Too Many Cooks in the Kitchen

Fostering Organizational Cohesion by Digitizing the RPD

A Technical Paper prepared for SCTE by

De Fu Li
Distinguished Engineer
Comcast
1701 JFK Boulevard Philadelphia, PA 19103
defu_li@comcast.com

Gregory Medders, Comcast

Eric Stonfer, Comcast

Bhanu Krishnamurthy, Comcast

Sinan Onder, Comcast

Mehul Patel, Comcast

Table of Contents

List of Figures

List of Tables

Abstract

The distributed access architecture is enabling Comcast's ambitious program to offer symmetric, multi-gig speeds to its customers over a full-duplex (FDX) hybrid fiber-coax network. Over the past years, we have learned many valuable lessons required to activate 1000s of remote physical layer devices (RPDs) a week that we will apply to the D4.0 rollout. However, a production scale deployment of this technology involves many interdependent processes and components, including planning and capacity management, plant design, logistics, construction, provisioning, and operations.

Furthermore, the FDX deployments often take place incrementally over infrastructure that already serves live customers, requiring precise coordination across teams and organizations to ensure a good customer experience. Left unchecked, teams may unintentionally interfere with one another and spoil the work of other groups, much as too many cooks in kitchen may spoil a meal.

Sharing our lessons learned in our mid-split transitions, we describe how a digital transformation of the underlying access network provides a mechanism to avoid the worst manifestations of Conway's law. Because the RPD lifecycle is complex, different domain experts are required at each phase. This can result in ambiguity over who is responsible for the RPD, since it changes over time.

Through a digital transformation of the RPD, we show how a federated data service can be used to democratize data and present guardrails that must be in place to ensure its integrity. Crucially, while the same teams still own the same aspects of the RPD lifecycle, by communicating through the common digital RPD, the scaling implied by Conway's law is mitigated, enabling a smoother transition to FDX. In the future, the digitization of the cable access networks will fundamentally change how Comcast operates, enhancing its ability to provide a reliable, cutting-edge experience to customers.

1. Introduction

Since the days of dial-up, internet service providers have balanced the need to provide reliable services in the current market while anticipating the future growth and data needs of the diverse applications enabled by the internet. Whether due to a surge of downloads prompted by a new gaming platform or the unprecedented shift to working from home during the COVID-19 pandemic, anticipating and exceeding customer internet connectivity requirements is a constant challenge. With the introduction of symmetric multi-gigabit speeds afforded by the most recent standards, cable operators have the opportunity to deliver the fastest speeds at the lowest latency, powering the residential and commercial computing needs of tomorrow.

1.1. iCMTS

The data over cable service interface specification (DOCSIS) standard defines how broadband internet can be distributed over a hybrid fiber-coax (HFC) network. In traditional DOCSIS, the key network elements are the cable modem (CM) and the cable modem termination system (CMTS). As a form of customer-premises equipment (CPE), the CM facilitates connectivity to the internet via the CMTS, a device typically deployed in cable operators' head-end facilities that can serve the control and data needs of thousands of CMs communicating over the shared HFC network.

Earlier generations of CMTSs were built as custom hardware platforms. These are often termed "integrated" CMTS (iCMTS) because they utilize an integrated DOCSIS MAC and PHY function.

More recently, the distributed access architecture (DAA) specification, also known as the modular head-end architecture version 2 (MHAv2), was introduced [1]. DAA splits the CMTS into two distinct components: the physical (PHY) function and the core function. The remote PHY device (RPD) provides the PHY function, while the core functions consist of CMTS operating on the medium access control (MAC) and internet protocol (IP) layers.

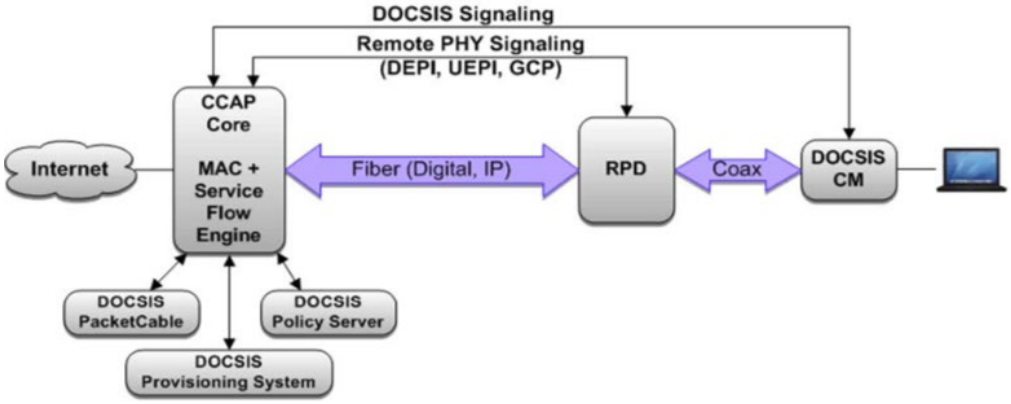


Figure 1 – MHAv2 Architecture (Source: [1])

For the purposes of this paper, iCMTS is used as a representation of custom hardware CMTS platforms.

1.2. vCMTS

The PHY/core split allows significant flexibility in how the CMTS functionality can be implemented. A key feature enabled by the split is the ability to run the CMTS on commercial off the shelf (COTS) hardware, rather than relying on custom hardware. This in turn provides an opportunity to modernize the CMTS, transitioning it from an integrated hardware solution to a collection of software applications or microservices, many of which can even run on a public cloud computing platform. Taken together, this architecture is referred to as the virtualized CMTS (vCMTS).

To smooth the transition from iCMTS, some vCMTS implementations were built to look and feel exactly like an iCMTS through familiar features such as the command line interface (CLI) and the SNMP (Simple Network Management Protocol) interface. In principle, these common interfaces abstract away the iCMTS/vCMTS distinction, allowing the multiple system operators (MSOs) to deploy and operate a vCMTS in a black box way as with iCMTS. However, treating a vCMTS like "just another CMTS" would neglect the significant opportunities provided by modern software and cloud-native architecture.

Adopting a modern, containerized architecture allows us to leverage open-source infrastructure tooling and software. The core vCMTS applications are deployed using Kubernetes, which enables features such as high-availability and in-service software upgrade. From an end-user perspective, one of the biggest advantages of the move to cloud-native architecture is real-time observability. For example, we monitor logs using the Elasticsearch, Fluentbit, and Kibana (EFK) stack. Rather than relying on SNMP polling, real-time telemetry about the customer experience is obtained using Prometheus, a time series database (TSDB). Additionally, end-to-end automation of a vCMTS is very broad in scope, including infrastructure as code, site stand up, deployment, incident and change management, alerting, and automated remediations.

While some of the microservices in the vCMTS ecosystem can (and do) run in the public cloud, there remain some aspects of a vCMTS that are difficult to operate in a public cloud environment. Very precise timing must be maintained between the RPD and the core (e.g., for DOCSIS upstream bandwidth allocation maps (MAPs)). For this reason, routing communication from the RPDs to the core over the internet backbone (or even over dedicated connections to a public cloud provider) introduces variation in latency and congestion that could impact the subscriber experience. For this reason, Comcast operates a hybrid model utilizing both the public and private cloud. The data plane always resides in the private cloud, however, the control and management plane functions are more flexible, allowing them to utilize the public cloud where appropriate.

1.3. Challenges

Achieving symmetric gigabit speeds is not without its challenges. In addition to the complexities of qualifying new CMs and CMTS for the DOCSIS 4.0 spec, FDX-capable RPDs and amps must be installed in the field. Network equipment between the RPD and the vCMTS, including switches, optics, and network interface cards (NICs), must have the capacity to support many customers with symmetric, multi-gigabit services. The utilization of the radio frequency (RF) spectrum must be changed, since legacy video services are often served in the same frequencies of the new FDX band.

While each of these changes alone is challenging, perhaps the most challenging aspect is the coordination required to deploy these deeply impactful changes while maintaining an outstanding customer experience. In this paper, we will discuss the strategies Comcast is using to deliver the most reliable services via next generation access technologies.

2. Scaling Out DAA Deployment

2.1. vCMTS Automation

The ability to leverage COTS hardware in a vCMTS provides many benefits, such as reduced cost, reduced power consumption, protection against vendor lock-in, and flexibility to change hardware during supply chain crunches, such as those that have occurred since the COVID-19 pandemic. However, these benefits come at a cost since operationalizing a vCMTS is more complicated than a black box iCMTS. At a high-level, the vCMTS edge cloud can be broken down into:

- The primary pod (PPOD): A collection of servers connected to a leaf-pair, collectively forming the Kubernetes cluster in which the vCMTS workloads run [2].
- A converged interconnect network (CIN): the network elements connecting the PPOD to the access network devices (e.g., the RPD).
- The RPDs themselves, which connect to the coax network and ultimately the end users.

In our previous paper “Humanoids Optional: Deploying vCMTS at Scale with Automation” [3], we discussed how learnings from the DevOps movement were used to automate the PPOD and CIN operations [4]. By focusing on flow (automated testing and deployment) and feedback (observability and risk-aware deployment), automation has enabled us to exponentially grow the number of vCMTS clusters deployed at Comcast.

From their inception, the PPODs are assembled in a standard way in the warehouse and shipped to the head-end. Cluster standup pipelines begin by validating all physical layer connections; from that point forward, no humans are directly involved in the initial cluster configuration or any subsequent changes. Rather than assigning operations engineers to manage clusters or network elements individually, the

PPOD and CIN join a logical fleet that makes up the vCMTS edge cloud. Software and network engineers are then empowered to initiate changes to software versions or network configuration templates against the entire edge cloud. The changes are incrementally rolled out across the fleet, with automated health checks ensuring that the change is successful and does not degrade services.

This extensive investment in automation is motivated by our initial experience operating a vCMTS. During our first vCMTS trials, we approached vCMTS operations much like we would a traditional iCMTS, with heavy reliance on operations engineers to configure and maintain PPODs. While the goal was to maintain identical clusters (aside from intrinsic differences, such as which IP addresses were assigned to the clusters), small differences in the hand-configured systems accumulated over time into more significant variation between the clusters.

Over time, this configuration drift made what should have been trivial changes very difficult to implement. Learning from these initial experiences, we invested heavily in end-to-end automation of a vCMTS. Now that each cluster is deployed identically and all changes are automated, we have essentially eliminated outages caused by human error. Additionally, the automated vCMTS standup enables us to bring new vCMTS clusters online within a few hours of it being connected in the head-end, enabling the rapid expansion of the vCMTS edge cloud (Figure 2).

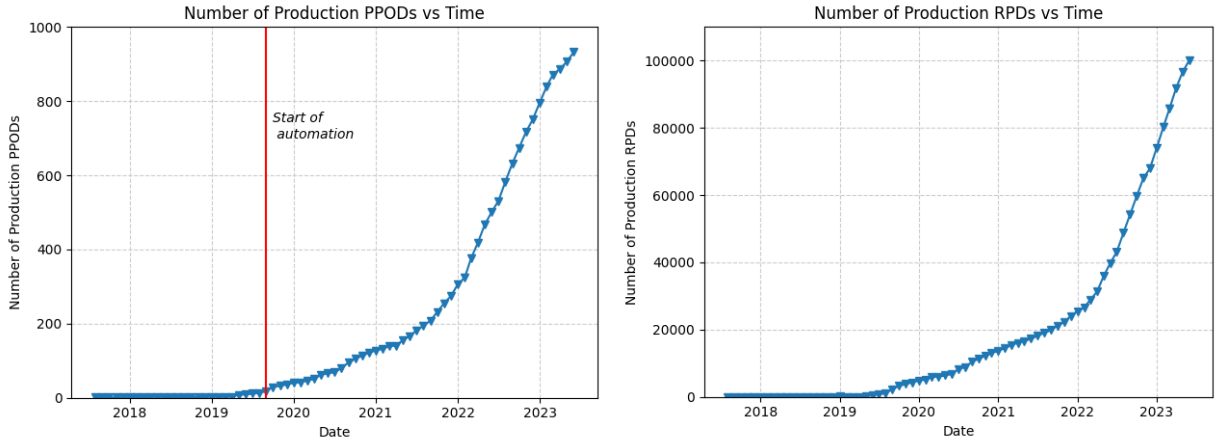


Figure 2 – Number of production PPODs and RPDs deployed vs time.

2.2. Microbursts

End-to-end automation simplifies the debugging of unexpected issues when they arise. Because every system is identical, issues are often either observed in all systems (in which case, the automated health checks that are integrated into every change catch the issue early in the deployment) or in one system.

When issues are present in only one system, the cause is typically attributed to differences in the underlying usage (e.g., number of RPDs, differences in CM models or behaviors). We encountered one such unexpected issue while scaling out our DAA deployment; in a single site, we began observing an increase in CM flapping, where modems are going offline and reregistering or going into partial mode.

To understand the cause of the CM flapping it is helpful to review the CIN architecture. As shown in the simplified figure below, the CIN is based on a typical data center leaf-spine architecture. The hosts are connected to a pair of leaf switches. The leaf switches are then connected to a pair of spine switches. Each

RPD is connected via a 10 Gigabit link to an aggregation switch; other than between the aggregation switch and the RPD in the field, each network hop has a redundant path in case of failure.

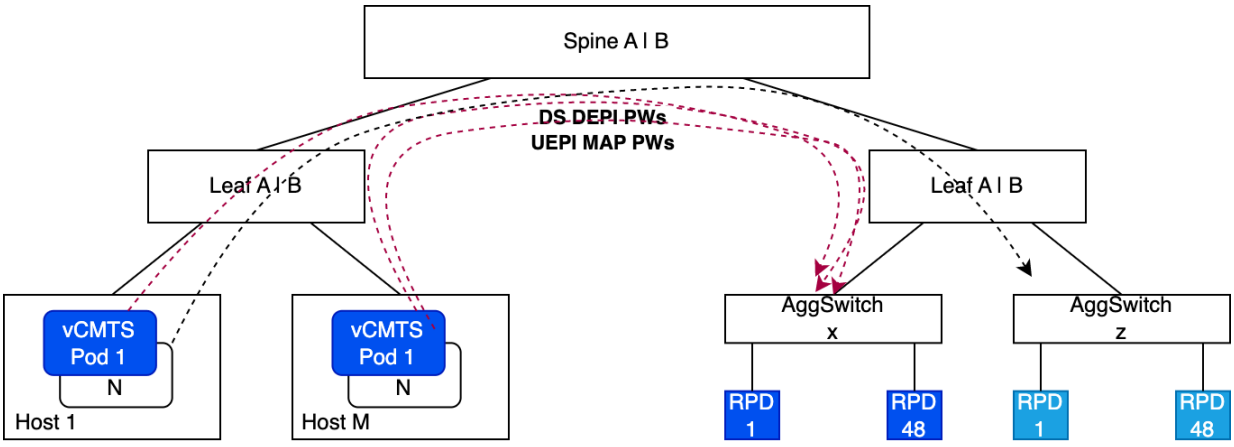


Figure 3 – CIN Based on a Leaf-Spine Network Architecture

Network traffic patterns tend to be site specific (i.e., tied to a specific CIN domain) and are influenced by subscriber and RPD population density. Traffic between the PPOD and RPD is also asymmetric, with a higher bitrate in the downstream direction from the provider toward the subscriber. When investigating the CM flapping, we noticed that this particular site was experiencing microbursts (short spikes in the network traffic) that were saturating the egress queue of the spine. These microbursts can lead to congestion and dropped packets; when the congestion across the CIN is severe enough, this manifests as momentary modem flapping and out-of-sequence packet events.

When investigating the microburst phenomenon, we found that in some situations the DOCSIS map traffic can be a key contributor to microburst activity. Below, we perform a simple calculation to illustrate the network traffic associated with maps. Table 1 shows an example 1x2 service group (SG) configuration with 48 downstream (DS) single carrier quadrature amplitude modulation (SC-QAM) and 6 upstream (US) SC-QAM per port. As was common at the time, every 4th DS SC-QAM was primary capable.

Table 1 – Service Group Parameters

DS SC-QAM	48	32, 36, 40, 44, 48	
Primary Capable SC (every 4th or 8th)	4	12	
US SC-QAM per Port	6	4, 5, or 6 upstreams	
US Ports	2	1x1 or 1x2 SGs	
Voice Calls	0		
Voice Frame	20	msec, G.711 or G.722, default to G.711 as the worst case.	
modems	500	D2.0 is less than 5%; assuming all are 8 channel capable 3.0+	
RNG SM-Internval	14	sec	
T4 Multiplier	2		
Map Size	4	msec	
MDD Interval	2	sec	Secondary MDD is small.
UCD interval	2	sec	
OCD on PLC	200	msec	
DPD on PLC	200	msec	
OCD on ProfiA	500	msec	
DPD on ProfA	500	msec	

To examine the impact of subscriber utilization of the node, we model the downstream data in Table 2. We note that packet size in the downstream direction tends to be bimodal, with packet sizes between 1024 to 1518 bytes accounting for the majority of downstream data. Combined with the layer 2 tunneling protocol (L2TP) header overhead, we arrive at the total packets per second and bytes per second in the downstream direction.

Table 2 – Downstream Data Traffic Model

Subscriber traffic parameters			
SG Capacity (bits/s)	SG Avg Utilization	L2TP Header Overhead - no pkt coalecing, IPv6 L2TP Header	
2,500,000,000	20.0%	67	
DS Packet Size (bytes)	Percent of DS traffic in that size bin	Packets per second (PPS)	BPS /w L2TP Header
1024 - 1518	87.65%	43,101	461,352,085
512 - 1023	1.26%	1,026	6,849,967
256 - 511	1.31%	2,135	7,694,329
128 - 255	2.90%	9,465	19,573,107
65 - 127	5.80%	37,760	49,239,583
< 64	1.08%	10,547	11,053,125
	Totals	104,034	555,762,196

Combining the results from previous tables, in Table 3 we break down the DS DOCSIS traffic for our representative service group into two categories: DOCSIS control mac management message (MMM) data and subscriber data. Notably, while the amount of data utilized by control messages is small relative to the actual subscriber data (7%), the overhead of handling the DOCSIS control data can be significant from a packet perspective. Under these parameters, the ratio of control to data packets was 41%. Given that a CIN may have thousands of SGs, DOCSIS map traffic can represent an unexpectedly large portion of the packets traversing the CIN.

Table 3 – DOCSIS MMMs vs Data Traffic Ratios

		1 SG		
		PPS	Avg Pkt	BPS
Committed Information Rate (CIR)	UEPI Maps to RPD	3,000	120	2,880,000
	MAPs to primary DS	39,000	120	37,440,000
	MDD	25	520	101,920
	UCD	30	530	127,200
	OCD	7	176	9,856
	DPD	7	114	6,384
	RNG-SM	107	211	180,857
	Voice	0	280	0
	Control Sub Total	42,176	121	40,746,217
Data DS PSP Flow 0	Priority - Best Effort	104,034	668	555,762,196
	Total	146,210		596,508,413
	Control vs Data Ratio	40.54%		7.33%

Even before the root cause of the modem flapping was understood, rolling out mitigations was easily accomplished via automation. Ultimately, changing the number of downstream capable channels in a service group significantly reduced the impact of the control packets on the CIN. This analysis has also allowed us to improve our CIN network capacity sizing, demand monitoring, and switch vendor selection criteria, prioritizing for egress buffer capacity. Together, the improved ability to troubleshoot and rapidly deploy fixes is part of why automation significantly improves the overall reliability of the vCMTS.

2.3. Profile Management Application

While end-to-end automation is a hallmark of our vCMTS strategy, some automation features such as modulation profile management are utilized in both an iCMTS and a vCMTS. Modulation profile management involves maximizing bandwidth capacity while minimizing transmission errors. Comcast has hundreds of thousands of node segments, each with its own unique radio frequency (RF) characteristics. Attempting to manually manage the modulation profile configuration of each DOCSIS RF channel for every node segment is futile. For this reason, we developed a profile management application (PMA) that programmatically analyzes the DOCSIS RF channels and adjusts their modulation profile settings.

The DOCSIS radio frequency (RF) channel types that require their modulation profiles to be managed are the downstream orthogonal frequency division multiplexing (OFDM) channel, the upstream advanced time division multiple access (A-TDMA) channel, and the upstream orthogonal frequency division multiple access (OFDMA) channel. Figure 4 shows a cable modem's receive modulation error ratio (RxMER) heatmap. The recommended profile built by the PMA is represented by the solid yellow line showing the subcarriers (bit loading) assignment matching to the RxMER heatmap. Details of the profile management techniques, their performance, and benefits can be found in [5], [6], and [7] .

While modulation profile management is used in both the iCMTS and vCMTS platforms, its implementation varies between platforms. Typically, an iCMTS has a built-in PMA feature. In the vCMTS application framework, a PMA is one of the many applications that interact with vCMTS but which are logically separate from vCMTS itself.

As a cloud-native solution, the vCMTS architecture favors the microservice design pattern. This allows us to decompose a complex system into domains that minimize coupling. As independently deployable microservices communicating through versioned APIs, applications such as PMAs are deployed independently from vCMTS and can run in the public cloud to reduce cost via elastic scaling.

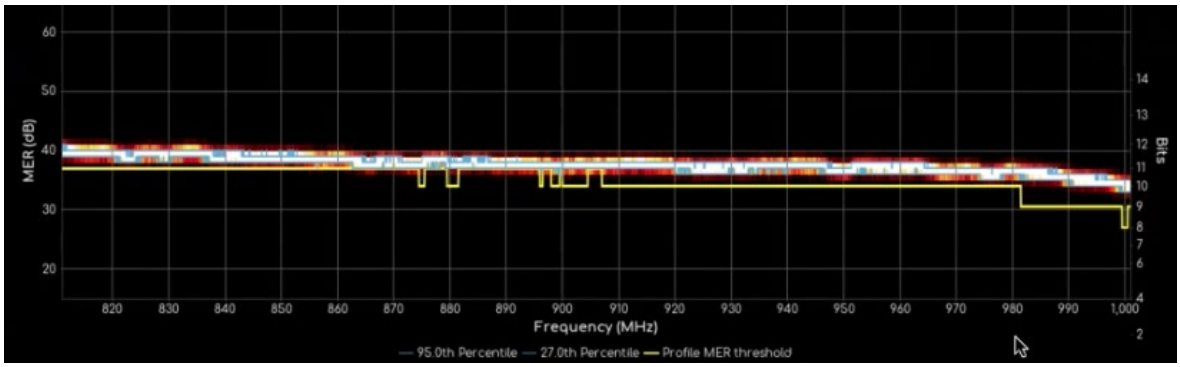


Figure 4 – RxMER Heatmap of a Cable Modem's OFDM Channel

Figure 5 illustrates how the vCMTS application architecture uses a hybrid public-private cloud. The Comcast Edge Cloud is a private cloud deployed as Kubernetes clusters that are physically located in head-ends. Instances of vCMTS are dynamically deployed as Kubernetes pods in the edge cloud and service one or more RPDs. There are currently more than 100,000 of these pods distributed across almost 1,000 Kubernetes clusters in the edge cloud. The PMA resides in the public cloud and ingests vCMTS telemetry data, performs analytics, and recommends profile updates as needed. The recommended profile is set via the service gateway (SGW), which routes the profile change request to the vCMTS pod's API endpoint, ultimately initiating the profile change procedure defined by DOCSIS.

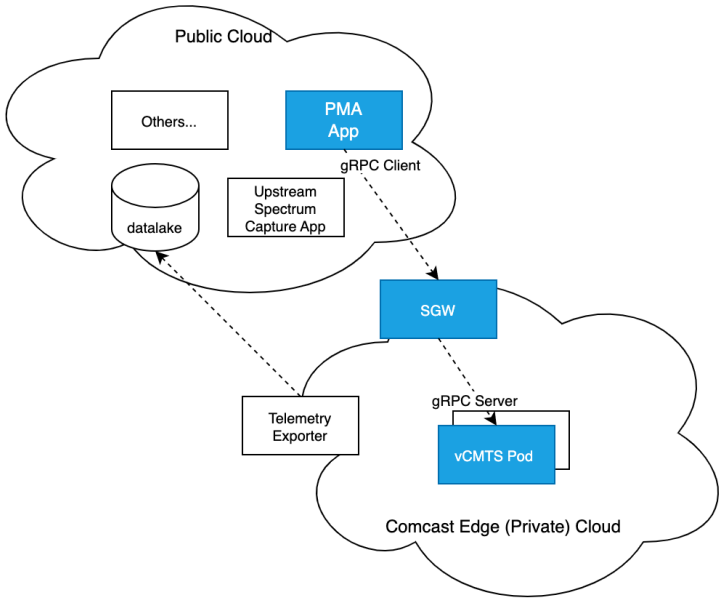


Figure 5 – vCMTS Application Architecture

As more and more nodes are being managed by the PMA, this architecture has scaled well. However, a PMA serving more nodes also means a diversity of customer and RPD devices and firmware versions, which has exposed several interoperability issues. To understand these issues, it is necessary to look a bit deeper into the profile update mechanisms.

When an OFDM profile is updated, the vCMTS must initiate the DS profile descriptor (DPD) change procedure; when an US SC-QAM or OFDMA modulation profile changes, the vCMTS must initiate the US channel descriptor (UCD) change procedure. The change procedures are specified in both the MAC and upper layer protocols interface (MULPI) and the RPHY specifications, since the RPD also plays a role in RPHY.

The diagram below shows the DPD change procedures for the iCMTS and vCMTS. There is a minimum of 500 msec lead time in sending the new DPD MMM to the cable modems. The vCMTS sends the DPD over the generic control plane (GCP) to the RPD, and the RPD updates the "C" bit in the next codeword pointer (NCP). The "C" bit update is basically an odd-even bit toggle to match the least significant bit (LSB) of the change count field in the DPD message, which is used to signify a successful completion of the change procedure.

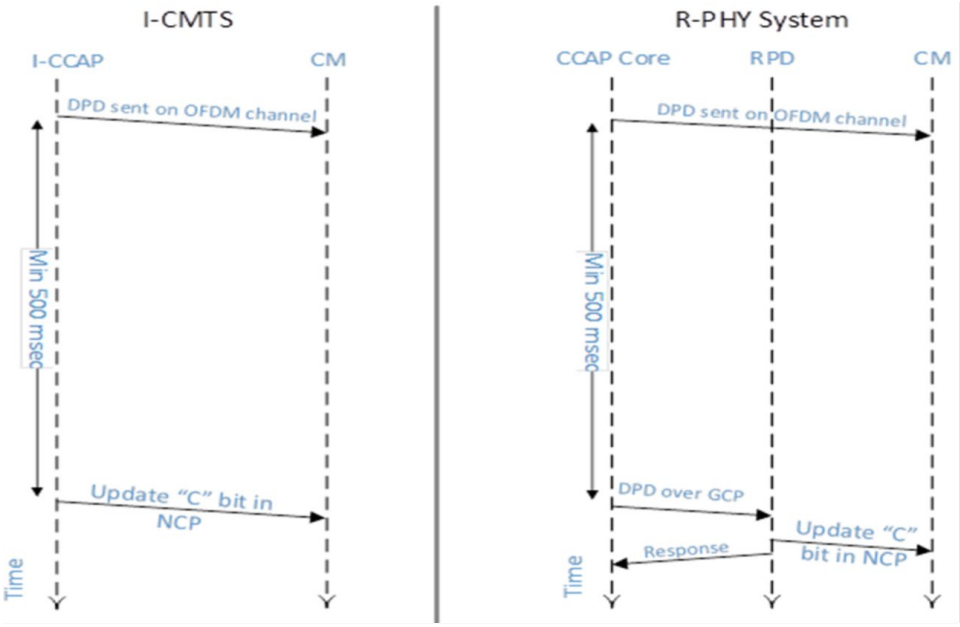


Figure 6 – OFDM profile update in iCMTS vs vCMTS (Source: [1])

When something goes wrong in the change procedure, cable modems utilize the DOCSIS protocol CM-STATUS to report error; this forms a closed loop and allows for recovery. The CM-STATUS reporting implementation varies among the modem chipset vendors. Similarly, RPDs from different vendors vary in implementation behaviors, such as timing sequence, concurrency of multiple DPD changes, timeout handling of "C" bit update.

The failure symptoms include modems going into OFDM partial mode or being unable to receive traffic on the OFDM channel. Figure 7 shows an incident of modems going into OFDM partial mode that was triggered by the PMA recommending a profile change. The time series data labeled "ofdmActual=1x0" (in blue) is the total modem population that acquired the OFDM channel. The other (in green) is the population in OFDM partial mode. As shown, many modems were stuck in OFDM partial mode after the profile change. Such behavior is not unique to the DS, as the UCD change procedures also have their interoperability challenges.

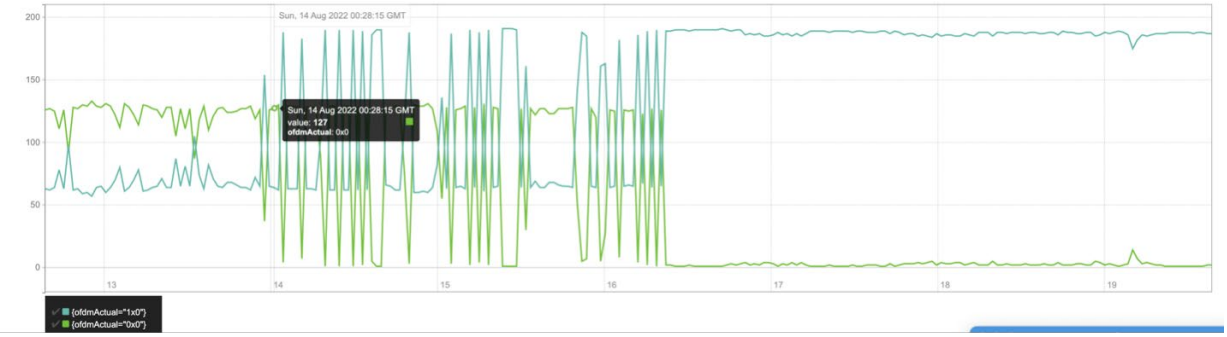


Figure 7 – Total Count of Modems in OFDM Partial vs Time for a Given RPD

We have been fortunate to have an incredible collaboration among both RPD vendors and modem chipset vendors with whom we partnered to identify and fix the root cause of these interoperability issues. Unfortunately, however, the release of new firmware to address issues discovered in the field tends to be a long process. The firmware release, verification, and deployment cycle can take weeks, if not several months.

This situation, where a root cause is known but the fix will take time to implement, is one instance where the end-to-end automation of a vCMTS has delivered enormous value. While a firmware issue may take months to resolve, in many cases, workarounds can be written in a matter of hours. As part of the automated remediations to alerts discussed in [3], the workarounds (often implemented as requests to the vCMTS API) can be triggered in response to an alert.

In iCMTS, a human must be paged to address an alert; in a vCMTS, machines page other machines (microservices) that mitigate the alert in seconds, significantly improving the customer experience by reducing downtime. While a PMA is a feature for both iCMTS and vCMTS, the ability to rapidly address issues encountered during profile management is a testament to the agility and resiliency of vCMTS.

2.4. Midsplit

RF plant capacity is an ever-pressing challenge for both iCMTS and vCMTS. The traditional hybrid fiber coax (HFC) plant has been in place for decades, utilizing a low-split (LS) with an upstream frequency range from 5 MHz to 42 MHz. More recently, mid-split (MS) HFC plants have been introduced with an upstream frequency range from 5 MHz to 85MHz. This additional spectral band can be configured for an OFDMA channel, increasing upstream capacity by more than two-fold.

The benefits and challenges of upgrading a LS plant to a MS one are well documented [8]. The in-Home Assessment Tool (iHAT) as proposed in Ref. [9] was introduced to address the challenges, including:

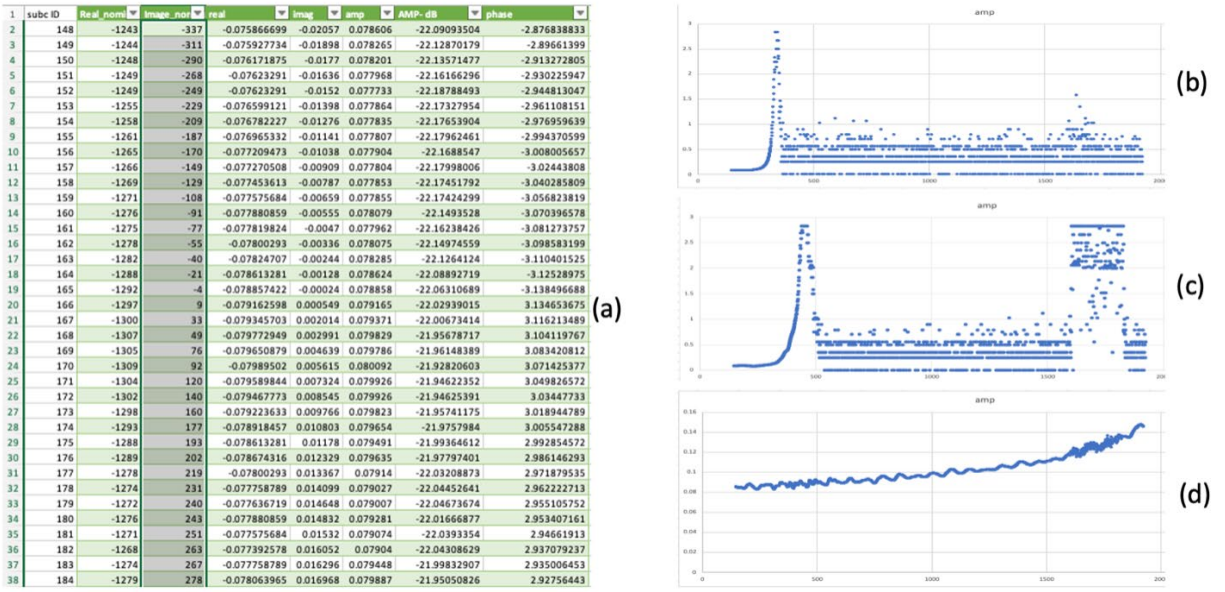
- With imperfect isolation, the MS-CPE's upstream OFDMA signal may leak into the LS-CPE's downstream.
- Active or passive components, such as amplifiers or splitters, can make MS-CPE inoperable on the OFDMA channel. Essentially, these components have a low-pass-filter, such as an LS diplexer. This issue was generalized in Ref. [8] as a 'drop-amp' issue.

iHAT was based on the vCMTS application framework. Utilizing an API provided by the vCMTS, iHAT schedules OFDMA upstream data profile bursts to detect these issues and steer modems' configuration to mitigate their impacts.

One challenge we have experienced with the transition to MS and iHAT was caused by a DOCSIS 3.1 interoperability issue between modems, RPDs, and the vCMTS. Central to this issue was the question: when a modem goes into OFDMA partial mode, how can one differentiate between a "drop-amp" and an actual interoperability issue?

The solution to this was to configure the OFDMA channel ranging zones below 42 MHz spectrum band, which allows the D3.1 modems to be detected. An unequalized probe is scheduled prior to the modem transitioning into station maintenance mode for the OFDMA channel ranging process. To achieve this, the vCMTS was modified to export the first sample of the unequalized probe data. Figure 8(a) shows a sample captured and converted real and imaginary samples into amplitude. Figure 8(b-c) show modems having a "drop-amp" issue. Figure 8(d) shows a cable modem's OFDMA channel response and successful operation on the OFDMA.

The ability to modify the vCMTS, deploy it via ISSU with no customer impact, and integrate it via API with iHAT allowed us to expose the unequalized probe data, which provided a mechanism to unambiguously diagnosis the drop-amp issue.



	subc ID	Real_nomi	Image_nor	real	imag	amp	AMP- dB	phase
2	148	-1243	-337	-0.075866699	-0.02057	0.078606	-22.09093504	-2.876838833
3	149	-1244	-311	-0.075927734	-0.01898	0.078265	-22.12870179	-2.89661399
4	150	-1248	-290	-0.076171875	-0.0177	0.078201	-22.13571477	-2.913272805
5	151	-1249	-268	-0.07623291	-0.01636	0.077968	-22.16166296	-2.930225947
6	152	-1249	-249	-0.07623291	-0.0152	0.077733	-22.18788493	-2.944813047
7	153	-1255	-229	-0.076599121	-0.01398	0.077864	-22.17327954	-2.961108151
8	154	-1258	-209	-0.076782227	-0.01276	0.077835	-22.17653904	-2.976959639
9	155	-1261	-187	-0.076965332	-0.01141	0.077807	-22.17962461	-2.994370599
10	156	-1265	-170	-0.077209473	-0.01038	0.077904	-22.1688547	-3.008005657
11	157	-1266	-149	-0.077270508	-0.00909	0.077804	-22.17998006	-3.02443808
12	158	-1269	-129	-0.077453613	-0.00787	0.077853	-22.17451792	-3.040285809
13	159	-1271	-108	-0.077575684	-0.00659	0.077855	-22.17424299	-3.056823819
14	160	-1276	-91	-0.077880859	-0.00555	0.078079	-22.1493528	-3.070396578
15	161	-1275	-77	-0.077819824	-0.0047	0.077962	-22.16238426	-3.081273757
16	162	-1278	-55	-0.07800293	-0.00336	0.078075	-22.14974559	-3.098583199
17	163	-1282	-40	-0.07824707	-0.00244	0.078285	-22.1264124	-3.110401525
18	164	-1288	-21	-0.078613281	-0.00128	0.078624	-22.08892719	-3.12528975
19	165	-1292	-4	-0.078857422	-0.00024	0.078858	-22.06310689	-3.138496688
20	166	-1297	9	-0.079162598	0.000549	0.079165	-22.02939015	3.134653675
21	167	-1300	33	-0.079345703	0.002014	0.079371	-22.00673414	3.116213489
22	168	-1307	49	-0.079772949	0.002991	0.079829	-21.95678717	3.104119767
23	169	-1305	76	-0.079650879	0.004639	0.079786	-21.96148389	3.083420812
24	170	-1309	92	-0.07989502	0.005615	0.080092	-21.92820603	3.071425377
25	171	-1304	120	-0.079589844	0.007324	0.079926	-21.94622352	3.049826572
26	172	-1302	140	-0.079467773	0.008545	0.079926	-21.94625391	3.03447733
27	173	-1298	160	-0.079223633	0.009766	0.079823	-21.95741175	3.018944789
28	174	-1293	177	-0.078918457	0.010803	0.079654	-21.9757984	3.005547288
29	175	-1288	193	-0.078613281	0.01178	0.079491	-21.99364612	2.992854572
30	176	-1289	202	-0.078674316	0.012329	0.079635	-21.97797401	2.986146293
31	177	-1278	219	-0.07800293	0.013367	0.07914	-22.03208873	2.971879535
32	178	-1274	231	-0.077758789	0.014099	0.079027	-22.04452641	2.962222713
33	179	-1272	240	-0.077636719	0.014648	0.079007	-22.04673674	2.955105752
34	180	-1276	243	-0.077880859	0.014832	0.079281	-22.01666877	2.953407161
35	181	-1271	251	-0.077575684	0.01532	0.079074	-22.0393354	2.94661913
36	182	-1268	263	-0.077392578	0.016052	0.07904	-22.04308629	2.937079237
37	183	-1274	267	-0.077758789	0.016296	0.079448	-21.99832907	2.935006453
38	184	-1279	278	-0.078063965	0.016968	0.079887	-21.95050826	2.92756443

(a)

Figure 8 – (a) Unequalized Probe Data; (b)-(d) Amplitude Response Samples from Three Different Modems

3. Digital Transformation of the Access Network Infrastructure

In the previous sections, we discussed how our end-to-end automation of a vCMTS has enabled Comcast to exponentially grow its vCMTS footprint, mitigate unexpected issues encountered during our path toward deploying more than 100,000 RPDs, and expand capacity over the existing HFC network. Automation, however, is a somewhat vague term since it could refer to anything from a shell script to simplify a repetitive task, to infrastructure as code. In this section, we would like to clarify that our usage of "automation" refers to a digital transformation of the access network infrastructure.

The term "digital transformation" seems to mean something different to every person. Indeed, research on organizations pursuing digital transformations has found at least "23 unique definitions" of the term [10], which was generalized into the following conceptual definition:

> *A digital transformation is "a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies"* [10].

Given the ambiguity, it's helpful to contextualize digital transformations using frameworks such as Quali's "Infrastructure Automation Maturity Model" [11]. While we were unaware of this maturity model during our previous work, it nicely summarizes many of our desired end goals, for example:

- **Dynamic lifecycle**: from conception to destruction, infrastructure is created, modified, or retired according to demand. While maintaining a private, physical edge cloud significantly raises the cost of adding or removing resources relative to simply provisioning/destroying virtual resources in a public cloud, once the business decision to modify the physical resources is made, the process must be automated and seamless.
- **Single source of truth**: there should be no ambiguity surrounding which resources exist or their status, capacity, and configuration.
- **Intelligent predictive change**: while the edge cloud fleet is currently comprised of almost 1000 Kubernetes clusters distributed across the US, they are logically identical and manipulated at the fleet level. Changes are intelligently rolled out to the clusters in accordance with risk, characteristics of different sites (e.g., RPD vendor), utilization, and the results of automated feedback. Furthermore, workloads adaptively utilize available resources, e.g., latency-sensitive traffic is prioritized onto shorter optical transport legs.

We attribute the high maturity level of our CIN and PPOD automation to a few factors. One advantage is that the vCMTS was a greenfield project, with a very strong focus from day one on limiting complexity and unnecessary variation from site to site. Another factor that has contributed to our success was our investment in achieving a true digital transformation of the CIN and PPOD. As we mentioned in our previous paper:

> *After our early work defining our infrastructure in code, we naively thought we had solved the difficult part of the problem. In fact, someone in our organization famously predicted it should take "a few weeks" to whip up the automation to allow us to manage vCMTS at scale. Instead, what began with a few engineers has evolved over the past three years into multiple teams comprised of 15 people. Building capabilities such as self-recovery, complex decision-making logic, risk-based scheduling to name a few took months of hard work by some of our highest performing engineers* [3].

3.1. Concepts

3.1.1. Digital Twin

Crucial to the success of our PPOD and CIN automation has been the creation of a digital twin. A digital twin refers to the concept that alongside the physical infrastructure existing in the real world, a digital representation of the physical entity (its "twin") allows users to inspect and manipulate the physical object itself. While the term digital twin may be relatively new, the idea has been present for a long time. For example, (digital) CAD drawings of a desk can be consumed by a computer numerical control (CNC) machine to transform a (physical) piece of wood into a desk.

The concept of digital twins, however, significantly expands on the machine shop example; the digital entity's lifecycle is in-sync with the physical entity. While a desk may be created via a CAD drawing, deletion of the CAD drawing file has no impact on the physical desk. In contrast, for digital entities such as a CIN switch, manipulating the configuration of the switch's digital twin triggers automation that enacts the corresponding change on the physical switch. And just as the creation of a physical switch begins with the creation of its digital twin, the deletion of the digital twin triggers the deprovisioning of the switch itself.

3.1.2. Global Object Store for Digital Entities

Central to the idea of a digital twin is the concept that the digital world should match the physical world. Echoing domain-driven design [12], this implies that if there is one physical device, there should be only one digital device. Indeed, the automation system that manages the PPOD is the "owner," or single source of truth for the PPOD. In a sense, we benefit from the greenfield nature of the vCMTS, which has resulted in relatively few teams who need to know about the PPOD and CIN. When external systems need information about the PPOD or CIN, they simply integrate with our APIs following the usual microservices pattern.

Because the node is an aggregation point for many backend systems in Comcast, there are many more stakeholders in the RPD lifecycle. As we were investigating how to accomplish a digital transformation of the RPD, it quickly became clear that we needed to rethink how we facilitate easy integration with other teams.

Rather than integrating directly with other teams, we are leveraging the data mesh concept [13]. Information about the RPD will be published into a global object store, allowing any interested consumer to fetch the current (or historical) information about any given RPD. Additionally, consumers can subscribe to notifications, allowing them to receive events when, for example, an RPD's operational status changes.

3.1.3. Data Catalog

In addition to the global store of the digital objects, we briefly want to note that other data is required to maintain the access network. Raw telemetry, for example, is collected by Prometheus and stored for analysis. While the analyzed data may be stored directly on the digital entities, the raw data itself is not persisted in the global object store; instead, it can be linked to the digital entities via references to the external data catalog stores (e.g., data lakes). This ensures that only valuable and well-understood data is exposed to the larger organization.

3.2. RPD Digitization

3.2.1. Use Cases

As was mentioned in Section 3.1.2, there are many stakeholders that either supply the information necessary to provision an RPD or which consume information about the RPD for the purposes of customer support, incident management, etc. In Figure 9, we depict a highly simplified scenario where a new RPD is introduced into the field. A new (digital) RPD is created during the augment planning stage. This triggers procurement and construction to install the physical RPD at the requested location; in parallel, the DOCSIS and video configuration is created for the RPD.

When the device is connected, the provisioning process begins, during which it is authenticated, configured, and ultimately is ready to serve customers. Telemetry is emitted for the RPD and consumed

by many teams, including the capacity and impairment monitoring teams. In the event of an impairment, an incident is created which may result in the device being rebooted or reconfigured. Meanwhile, if the capacity team detects that the RPD is oversubscribed, a request can be made to the augment planning team to start the process of adding a new RPD.

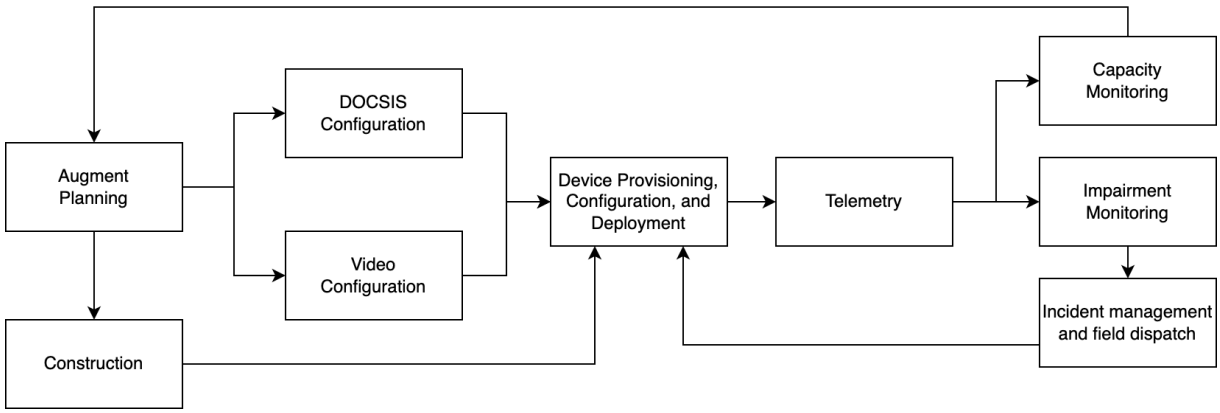


Figure 9 – Simplified Process for Provisioning a New RPD Based on Capacity Demands

Although simplified, this RPD lifecycle shows that many subject matter experts (SMEs) are required to manage an RPD. Teams are often organized around each of those domains. This leads to ambiguity around who is the "owner" of the RPD at any given phase of the lifecycle, since each team owns a different RPD subdomain. Since machine-to-machine communication is required in any highly scalable system, each team will typically have a microservice that exposes information about their subdomain.

However, each also depends on the other teams to build a complete picture of the RPD. The complexity of this model (where each subdomain needs to know about every subdomain to assemble the complete information about the RPD) grows as $O(N^2)$, where N is the number of subdomains or teams involved in the RPD lifecycle.

Phrased in the context of a digital transformation, while there is a single physical RPD, a process-driven framework naturally evolves into a multiverse of digital RPDs. Careful collaboration could in principal result in eventual consistency among these different teams, but achieving such consistency is a daunting challenge.

3.2.2. Model

In addition to employing microservices to separate domains, we employ a data mesh in the operational plane. This allows us to extend the microservices approach while addressing its shortcomings when it comes to scaling in the enterprise, where there can be thousands of microservices, and data consistency becomes a tremendous challenge. Here, the data mesh provides a federated data services framework; as a schema-driven service, the data mesh focuses application teams on defining their data as a product. That data can be linked to other entities via references in the global object store (Section 3.1.2).

For example, in Figure 10, a digital RPD is represented as a composition of device information (e.g., construction plan, its location, etc.), DOCSIS and video config, impairment status, and capacity information. While the provisioning process for an RPD is unchanged from Figure 9, the subdomain teams store their information in the data mesh. Because a complete RPD state can be obtained by traversing the connected objects, this eliminates the need for each team to maintain consistency with every other team. When the DOCSIS configuration changes, for example, the team that manages the

DOCSIS config simply publishes the data to the data mesh for that RPD; teams that subscribe to the DOCSIS config events are notified of the change, and any subsequent query to the data mesh automatically reflects the updated configuration.

Due to its focus on lifecycle management, data consistency, and machine-to-machine communication, we anticipate that a full adoption of infrastructure digitization will significantly strengthen our ability to deliver the highest quality services to our subscribers.

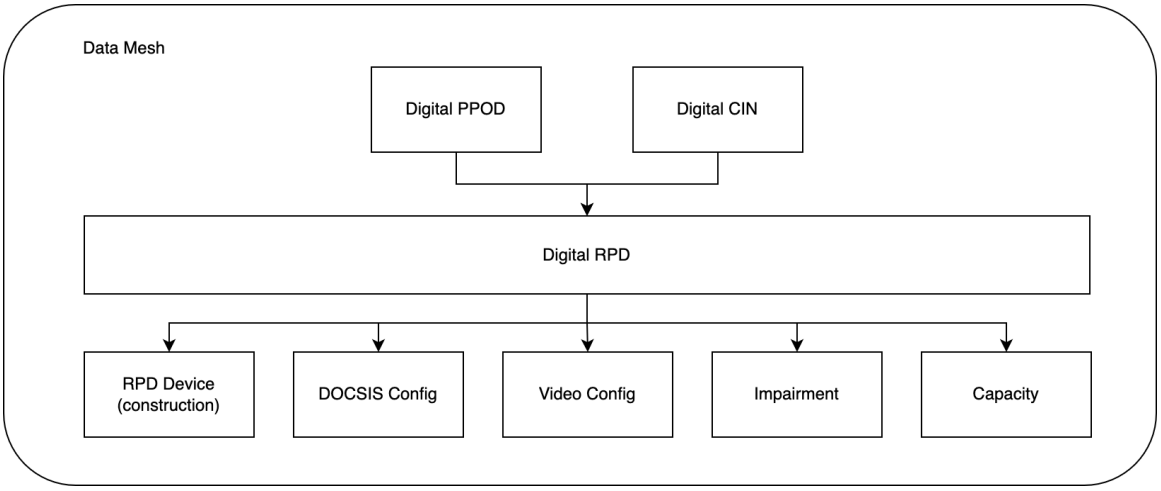


Figure 10 – Digitized RPD and its Subdomains in a Data Mesh

4. Conclusion

At Comcast, we are committed to providing the highest speeds on the most reliable network. We believe that virtualization of the CMTS is a necessary first step toward reducing cost and power consumption of internet delivery. However, the path toward providing symmetric, multi-gigabit services while maintaining an exceptional customer experience requires very tight coordination across the many different domains of the access network. A digital transformation of the access network infrastructure is critical to achieving the level of coordination and observability required to enable a smooth transition to DOCSIS 4.0 and FDX.

Abbreviations

ATDMA	Advanced time division multiple access
CIN	converged interconnect network
CLI	command line interface
CM	cable modem
CMTS	cable modem termination system
COTS	commercial off the shelf
CPE	customer premises equipment
DAA	distributed access architecture
DPD	downstream profile descriptor
DOCSIS	data over cable service interface specification
DS	downstream
EFK	Elasticsearch, Fluentbit, and Kibana
FDX	full-duplex DOCSIS
HFC	hybrid fiber-coax
iCMTS	integrated CMTS
iHAT	in-Home Assessment Tool
IP	internet protocol
L2TP	layer 2 tunneling protocol
LSB	least significant bit
MAC	medium access control layer
MAP	upstream bandwidth allocation map
MHAv2	modular headend architecture version 2
MMM	MAC management message
MS	mid-split
MSO	multiple system operator
MULPI	mac and upper layer protocols interface
NIC	network interface card
OFDM	downstream orthogonal frequency division multiplexing
OFDMA	upstream orthogonal frequency division multiple access
PHY	physical layer
PMA	profile management application
PPOD	primary pod (server rack containing vCMTS)
RF	radio frequency
RPD	remote PHY device
RPHY	remote PHY
RxMER	receive modulation error ratio
SC-QAM	single carrier quadrature amplitude modulation
SNMP	simple network management protocol
SG	service group
SGW	service gateway
LS	low-split
TSDB	time series database
UCD	upstream channel descriptor
US	upstream
vCMTS	virtualized CMTS

Bibliography & References

[1] Cable Labs, "Remote PHY Specification," in *CM-SP-R-PHY-I14-200323*, 2020.

[2] V. Mutalik, B. Gaydos, D. Rice and D. Combs, "Fifty Shades of Grey Optics: A Roadmap for Next Generation Access Networks," in *SCTE*, 2019.

[3] B. Krishnamurthy and G. Medders, "Humanoids Optional: Deploying vCMTS at Scale with Automation," in *SCTE*, 2021.

[4] G. Kim, J. Humble, P. Debois and J. Willis, The DevOps Handbook, Portland: IT Revolution Press, LLC, 2016.

[5] J. Leech and A. Martushev, "PMA Improvements - Strategies Employed for Faster Mitigation, Increased Capacity, and Cost Savings," in *SCTE*, 2022.

[6] M. Harb, J. Ferreira, D. Rice, B. Santangelo and R. Spanbauer, "A Machine Learning Pipeline for D3.1 Profile Management," in *SCTE*, 2019.

[7] K. Sundaresan, J. Zhu, M. Mishra and J. Lin, "Practical Lessons from D3.1 Deployments and a Profile Management Application (PMA)," in *SCTE*, 2019.

[8] L. Zhou, R. Thompson, R. Howald, J. Chrostowski and D. Rice, "A Proactive Network Management Scheme for Mid-split Deployment," in *SCTE*, 2020.

[9] R. Thompson, R. Howald, J. Chrostowski, D. Rice, A. Vieira, R. Vugumudi and L. Zhen, "Rapid and Automated Production Scale Activation of Expanded Upstream Bandwidth," in *SCTE*, 2021.

[10] G. Vial, "Understanding digital transformation: A review and a research agenda," *Journal of Strategic Information Systems,* vol. 28, pp. 118-144, 2019.

[11] Quali, "Infrastructure Automation Maturity Model," 2021.

[12] E. Evans, Domain-Driven Design: Tackling Complexity in the Heart of Software, Addison-Wesley Professional, 2003.

[13] Z. Dehghani, Data Mesh: Delivering Data-Driven Value at Scale, O'Reilly Media, 2022.

