its availability of electrical power sources to distribute cells, Remote-PHY Devices (RPD) / Remote-MACPHY Devices (RMD), and aggregation points for CPEs (Chamberlain, 2018). From this perspective, the industry is ripe for full infrastructure convergence for the services mentioned above leveraging investments in an optimized and streamline manner.

From the user perspective convergence assists in typical usage methods. Take for example the smart phone user maintaining a session at a home, with Wi-Fi enabled backhaul, proceeding to a car driving down the road all on seamless session. Or the parent working from home at the same time and on the same high-speed internet that provides over-the-top video entertainment for the kids. These examples include seamless handoffs between different service types with different usage priorities.

2.1. Network as a Service

The drivers for network convergence not only include reduction of OPEX and improved general efficiency, it also includes the capability to add new revenue streams. Network convergence organizes the qualities of the network such that the network itself is an asset to sell. When the network understands the user transmission profile and its tendencies, then the path is set for creating a robust platform for network as a service (NaaS), where the operator offers highly customizable Virtual Network Functions (VNF) for customers who need a broad range of network capabilities (Hodges, 2019). A complete description of NaaS capability is left for another work, but necessary to mention in the context of a network slicing byproduct.

3. Solution Statement

The nemesis to converging residential, private line business services, and mobile services on common network infrastructure boils down to a possible acrimonious sharing of resources. After all, the networks mentioned above have grown up independently, with full control of their infrastructure, customer base, their priority structure and SLA compliance. These are valid concerns of a converged environment that merit a comprehensive solution. This paper proposes a method for developing a non-acrimonious solution for network and service convergence where the sharing of resources is possible. To do this we borrow from the principles of network slicing as recently defined for 5G. Naturally, an implementation of network slicing for the MSO will have its own idiosyncrasies consistent with our needs and expectations.

Before we discuss network slicing, we review a few technical principles that are key to understanding network slicing.

3.1. Convergence View

It is useful to catalog the extent of the convergence possibilities. Figure 1 provides a list of the different type of physical networks, services, and subservices that would be part of a converged end to end network. We list three access network types, the HFC plant, a point to point fiber plant, and a passive optical network. All these access types would converge on the same aggregation point C, at a hub or in the field, such as an unmanned cabinet or a strand-mounted device. Aggregation at point C could be done in fiber, per wavelength, in straight forward Time Division Multiplexing (TDM) using Optical Transport Network (OTN) or Flex-Ethernet (FlexE) framing, in layer 2 switching using Ethernet, or layer 3 routing using Internet Protocols. Northbound the aggregated signal could range from 10 Gbps to 400 Gbps

depending on optical technology and desired level of signal concurrency (Villarruel, 2018). The aggregated signal would then typically terminate in another aggregator, point B, whose job is to add/drop signals locally if a service core is present, or further aggregate signals onto a metro type network with signal rates in the 100 – 800 Gbps range. The technologies northbound are typically IP with embedded high throughput optics in the rage of 100-400 Gbps or a more advanced photonic layer in the range of 100-800 Gbps. Aggregation point A is then the termination point for the signals that originated in the access. Here they are evaluated in accordance to session and subscriber management at independent services cores. The Converged Cale Access Platform (CCAP) takes care of the DOCSIS signaling and represents in this case other auxiliary cores of other residential type functions, such as out of band set top box signaling and test and measurement. The video core cares for the distribution of MPEG sessions, broadcast and narrowcast. The 5G core cares for the mobility signaling. With the understanding that there is an ongoing evolution from LTE to 5G core, we shortcut to 5G for sake of expediency in this discussion.) The Metro Ethernet Forum (MEF) core stands for the session management for enterprise and applicable business customers. The Broadband Network Gateway (BNG) is the session and subscriber manager for the data subscribers of the Passive Optical Network (PON) network.

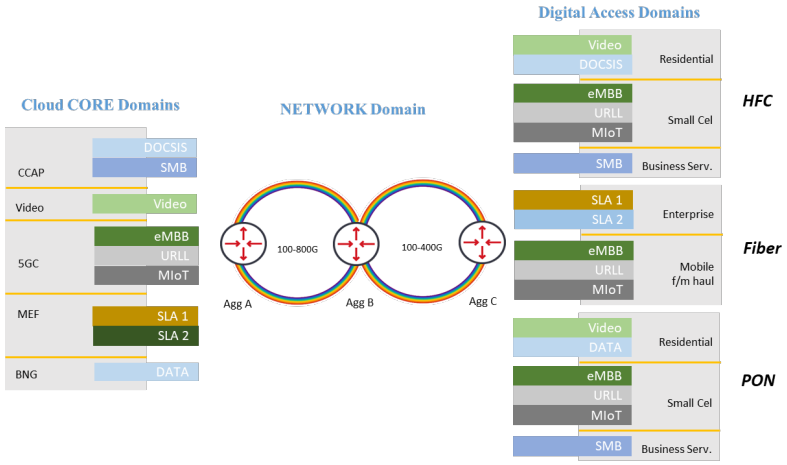

Figure 1 - Convergence of services for MSOs

Within each access network there are different types of services. Figure 1 shows a non-exhaustive list of service types. In the HFC access, for example, there are wireless small cells, which are also expected to use DOCSIS as backhaul (Andreoli-Fang, 2019), along with business customers. For Fiber access there are enterprise customers and mobile front / midhaul or backhaul customers. For PON access there can be residential, small cell and business services customers.

Figure 1 also shows that within each service vertical there are subservices, this is also a non-exhaustive list. These are unique user applications that have specific demands of their parent network consistent with the end user expectations. Within HFC/residential there is Video and DOCSIS data. In the HFC/small cell domain we list enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communication (URLLC), and Massive IoT (MIoT), which are specifically defined sub-service types for 5G. For Fiber/enterprise there are SLAs and in both PON and HFC there are SMB Business Services. Note that the same type of subservice could live within different access domains.

Now these different types of customers/services are managed by their respective core domain. In practice, and for the foreseeable future, we can assume that even though there is a transition to cloud-based cores, the cores would remain independent, i.e. not sharing compute or memory resources. Figure

1 however presents the cores as converged in the minimum sharing resources, (while keeping functional independence), but at some point, in time possibly sharing session and subscriber management functions.

Figure 1 also shows that subservice types have a unique termination point on the core, independent of access type they originate from. This is also a forward-looking proposition, as we recognize that each subservice now subtends to its own specific service core.

4. Convergence Drivers

4.1. Distributed Access Architectures

DAA is a technical driver for convergence. This is a topic that has received much due attention in the past few years and here we refer to work that has already been done. DAA is driven by the evolution of the residential network from analog fiber to digital fiber by extracting the physical RF layer from CMTS or effectively extending the Quadrature Amplitude Modulation (QAM) modulation platforms and positioning them in a separate location, where separate can mean a different shelf, in a different hub, or at the end of the deeper fiber additions in the outside plant, in a street cabinet or on strand mounted node like platforms. DAA technologies create an Ethernet / IP network where several aggregation points are needed to distribute or collect signaling. DAA because of its standards based digital transmission, is a natural platform for usage of other access endpoints that coincide using ethernet or IP access signaling (Villarruel, 2014). Ethernet point to point signals for business services apply, and so do recent pluggable OLT technologies that are granular OLTs, sprouting PON networks from any given 10GbE switch port (Villarruel, 2015). Figure 2 shows the evolution of the residential access plant, starting with analog based fiber then adding digital endpoints, followed by aggregation in the field and ultimately convergence of other services. On the core side it begins with legacy video and CMTS platforms evolving to cloud native cores along with a possible centralization of cores from smaller hubs to larger hubs or from hubs to headends, maximizing efficiency of compute. Not shown but also part of the DAA discussion is partitioning of software functions so not all of the DOCSIS stack resides in one location, this for example is the effort being done in the Flexible MAC Architecture (FMA) working group at CableLabs (BTR , 2019).

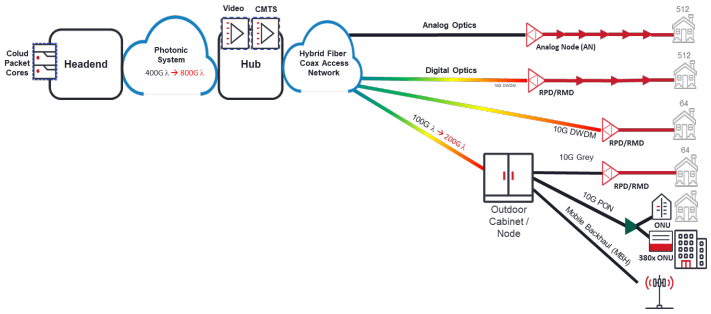

Figure 2 - Evolution to Distributed Access Architecture, DAA

4.2. Cloud Native Service Cores

The separation of specific PHY implementations in the CMTS has allowed a rethinking of its software architecture. Effectively virtualizing the CMTS by the decoupling of the software stack from vendor specific hardware. Beyond just virtualizing however there is a trend towards even more software flexibility by implementing a cloud native architecture where software is broken up by functions, and

these functions are presented as independent containers that run on a container platform and are orchestrated by a framework that facilitates their interrelation and thus present a complete service solution. Interestingly, these container frameworks are built with open source tools so both the hardware and software infrastructure for service cores is now open, in principle. Figure 3 is a simple view of popular cloud native implementations, where Linux is the operating system, Docker is the container platform and Kubernetes is the orchestrator, all of which were derived in open source communities.

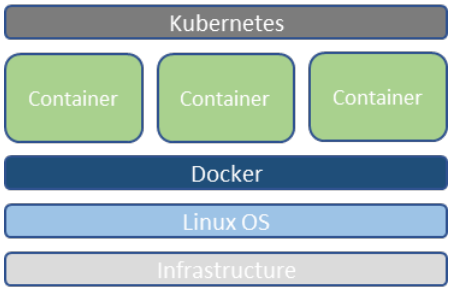

Figure 3 - Cloud Native Framework

The nature of containerization also allows for flexibility in geographical placement of containers, allowing for certain core functions to be closer to the user as needed. This is the science of edge compute, derived from cloud native architectures. This topic is beyond the scope of this paper, but it is a discussion that comes up in advanced implementations of distributed service cores and network slicing.

Also, of interest is the view of cloud native cores in deployment as shown in Figure 4. We see the system as a whole is broken up into functionalities. There is the Data Plane, the part of the software that processes and executes on data transmission requests, this notably includes the PHY layers and forwarding layers. There is the Control Plane that creates the configuration environment for the data. This includes things like path controls for a switched or routed network, session and subscriber control for the end user, bandwidth shaping for the network and authentication of endpoints. Beyond the data and control plane there is typically a Management Plane, which can also be part of the control plane, but here we call it out separately to make a point. This management plane facilitates the interdependencies of the control plane within itself and to the physical network items. In scale this typically cannot be done with simple network management system tools, and so an overall orchestration system is implemented. This orchestration system can have read write capabilities for all hardware and software elements and can provide a stateful view of the overall network to northbound operational and billing systems. This generalized view of cloud native deployments is useful as it sets up possible analogies to 5G.

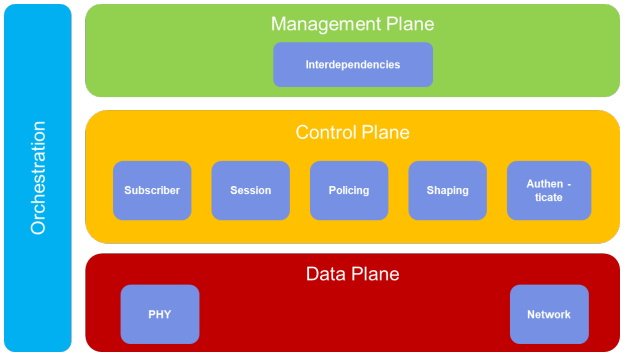

Figure 4 - Typical Cloud Native Core Deployment

5. 5G and Network Slicing

The recent and very popular evolution to 5G in the mobile space provides us a good reference to learn from as we look at the convergence of multiple services and subservices. 5G is the mobile evolution from 4G LTE with drivers coming from increased bandwidth throughput, many more subscribed endpoints, and most importantly for us a widely varied set of usage scenarios, effectively creating a wealth of service types that have broadly varied expectations from the 5G network, see Figure 5. All these varied services are set to run simultaneously making the 5G network a largely different proposition than its predecessors.

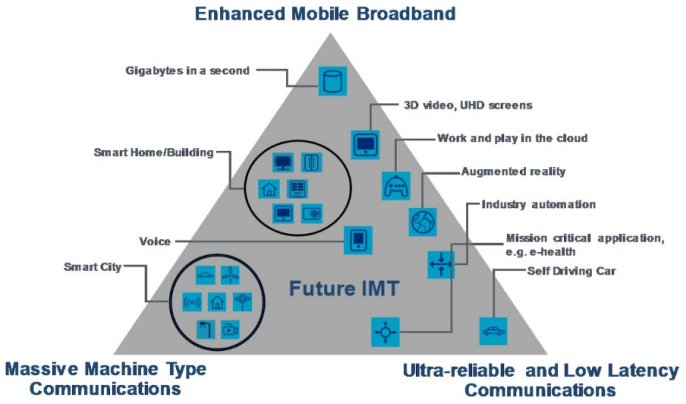

Figure 5 - 5G Usage Scenarios, (SANTO, 2017)

The capability for 5G to carry various services types has been expressed not only in the radio access network, with many more cell sites and increased bandwidth but also in the way the 5G core has been architected. In Figure 6 we show the 5G core architecture definition according to the third generation partnership project (3GPP), (Mademann, 2017).

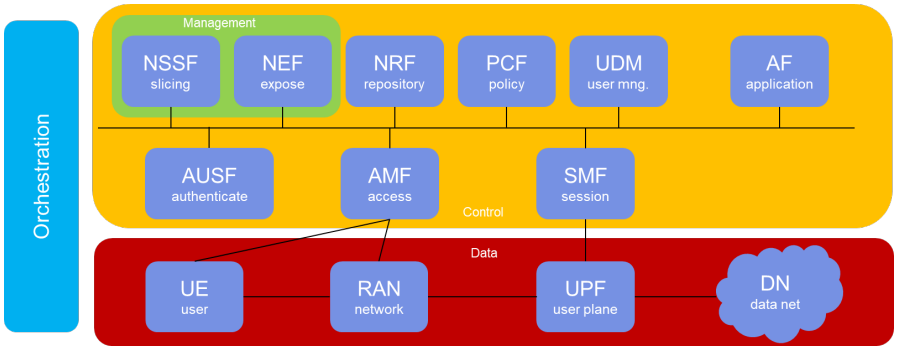

Figure 6 - Cloud Native 5G Core

A detailed review of the 5G NG architecture is beyond the scope of this paper but there are three notable evolutionary items making it different from its predecessor. They are listed below. In the next section we briefly describe each function.

- The 5G core is a cloud native, service based architecture, where architecture elements are defined as network functions with standard APIs such that they can be called on by any other network function with permission to do so (Felix, 2018). The standardized interfaces allow for multiple vendors to participate in the core, according to their capability in these discrete functions.

- 5G has separated the user plane and the control plane, allowing for top of the line hardware and software technologies and a flexibility of deployment models. This is referred to as Control User Plane Separation (CUPS). In this discussion we refer to the user plane as the data plane, see Figure 6.

- Most of the functions in the 5G architecture are carry over from LTE with some variations and containerization, however, some of them are new and specific to 5G. The new functions, we list under the management plane in Figure 6 are The Network Slice Selection function (NSSF), the Network Exposure Function (NEF). These functions were created specifically with the envisioned multiservice capability of 5G (variety of device types), and the dynamic nature of these services over time. Note that the calling out of NSSF/NEF as a management plane in Figure 6 is an exercise done for this paper to highlight the nature of these functions. The formal 5G architecture keeps them as just part of the control plane.

5.1. 5G Functions and Descriptions

Below is a brief description of the 5G core functions.

- UE: User Equipment. Any other device with mobile connectivity, such as smart phones.
- RAN: Radio Access Network. This is the network that connects user equipment to other parts of a mobile network via a radio connection.
- UPF: User Plane Function. Features to support packet routing and forwarding, interconnection to other data networks, and policy enforcement. Similar to the packet gateway in 4G LTE.
- DN: Data Network. Broader service provider network, "the internet."
- AUSF: Authentication Server Function. Authenticates UEs and stores authentication keys.
- AMF: Access Management Function. manages user equipment registration, authentication, identification
- SMF: Session Management Function. Establishes and manages UE sessions, static and in movement, allocate IP addressing, informs on quality of service.
- PCF: Policy Control Function. Provides policy rules to control plane functions.
- UDM: Unified Data Management. Stores subscriber data and profiles.
- AF: Application Function. Connectivity point for management functions and control plane and data plane.
- NRF: Network Repository Function. Registration and discovery of network functions so that they can find each other.
- NSSF: Network Slice Selection Function. Has the task of selecting and directing network traffic to the use of particular network slice? Its assignments are determined by allowed usage per a network slice library found in the network slice selection assistance information (NSSAI). It also determines the access and mobility function (AMF) settings applicable to a user entity.
- NEF: Network Exposure Function. A mechanism that securely exposes state information of the 5G core, which includes capabilities and events, packet flow descriptors, and translation services for the flow of internal external information.

Of greatest interest in this discussion are the NSSF and the NEF, as together the allow for a dynamic network slicing mechanism over time.

5.2. What is Network Slicing

Network slicing is formally a method to run multiple end-to-end logical networks on a common set of resources. Network slicing in the most basic sense is not new. We have to date used to layer 1 slices with

managed OTN, or layer 2 slices with managed Ethernet Virtual LAN Networks (VLAN), or layer 3 with managed IP virtual private networks (VPN). This is certainly a part of network slicing, but in the more general case the slicing logic also happens end to end, which includes the partitioning of not only of packet network resources but also software control and management resources. Figure 7 depicts a specific set of resources being partitioned for the whole network. This end to end network slice would be available for a service to subscribe.

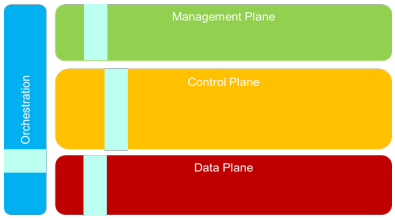

Figure 7 - End to end network slicing representation

There are particular service and operational expectations for 5G slices according to the 3GPP definition (3GPP, 2016). We summarize the expectations below to promote an intuitive understanding of how network slicing is leveraged in deployments. Ultimately as the concept of network slicing matures in the cable space there will need to be a well understood set of expectations, not exactly like the one listed below, but most likely similar with variations addressing the specific needs of MSO systems.

5.2.1. Service Requirements

1. Slices are unique sets of network functions and configurations. The operator can create network slices from a multivendor environment of functions.
2. Network slicing is a dynamic exercise in an autonomous system applicable for diverse market scenarios.
3. A system can recognize a UE and its associated network slice.
4. A system allows the UE to subscribe to a specific network slice for service.

5.2.2. Operation Requirements

1. The operator can operate different network slices in parallel with isolation per slice.
2. Slices maintain the security profile expected of the service that uses it.
3. Network slices are isolated such that a cyber-attack would be confined to one slice.
4. Operators can authorize a third party to manage the network slicing environment, per suitable APIs.
5. The network slicing system is to scale in capacity with no impact to its service or other slices.
6. The system can accept changes to slices with minimal impact to subscriber services.

5.3. Service Convergence and Slicing

A service is the instantiation of a set of features, policies, and configurations. A service is facilitated using a set of well-defined network functions and resources. In the context of network slicing we refer to a service as having a unique “blueprint”, and a blueprint can be paired to one or more slices.

Service convergence then is the practice of multiple services (or subservices) coexisting in a well-defined manner on shared networking resources, (not unlike DOCSIS and Video do in legacy HFC). Service convergence is then simply the management of a collection of blueprints. Which implies, from previous sections, the management of parallel or interrelated slices.

For service convergence it is useful to take the birds eye view of a system, its multiple domains, slices, and their relationship. Figure 8 shows the components of a network slicing system for a typical service provider. There are several types of domains. The access domains with its multiple type of UEs which in the case of the MSO this would include the residential, enterprise and mobile access. We also have the network function domain, in practice this could include secure initialization, session bring up and tear down, encryption, etc. The network domain would be the resources for connectivity, in this domain there can be many slice types or various use cases available. This would be the case where you find distinctions for VLAN or VPNs, transmission methods for MPLS or Segment Routing (SR), each profile having its own slice. The core domain is a set of slices that give each service a unique profile southbound to the UE and interpret that service northbound to the operational and billing systems.

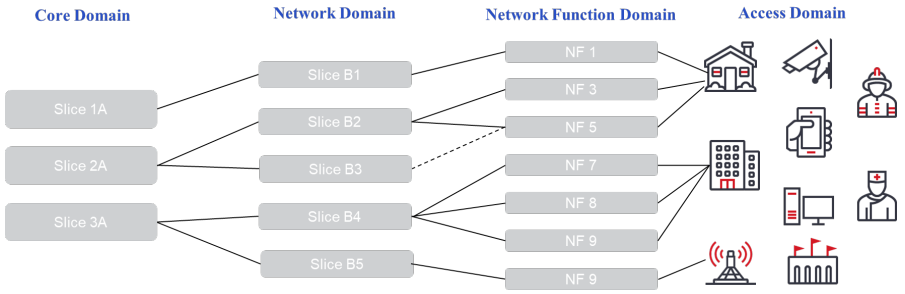

Figure 8 - Domain and Slices Relationship

What is important is not just the existence of slices but their inter-relationship in the system. We note that each service, or subservice can use a collection of network functions and subscribe to one or more network slices in a domain thus creating a specific blueprint. Figure 9, for example, shows a possible service blueprint in the highlighted mustard progression. Consider a security camera in the residential network. This camera needs resources from the network functions including the need for a timing service and local encryption. From the network domains it needs a VLAN or VPN tunnel, and from the core it needs a unique session and caching memory. This unique collection of resources is the blueprint for the security camera service.

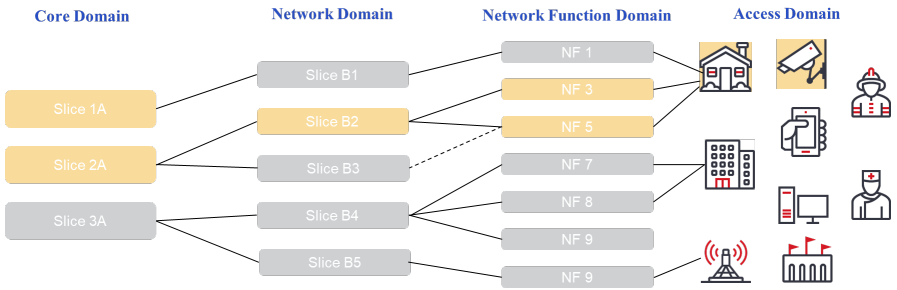

Figure 9 - Service Footprint

While the functions, slice and slice pairings in Figure 9 are unique they are also dynamic in nature, as the collection of needed slices and network functions can change over the lifecycle of the service. Thus blueprints create scenarios where the relationship across domains can be collaborative, interdependent, and pre-configured to particular scenarios.

The configuration of unique scenarios, as the example given in Figure 9, along with the end to end network slice management necessary over its lifecycle is what is referred to network as a service,

commonly referred to as NaaS, where the creation of particular revenue streams is now possible because there is an infrastructure to create and manage different blueprints. This can certainly be a usable capability for the MSO space moving forward. Consider the blurring lines between home and work because of the COVID-19 pandemic. As we considered the earlier example there is a case to differentiate, though the whole network, work traffic at home versus entertainment traffic at home—effectively two different blueprints.

6. MSO and 5G coexistance

In Figure 1, of section 3.1 we supposed the reuse of redundant functions in service cores, thus creating end-to-end convergence. But this goal is aspirational, and to thing about an interim path is necessary. We refer to the work done to support the momentum for integration of small cells onto the cable landscape, with DOCSIS backhaul (CableLabs, 2020). Per definition there is a generic satellite function to the 5G core that allows for a wireline system, to present itself as a user entity to the 5G core (3GPP, 2018). This function is called the Wireline Access Gateway Function, (W-GAF) and has standard interfaces, transmitting data plane traffic to the user plane function and control traffic to the 5G core control plane. This in effect allows for any non-5G system to have an embedded RAN and subtended devices, with the expectation that the wireline core has the capability to decipher separate data and control plane traffic.

In the case of cable, as shown in Figure 10, the W-GAF facilitates having small cells within the CMTS/DOCSIS system and creates an environment for UE's, behind the CMTS, to be treated as native participants of the 5G core. This is made possible in conjunction with the work done for DOCSIS MAC timing and latency functions (Cablelabs Timing, 2020) (Cablelabs Xhaul, 2020)

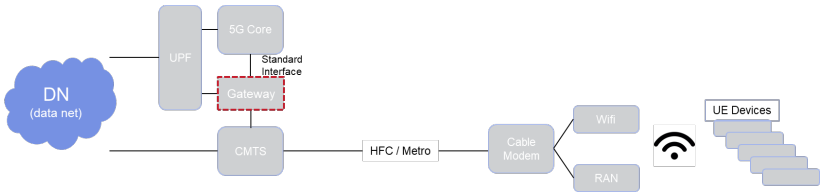

Figure 10 - CMTS / 5G cores coexistence

The working model for the coexistence of the CMTS and 5G, via a standardized gateway, gives a glimpse on the possibility to reuse the existing NSSF and NEF as a generalized slicing solution for other cores. The work at 3GPP championed by CableLabs, reported in the document "5G Wireless Wireline Converged Core Architecture Technical Report" (CableLabs, 2020) shows that there is already an established data model for an HFC system with a RAN that can subscribe to, be authenticated by, and admitted into the 5G core. Further there is a proposed method to extend not only subscription to the core but extend the QoS mechanism from the 5G core, available through the wireline gateway to the CMTS and its constituents. This is somewhat straight forward as the RAN components in the HFC have native wireless tendencies towards the control and data plane structure of the 5GC.

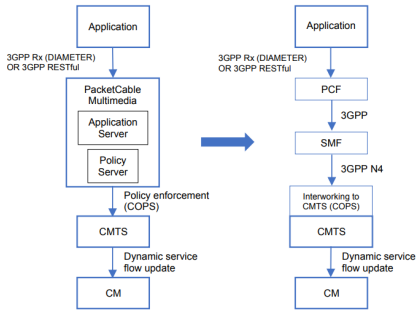

Figure 11 - General QoS mapping CMTS and 5G core (CableLabs, 2020)

Outstanding however is how to incorporate the non-wireless traffic into the global view for network slicing. What would be useful is a global QoS mapping to all CMTS traffic, as expressed in the DOCSIS 3.1 MAC Upper Layer Protocols Interface (MULPIv3.1) and Packet Cable Multimedia (PCMM), creating a complete map for all services, see Figure 11. The extension of QoS for native DOCSIS traffic can then be interpreted by 5G core Session Management Function (SFM) via a Common Open Policy Service (COPS) (CableLabs, 2020). With a global QoS structure in place the creation of MSO non-mobile specific slice types can be defined. The management of both HFC and 5G slices can then be done by the already existing SFFM.

It is worth mentioning that now QoS tagging is generally not a practice for DOCSIS packets in general and not over the new digital access fiber infrastructure. This is because it relies on the time map mechanism between CMTS and modems. There is a QoS mechanism that is embedded into the Xhaul specification, which will be necessary when transmitting small cells over DOCSIS and could prove useful if there is a general application of QoS for all residential signals would be necessary. An extension of QoS tagging to the Ethernet / IP network could prove useful or necessary as the MAC layer evolves in positioning, per the efforts of the FMA. In these cases, the network can participate in more intricate slices along with other elements of the end-to-end network.

An extension of this principle can be applied to the enterprise vertical, adopting enterprise bandwidth profiles to the global QoS mapping, starting at the first network aggregation point. The network slices for the enterprise would in principle use the structure laid out by native OAM signaling.

7. Slicing methods

7.1. Soft Slicing

Soft slicing refers to the practice of sharing resources with system enabled flexibility. In soft slicing the slices are statistically separate and transmit by logically multiplexing the data plane over some physical channel (IETF, 2018). Soft slicing makes use of the layer 2 Ethernet and layer 3 Internet Protocols we have grown accustomed to, dynamic MPLS, Segment Routing, other tunneling mechanisms, and even RAN frequency sharing.

Figure 12 shows a soft slicing example for convergence, with 5G, DOCSIS and MEF Enterprise services represented. Note that within each core there is a QoS priority structure, native to each one and the cores may or may not be sharing functional resources on the control plane. On the data plane however we see that there is an expression of an overall QoS per the management of the NSSF and the dynamic

management of the NEF. This exercise implies that the networking infrastructure is programmable with stateful telemetry to the management systems.

Soft slicing is useful when the operator is trying to maximize the return on investment on networking equipment or tying to minimize carbon and physical footprint of the data plane due to other forces. Note that the prioritization in Figure 12 necessarily implies a policy mechanism by the operator that sets the priority structure for the signaling and while technically possible it can have its hurdles per historical precedents discussed in Section 3.

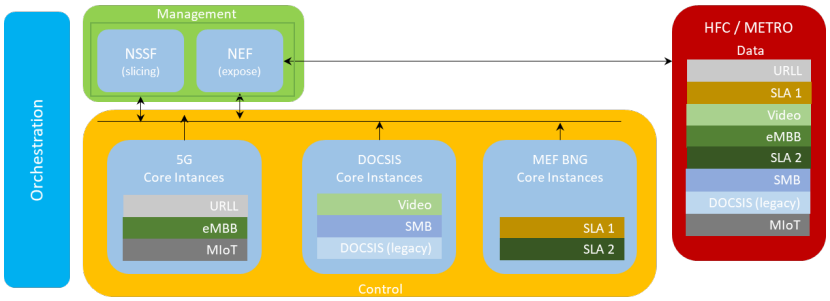

Figure 12 - Soft Slicing Example for Convergence

7.2. Hard Slicing

Hard slicing refers to the provision of resources in such a way that they are dedicated to a blueprint instantiation. This refers to a slice being assigned a particular lambda or ONT muxponder or Flex Ethernet (FlexE) channel.

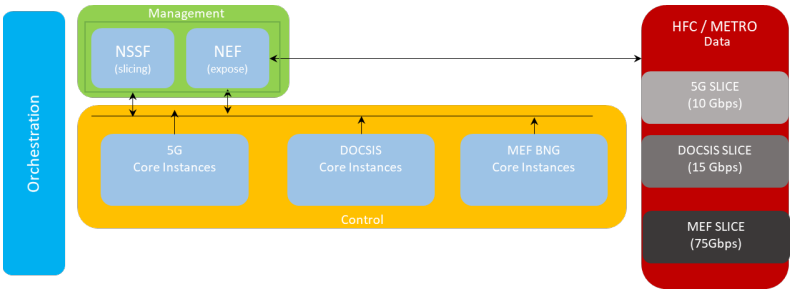

Figure 13 - Hard Slicing Example for Convergence

Figure 13 shows an example of convergence through hard slicing, using the mechanism of FlexE, as defined in the specifications from the Optical Internet Forum (OIF) and International Telecommunications Union (ITU): OIF-FLEXE-02.0IA and ITU-T G.mtn, respectively. FlexE is basically a time domain multiplexing of Ethernet signals where, as part of its capabilities, signals of lower bandwidths can aggregate to a higher bandwidth signal and vice versa, this channelization effect can be maintained throughout the Ethernet deployment, with 5 Gbps of granularity. Formally FlexE adds a flexible shim layer between the MAC and the physical coding sublayer, see Figure 14. Note that the implementation of FlexE is in a large part made possible by the arrival and availability of ZR type coherent optics in the range of 100-400 Gbps, which allow for robust access backhaul solutions.

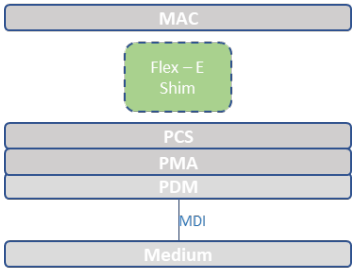

Figure 14 - Flexible Ethernet Operational Definition

In the Figure 13 example there is no system wide knowledge of the prioritization of the subservices within the 5G, DOCSIS or Enterprise services. Instead there is a bandwidth expectation for each service and a FlexE channelization assignment and an understanding from the NSSF as to the assigning of bandwidth per service. The bandwidth assignments themselves are flexible over the lifecycle of the service, however. Hard slicing is useful in cases where resource sharing is needed but hard boundaries between them are necessary. In the MSO case, this could prove a worthwhile steppingstone as the industry moves away from siloed operations per service.

7.3. Hard And Soft Slicing

The combination of hard and soft slicing for the MSO operator could be an implementable middle ground to optimize resources and create the right sort of boundary structure in the first implementation of convergence. As shown in Figure 14 soft slicing is limited to the purview of subservices within each core and in this manner there is no overreaching policy that pins a subservice on one core versus a subservice on a different core. Simultaneously an implementation of FlexE per core is introduced such that each core has a dedicated bandwidth to work with, allowing no confusion on usage limitation per QoS assignments of its subservices. In this case the NSSF is executing on blueprints that include several protocol layers, and as is expected the system is dynamic with the capabilities to evolve hard and soft slicing assignments over time. For the MSOs this can be a long term or near-term solution for convergence.

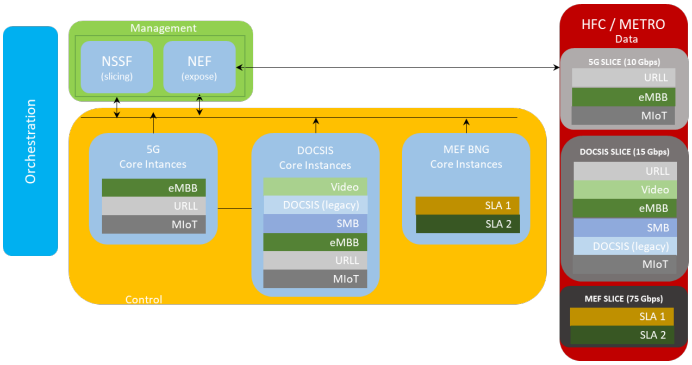

Figure 15 - Hard and Soft Slicing Example

8. Network Slice Lifecycle

It is useful to understand the lifecycle of a network slice. Note that the existence of the network exposure function is key here as by its gathering and maintaining of stateful system information the capacity of slices can be anticipated and thus react accordingly. Below we propose the 5 steps necessary for a lifecycle. These steps help the operator to appreciate the effort necessary in dealing with a network sliced system and creating an expectation for native or third-party enablement of these services.

- Creation
  - In this step a machine-readable blueprint definition is created. This is created by an operator or a third-party planner. The blueprint includes component resources, enabling features, workflow assignments and lifecycle expectation.
- Instantiation
  - At this step resource inventory and availability are counted. It is where function discovery of native or third party VNFs take place, and the orchestration of the system pushes for creation of slices.
- Scaling
  - This includes monitoring network throughput and trigger events. Allows for the extension or reduction of slice capabilities per need. All this is done with zero touch, it is an automated process.
- Isolation
  - This step manages resource impact of scaling on neighbor slices, including items like traffic, bandwidth among other processes. It insures parallel usage.
- Maintenance
  - This step facilitates in instantiation or tear down of slices, redirecting traffic to alternate slices as necessary with minimal or no interruption, and conducts proactive testing.

9. Industry Reccommendations

We list several helpful steps that are meant to help the industry facilitate convergence. The list below is non-exhaustive but does provide the industry with goals worth considering.

- Move towards industry-based open interfaces for cloud native core functions.
  - In the minimum open interfaces for network slicing like functions and network exposure like functions.
- Practice QoS and involve the packet network.
  - Even though QoS capabilities are inherent in DOCSIS for instance, they are not used all the time, nor are they generally reflected in the digital access network.
- Create MSO specific list of slice expectations and requirements.
  - This was an initial and necessary step in the creating of network slicing for the mobile industry. This is an activity CableLabs could lead for example.
- Implement hard or hard and soft slicing as a beginning step.
  - The cable industry must work with, not against, the reality of heavily siloed services, and move from there. Hard slicing is a useful tool in this direction.
- Move towards full functional convergence at the core.
  - This is a long-term goal, but it begins with definitions. The work done in the FMA of standard interfaces for certain MAC functions is a good example of how to start. Ultimately, legacy cable functions are competing with a system in 5G that is written for cloud and with standardized interfaces.

10. Conclusion

The topic of network slicing for convergence of services is novel and necessary. In this paper we have proposed a framework for service convergence using network slicing. We have reviewed the network slicing mechanisms for 5G and pointed out possible analogies that aid in developing slicing for MSO systems containing residential, business, and mobile services. We have covered the concepts of network slicing functions which organize and partition available network resources. We have described hard and soft slicing mechanisms and the necessary steps to maintain end-to-end slice visibility and usability over their lifecycle.

Lastly, we acknowledge the fruitful discussions we've had developing the content of this paper with Raghu Ranganathan and Darren McKinney from Ciena, along with Bernard McKibben from CableLabs.

Abbreviations

3GPP	3rd Generation Partnership Project
5G	5th Generation Mobile Network
AF	Application Function
AMF	Access Management Function
API	Application Programming Interface
AUSF	Authentication Server Function
BNG	Broadband Network Gateway
CCAP	Converged Cable Access Platform
CMTS	Cable Modem Termination System
CPE	Customer Premise Equipment
CUPS	Control User Plane Separation
DAA	Distributed Access Architecture
DN	Distribution Network
DOCSIS	Data Over Cable Service Interface Specification
eMBB	Enhanced Mobile Broadband
FlexE	Flex Ethernet
FMA	Flexible MAC Architecture
GbE	Gigabit Ethernet
Gbps	Gigabits per second
HFC	Hybrid Fiber Coaxial network
IP	Internet Protocol
ISBE	International Society of Broadband Experts
MAC	Media Access Control
MEF	Metro Ethernet Forum
MIoT	Massive Internet of Things
MPEG	Moving Picture Experts Group

MPLS	Multiprotocol Label Switching
MSO	Multiple-System Operator
NaaS	Network As A Service
NEF	Network Exposure Function
NRF	Network Repository Function
NSSF	Network Slice Selection Function
OLT	Optical Line Terminal
OPEX	Operational Expenses
OTN	Optical Transport Network
PCF	Policy Control Function
PON	Passive Optical Network
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RMD	Remote MAC-PHY Device
RPD	Remote PHY Device
SCTE	Society of Cable Television Engineers
SFM	Session Management Function
SLA	Service Level Agreement
SMB	Small Medium Business
SR	Segment Routing
TDM	Time Domain Multiplexing
UDM	Unified Data Management
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra Reliable Low Latency Communication
VNF	Virtual Network Function
VSG	Vertical Systems Group
Wi-Fi	Wireless Fidelity based on the IEEE 802.11

Bibliography

3GPP. (2016). *3rd Generation Partnership Project TR 22.891 V14.2.0.* Valbonne: 3GPP Organizational Partners.

3GPP. (2018). *Wireless and wireline convergence access support for the 5G System.* Valbonne: 3rd Generation Partnership Project.

Andreoli-Fang, J. (2019, September 10). *Enabling 5G with 10G Low Latency Xhaul (LLX) Over DOCSIS® Technology.* Retrieved from Informed Blog by Cablelabs: https://www.cablelabs.com/enabling-5g-10g-low-latency-xhaul-llx-docsis-technology

BAUMGARTNER, J. (2019, April 16). *Comcast, Charter MVNO Deals Are Bad for Everyone – Analyst.* Retrieved from Light Reading: JEFF BAUMGARTNER

BTR . (2019, December 24). *Vecima demos FMA API interoperability for CableLabs.* Retrieved from Broadband Technology Report: https://www.broadbandtechreport.com/docsis/article/14074102/vecima-demos-fma-api-interoperability-for-cablelabs

CableLabs. (2020). *5G Wireless Wireline Converged Core Architecture Technical Report.* Louisville: Cable Television Laboratoris Inc.

Cablelabs Timing. (2020). *Synchronization Techniques for DOCSIS® Technology.* Louisville: Cable Television Laboratories, Inc.

Cablelabs Xhaul. (2020). *Low Latency Mobile Xhaul over DOCSIS Technology.* Louisville: Cable Television Laboratories, Inc.

Chamberlain, J. (2018, January 16). *Necessities for Network Convergence in 2018 and Beyond (Part 1).* Retrieved from Commscope: https://www.commscope.com/blog/2018/necessities-for-network-convergence-in-2018-and-beyond-part-1/

Charter . (2020, July 31). *Charter Announces Second Quarter 2020 Results.* Retrieved from Charter Communications News: https://ir.charter.com/static-files/8402d27e-e891-41ce-ba11-b8fd55f79709

Comcast . (2020, Jule 30). *Comcast Reports 2nd Quarter 2020 Results.* Retrieved from Investor News Details : https://www.cmcsa.com/news-releases/news-release-details/comcast-reports-2nd-quarter-2020-results

Felix, E. (2018, October 20). *5G Service-Based Architecture (SBA).* Retrieved from Medium: https://medium.com/5g-nr/5g-service-based-architecture-sba-47900b0ded0a

Hodges, J. (2019, June 4). *The Rise of Network-as-a-Service.* Retrieved from Light Reading Cloud Services: https://www.lightreading.com/services/cloud-services/the-rise-of-network-as-a-service/a/d-id/752185

IETF. (2018, January 4). *Network Working Group.* Retrieved from Network Slicing Architecture, draft-geng-netslices-architecture-02: https://tools.ietf.org/id/draft-geng-netslices-architecture-02.html#rfc.section.4.3

Mademann, F. (2017, December 21). *System architecture milestone of 5G Phase 1 is achieved.* Retrieved from 3GPP A Global Initiative: https://www.3gpp.org/news-events/1930-sys_architecture

SANTO, B. (2017, August 25). *The 5 best 5G use cases.* Retrieved from EDN: https://www.edn.com/the-5-best-5g-use-cases/

Vertical Systems Group. (2020, April 22). *2019 U.S. Carrier Managed SD-WAN LEADERBOARD.* Retrieved from Vertical Systems Group: https://www.verticalsystems.com/2020/04/21/2019-us-sd-wan-leaderboard/

Villarruel, F. (2014). Plasticity of the New HFC Network Engineering for Remote-PHY and FTTP. *SCTE Cable Tech Expo 14* (pp. 1-23). Denver: SCTE.

Villarruel, F. (2015). Virtual PON Network A Practical Guide For The Network Planner. *Cable-Tec Expo 15* (pp. 1-22). New Orleans: SCTE.

Villarruel, F. (2018). Capacity and Technology Considerations in DAA. *SCTE ISBE NCTA CABLELABS 2018 Fall Technical Forum* (pp. 1-23). Atlanta: SCTE*ISBE.

