What Can Your CPE Tell You?

Use Cases That Matter When Deriving ML Data from RDK-based Set-tops and Gateways

A Technical Paper prepared for SCTE•ISBE by

Bryan Kelly
Executive Director, Customer Experience Personalization
Comcast Cable
183 Inverness Drive West, Englewood, CO
+1 (303) 658-7561
Bryan_Kelly@Comcast.com

Table of Contents

List of Figures

List of Tables

Introduction

As the landscape of "big data" resolves into tangible progress using machine-level (ML) data and artificial intelligence (AI) to proactively optimize the network, service providers using gateways, digital set-tops and related cloud components that are based on the Reference Design Kit (RDK) are reaping beneficial returns. This paper highlights multiple use cases involving RDK-derived data to provide context about, predict and fix problems before they impact customers, recovery with alacrity when problems occur, and apply anomaly detection to resolve "edge cases" and related challenges that occur across mixed cloud environments.

Operators around the globe are making use of machine-level (ML) data, derived from RDK devices, to feed artificial intelligence (AI) engines and algorithms. Those algorithms, in turn, fuel business- and care-facing dashboards designed to continuously monitor and proactively address events that impact customers. That machine-level data is increasingly mined from in-home devices, including digital set-top boxes and broadband gateways, and based on the RDK's open-source software stack inside system-on-chip (SoC) silicon, both in set-tops and broadband gateways.

For the past few years, representatives from the RDK community have addressed Cable-Tec Expo audiences, detailing progress in business and technical operations and involving Artificial Intelligence (AI) algorithms fed by RDK-sourced machine data. This year, Comcast, as one of the RDK's founders, will detail how it developed a customer experience optimization data platform, called Timeline, which builds a continuous and linear view of each interaction a customer has with the company – be it a series of nonscheduled equipment reboots indicating service problems, or an actual call.

This paper describes how the Timeline software platform, developed by Comcast, creates an empathetic and linear view of its customers' experiences in terms of their interactions with the company including but not limited to signal quality, care-related incidents, product usage and assistance. It was built with APIs that are leveraged by many departments within the company, from product, to care, to technical operations.

Within, we describe three use cases, some in-production and some in-development. These illustrate how RDK-based data derived from set-tops and gateways is being used within the Timeline platform: 1) Determinations of whether broadband gateways are healthy; 2) Measurements of the aggregate and individual health of WiFi-connected devices; and 3) Proactive analysis of error conditions within set-tops to obviate truck rolls and improve the customer experience.

1. The Timeline Story

The Timeline platform began as a simple quest for empathy. With data about customers' experiences invariably scattered across a company, it was difficult (understatement) – both for consumers and service providers -- to really know the whole story about how different products and services were doing. Consider: There are dozens of touch points with consumers, and depending on the nature of the journey, a customer may have questions about or interactions with their existing products, marketing touchpoints, product constructs, online forums, frontline agents (chat, phone) and technicians.

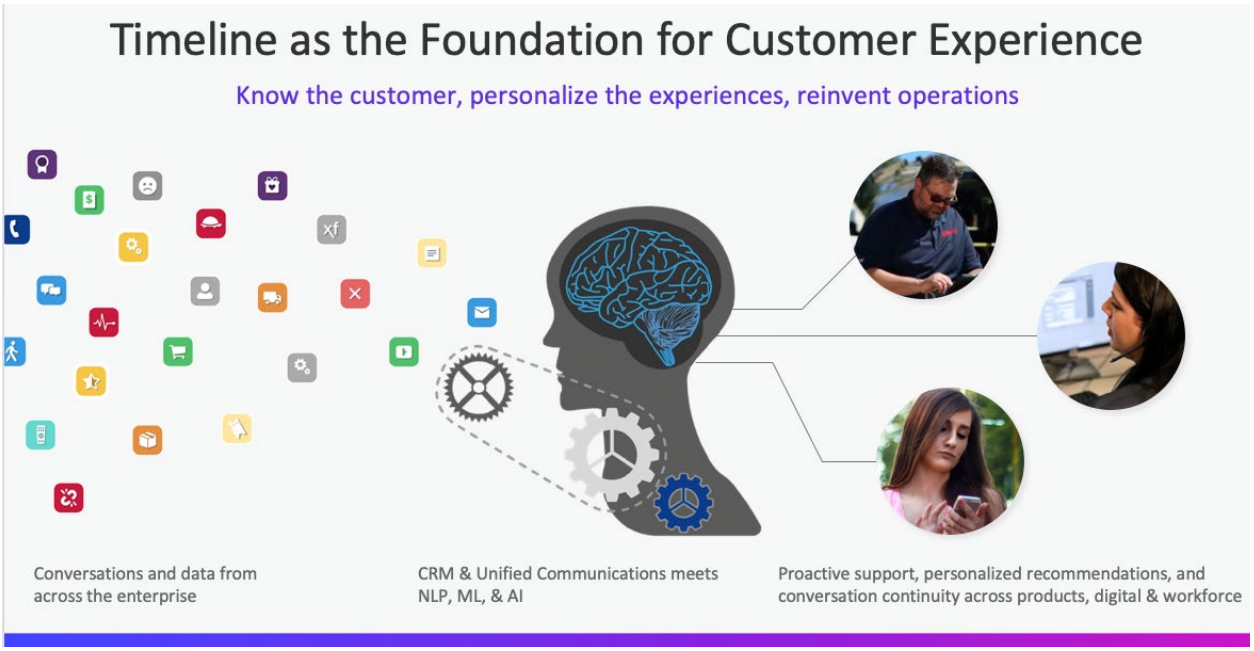

Figure 1: The Timeline Platform

What was missing was a "beginning-to-now" chronological platform, capable of connecting products, workforce tools and any customer interaction, and serving as the foundation for personalized, consistent customer experiences. It needed to be visual, like a dashboard, so that care, dispatch, and any organization focused on making the customer experience our best product could see the whole experience, in one place – both past history and interactions in real-time. With the addition of APIs, any system would be able to leverage that data to provide improved experience context, informing agents why customers may be calling, send proactive communications to customers, offer powerful journey analytics to the business to understand the impact of key CX initiatives and many other data driven initiatives designed to provide differentiated treatment.

That's what triggered what became the Timeline platform, in 2015; since then, it has captured over 40 billion interactions from over 50 million customers, and fields over 25 billion consumer inquiries per month. Internally, it is used by more than 35,000 employees monthly, and has been used by more than 130,000 employees across the Comcast organization. Timeline is integrated with dozens of internal tools, including Einstein 360, Tech360, Xfinity.com and many more. Operationally the platform has been integrated with multiple systems including Slack, Salesforce, DataDog and Pagerduty, providing a seamless, self-service based operational environment across data producers, platform engineers, software developers, product owners and consumers of the Timeline products. In 2016, it was voted "most impactful CX tool," in an internal recognition program; since its inception, it has saved the company millions of dollars by improving first call resolution, reducing truck rolls and providing improvements across the customer experience.

The Timeline Platform has grown to include six product subcategories, detailed below.

1.1. CX Data Platform-as-a-Service

"Timeline Elements" is a data ingest tool that provides (internal) users a simple and self-service means to connect their data. The foundation of any customer experience exchange platform is its data; Elements is envisioned to augment the applicability of an enterprise data lake management tool with data ingest, prep, presentation management, metadata and messaging management, governance, security and APIs. Internal design goals included making it easy and frictionless for various entities within Comcast to submit configured data, so that Elements can automatically build the schemas and trigger the processes necessary to start ingesting, validating, monitoring and visualizing the data (in a non-production environment.)

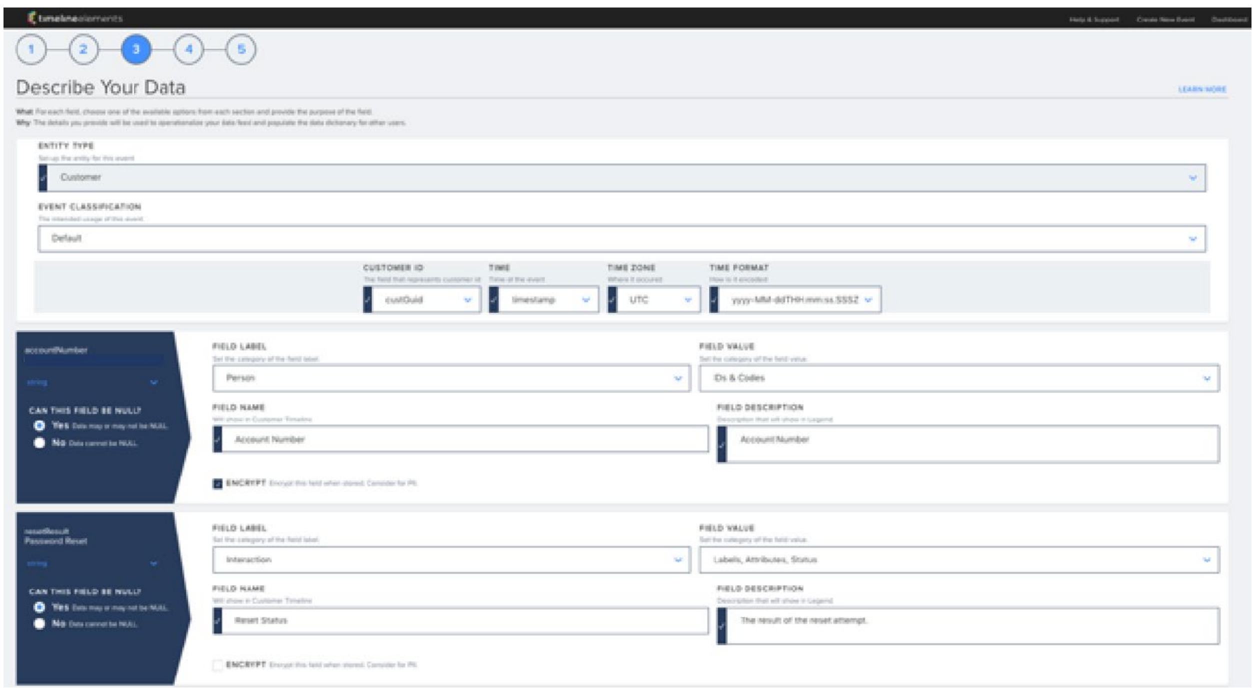

Figure 2: A descriptive data screen from Timeline

1.2. Visualization

"Customer Timeline" is a journey visualization user interface (UI) that sequences customer interactions, including their device(s)/service(s) health, so as to better understand and handle each customer experience. It is an intuitive, 360-degree and lifetime customer view, for employees to see historical and real-time status information, in one place. The intent is to be able to see the big picture, target trouble areas, and dive into specific events. Event details include information about timing, who was involved, and what was done, with an eye toward solving problems faster.

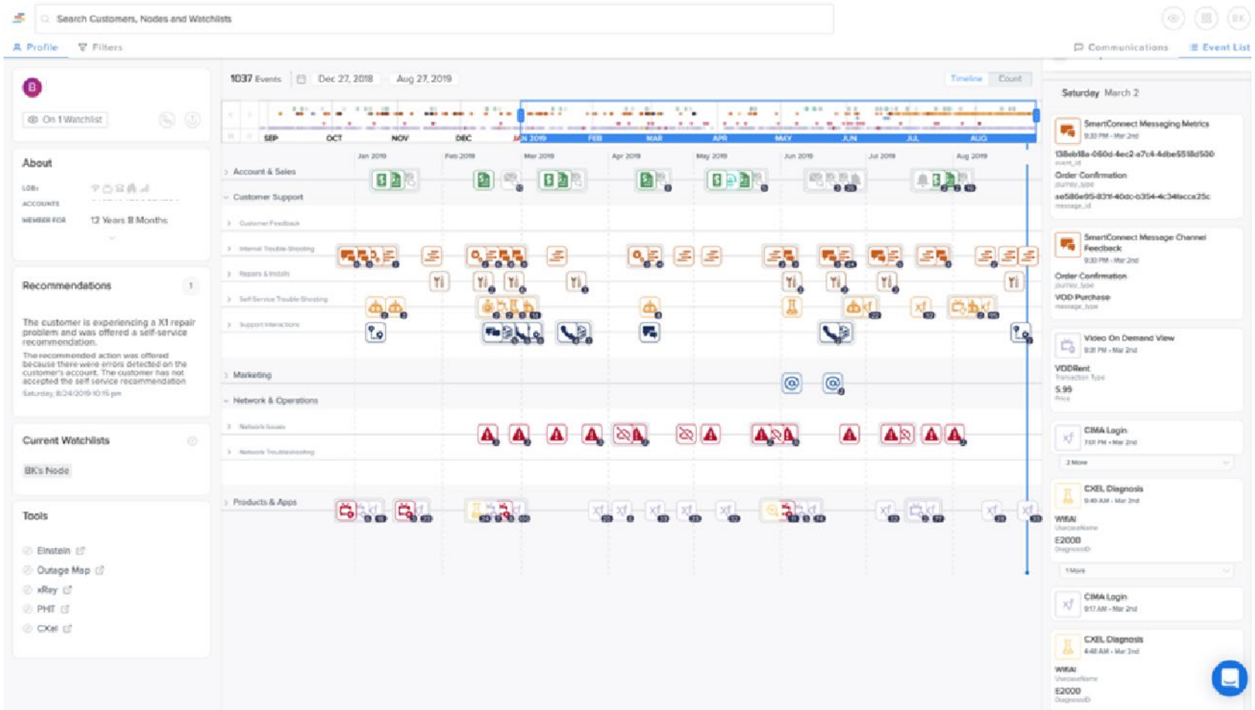

Figure 3: How data is visualized in Timeline

1.3. Messaging

"Smart Connect" is a real-time, automated messaging and notification subsystem, designed to engage the workforce and customers with messaging and conversations supported by machine learning (ML) and artificial intelligence (AI)-based business rules. The reasoning: reactive communications do not don't cut it anymore. Customers are where they are, and are smartphone-trained to start a conversation within an app or other digital experience. The Smart Connect system leverages real-time event data and provides the curated best-next-step to any consuming product, channel or tool via an API. From "Tech ETA" messaging, that provides visual and near-real-time arrival information during service calls, to an automated appointment waitlist, when working from home and available for a sooner service window, to frontline agent, next best action recommendations.

1.4. Proactive Care

"Timeline Watchlist" is a customizable, real-time monitoring tool designed to proactively organize, track and support specific segments of customers enabling differentiated treatment. For context, the feature arose out of a constant frustration amongst field and care representatives, who wanted better ways to stay in touch with customers having problems, to see the matter through. Watchlist is driven by real-time data feeds, so that employees are updated the moment a customer "on the Watchlist" is having a problem. With mainstream messaging and email integrations, employees are empowered to take action on their customers' behalves. The Watchlist dashboard enables live lists, that automatically refresh, to track a static list of customers in care situations, or build lists that dynamically changes, based on set preferences.

1.5. Conversations

"Timeline Talk" will connect conversations across all communications channels (web, phone, SMS, chat, email) to reduce customer effort and personalize experiences. It puts all customer conversations in one place, no matter who talks with a customer, or why. Anyone can initiate and review calls, SMS and email interactions, in one thread. Talk also lets team members to communicate, and update the customer, without switching windows. The reasoning: With cross channel shared context, everyone is on the same page, conversations are continuous and problem solving is simplified. Talk will leverage an integrated and AI-powered virtual agent to triage customer issues and quickly connect them with the correct resolution team. Calls and written correspondence are stored in one place, with advanced search functionality, so that prior interactions and current status data can be retrieved in seconds.

1.6. Analytics

Timeline's analytics platform provides a dedicated analytics data enviornemnt with a standardized structuring, enabling partnerships with leading journey analytics providers Data flows through the Timeline platform and into the analytics environment in real time. This empowers the business to understand the capability of their processes designed to improve the customer experience, the impact of those initiatives on broader business objectives, including simplifhing customer effort through digital transformation and reducing churn. Most importantly, the anlatyics platorm combined with the advanced journey analytics tools, enables the business to leverage data as a guide to understanding the next opportunities for improving the customer experience. For example, an "outlier management" tool highlights customers based on their experiences, and illuminates those experiencing the most frustration. It provides fast insights into specific cohorts of customers – those connected to a downed node, those with chronic problems related to a specific equipment type – to prioritize and take corrective action.

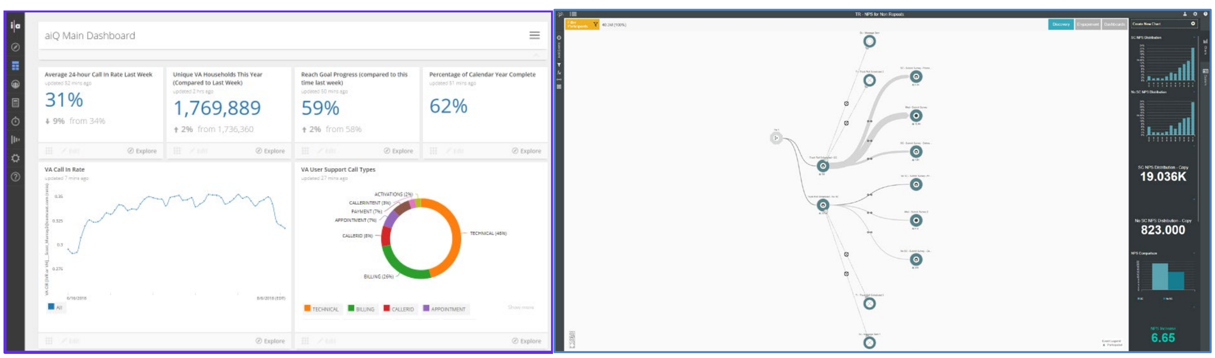

Figure 4: An analytics dashboard in Timeline

2. The RDK Story

The first formal gathering of the RDK community happened in December 2012, in New York, and was co-hosted by Comcast and what was then Time Warner Cable; Liberty Global joined soon after. Its intent, then and now, was to hasten service agility and gain more visibility into the inner workings of digital set-tops, to begin, and later, broadband gateways. Characterized as a royalty-free distribution of shared source components, RDK is intended to be a template and software distro for building CPE software. For system operators, RDK is designed to enable them to control their own destinies, using software and data.

The first well-known outcome of the RDK was Comcast's X1 video experience, at the time a complete break from what had been a blue grid-styled guide. Since then, Liberty's Horizon, Cox's "Contour," Rogers' "Ignite TV" and Shaw's "BlueSky TV" have all been powered by the RDK.

Subsequently, a broadband stack was developed by the RDK, as well as additional profiles to cover additional connected devices. In that sense, RDK is designed to be a modular, portable and customizable open source software suite for the connected home. It standardizes core functions used in video, broadband and IoT devices, providing a common data model and enabling operators to manage those devices and easily customize their UIs and apps. By mid-2019, more than 50 million RDK-based devices had been installed in homes throughout the world.

The RDK software stack is depicted in Figure 5, below.

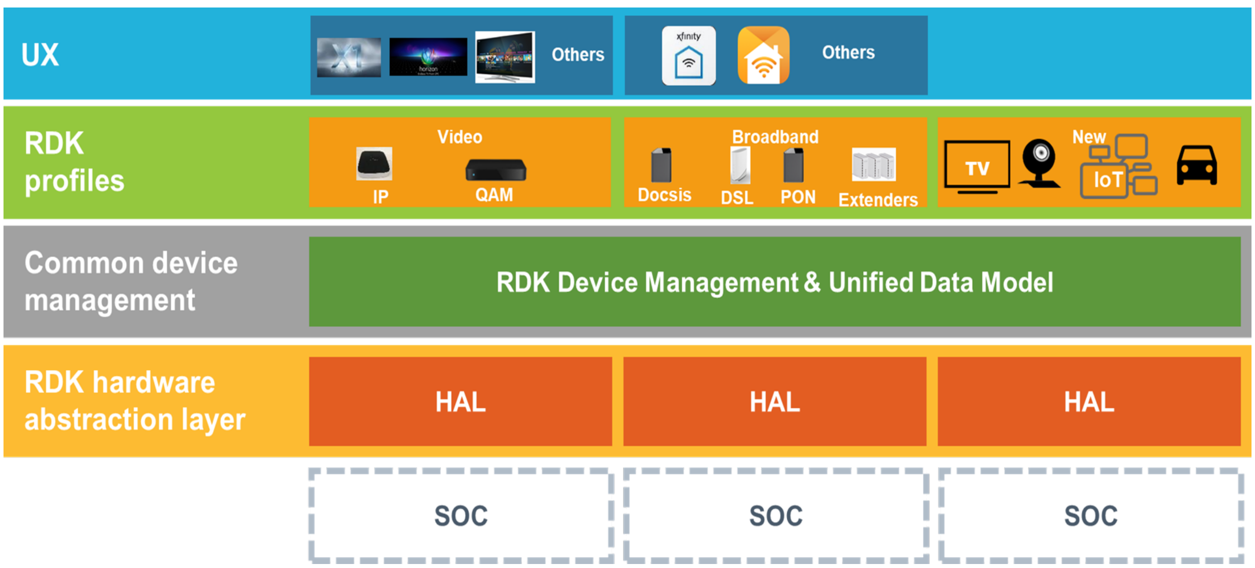

Figure 5: The RDK Stack

For the purposes of this paper, the RDK's contributions to analytics and telemetry, to improve the customer experience and drive business results, will be the focus. Most of that data is derived from the lowest layers of the stack shown in Figure 1 – the RDK hardware abstraction layer (HAL) and device management layer.

Within Comcast's RDK team, some 30 engineers and data scientists routinely monitor telemetry markers, derived from the double-digit millions of fielded digital set-tops, gateways/routers, and connected devices, like cameras. For broadband devices, as many as 30 criteria are gathered, and classified by rules that indicate whether a device's health is good (green) or bad (red.) For instance, a telemetry marker for a home may show that the gateway is green, but the router (and therefore the WiFi) is down. That information is then relayed to internal tools used by care agents.

3. Machine Data Types

Considerable informative data can be extracted from the RDK-based firmware inside digital set-tops and broadband gateways, which can be correlated with other data sets – customer care, infrastructure metrics, in-home signal strength – to proactively optimize CPE and related business systems.

Among the data that can be gathered and analyzed: Dropped packets, channel change responsiveness, voice recognition timeliness, MPEG/IP layer information about video quality, tune/stream errors, and home network fault segmentation (WiFi or wired LAN).

RDK-based analytics can enable service providers to better manage the functions listed in Table 1:

Table 1: Partial list of analytic processes fueled by machine-level RDK data.

Device Trials/Rollouts	Feature Trials/Rollouts
Release Rollouts	Field Issue Triage
Customer Call Reduction	Truck Roll Reduction
Customer Experience Improvements	Functional/Performance Analysis
Trend Analysis	Log/Crash Analysis
Device Optimization	Network Optimization

4. Use Case 1: Broadband Gateway Health

Among the anecdotal evidence that triggered the original RDK, then focused entirely on the video experience, is that prior to initiatives like RDK, it was difficult, if not impossible, to "see" into various software stacks vital to the customer experience. Because of then-monolithic and proprietary code, sourced from multiple internal and external providers, conducting service or equipment triage generally involved two things: Time, and finger-pointing. It follows that a core design goal with the original (video-oriented) and consequent RDK profiles is stack visibility, so as to continuously monitor and assess performance, and proactively mitigate device or service issues.

A category of machine level diagnostic data is being developed within Comcast and designated a "CXEL," pronounced "sixel." It's a linguistic mashup of CX, customer experience, and pixel, to mean the baseline unit of a video image. CXELs can come from several sources, and "go red" when operating out of various parameters. Used to indicate overall health, gateway CXELs inform just part of the picture. When combined with CXELs sourced from the layer3 network and the access/last mile network, they make it possible to know where problems are occurring – is the problem with the WiFi, the last mile network, or the "big" Internet?

The machine-level criteria used to define a gateway-related CXEL is depicted in Figure 6. Across RDK-based gateways and WiFi devices, more than 30 criteria are collected every four hours, to ultimately assess device and overall network health. Markers (left column) that start with "RF" generally indicate DOCSIS-related data. Markers that begin with "SYS" correlate with RDK broadband data. "Range" (middle column) indicates the number of times the markers are reported. The markers and range data is combined to measure the health of the gateway, via the CX-el. If a device meets an aggregate score of 80+, that gateway's CXEL is flagged as red (problematic). Similarly, all markers that begin with "WiFi" (not pictured) are used to develop WiFi-targeted CXELs. If the combined CXEL score for WiFi markers/range is 30+, the device is marked as red / problematic.

A	B	C
Marker	RANGE	Weight
SYS_ERROR_DNSHostname_Error	27	30
SYS_INFO_Transition_RedLED	15	30
RF_ERROR_WAN_stop	3	30
SYS_INFO_Transition_WhiteLED	8	20
SYS_ERROR_Zero_CID	2	60
RF_ERROR_T4_TIMEOUTS	15	30
RF_ERROR_wan_restart	3	30
RF_ERROR_LAN_stop	1	70
SYS_ERROR_CMTSretry_lock_restart	200	20
RDK-10037	15	30
SYS_ERROR_Webpareconnect_PingMiss	4	30
RF_ERROR_IPV4PingFailed	2	40
SYS_ERROR_erouter0link_not_ready	6	30
RF_INFO_CMSTATUS21	15	20
SYS_SH_RDKB_FIREWALL_RESTART	10	30
SYS_INFO_XI5_detect_enabled	2	40
RF_ERROR_WAN_stopped	4	40
SYS_ERROR_erouterLink_down	3	30
SYS_ERROR_Zombie_dnsmasq	2	40
SYS_ERROR_FPM_Pool_InvalidToken	6	30
SYS_ERROR_LOGUPLOAD_FAILED	5	40
RF_ERROR_IPV6PingFailed	1	20
SYS_ERROR_MemAbove600	10	20
RF_ERROR_MDDLost	5	40
RF_ERROR_Wan_down	4	30
SYS_ERROR_Nvram_spacefull	30	20
RF_INFO_CableCut_event	2	30

Figure 6: Machine diagnostic data used to inform CXELs

5. Use Case 2: WiFi & Connected Device Health

When service providers first began offering WiFi connectivity, the most frequent customer calls asked "what's my SSID." After that: "Why isn't the WiFi reaching the back bedroom" / the farther reaches of a home. At the time, the machine-level "hooks" didn't exist in a way that could be correlated, interpreted, and improved. More recently, queries have advanced to cluster around "why isn't the Nth connected device working"?

A CX phenomenon well known to service technicians is the speed test, almost never conducted to actually assess received speeds, but rather because something is working right – a streaming video is buffering, a download is taking too long.

In all cases, the advancement of behind-the-scenes analytics, based on telemetry data, can vastly improve the service experience, to the point that a reduction in speed tests could likely be identified as a key performance index (KPI).

At the 2018 Cable-Tec Expo, panelists on an RDK session discussed at length the notion of a "WiFi Happiness Index," developed as a way to create a weighted, realistic view of WiFi behavior in a home. It involves more than 50 parameters, and enormous amounts of machine-derived data – making it an excellent candidate for AI handling, which gets stronger with increased data loads.

Since then, work has advanced to fine-tune the business of interpreting gateway health, in ways that feed apps like, in Comcast's case, the xFi Assistant, which shows customers things like how many devices are connected, network activity, and how to troubleshoot devices. (It is impossible to discuss xFi Assistant without two bits of color commentary. First: A "pause WiFi" component is known internally as "the child locator," because they come running when Mom pauses the signal! Second, last summer, when an xFi customer retreated to the Colorado mountains with her husband for the weekend, she quickly clued into the party her teenager was hosting at home – by the volume of "device joins" reported via the app; a quick peek into the connected back door camera, after phoning home, showed a pile of teenagers exiting the premises. The story goes on: Later that night, the xFi app's camera captured an intruder, and triggered a photo notification: A Colorado black bear was on the back patio.)

Another of the AI/ML advancements, since last Expo, is the collection of additional telemetry data from broadband and WiFi gateways, then correlating it with other gateways on the same (logical) node, to ascertain whether groups are "going red," indicating localized or systemic problems. When a device or logical node of devices turns red, RDK components are being developed to report descriptive information, which can be exported into ML and AI engines, like Timeline, along with recommended actions – either proactive self-help (along the lines of "we see you're having a problem, try this") or to prescribe corrective action to technicians connected to the Timeline platform via other connected enterprise apps.

It is one thing to see a logical node of gateways "go red," however, and quite another to know if those gateways have truly "gone down." Significant analysis was performed to correlate any red logical nodes, comprised of RDK-based gateways, with other known "bad events," such as unscheduled reboots over a period of time, or T3 timeouts, which are associated with the connection between a DOCSIS device to its CMTS. (Over a certain "pain threshold," too many T3 timeouts signal a larger problem.) Such machine-derived data is being fed into related decision engines, used by other care-oriented engines, to continuously interpret, cross-check and eliminate customer-impacting anomalies.

Work is similarly underway to advance the WiFi Happiness Index (WFHI). As usage increased, it became apparent that viewing the household's WiFi environment, in aggregate, was sub-optimal. Viewing signal clarity and reach, based on certain mean and median throughput thresholds, was insufficient, such as in terms of mean values for RFFI and channel utilization. In essence, prior versions of the WFHI aggregated measurements along different parameters, assessing an overall household snapshot that was largely inaccurate, from a connected device perspective. In reality, and in constant flux, the needs of embedded WiFi radios differ: An iPhone 5 can only do so much, relative to an iPhone XS; likewise for laptops and all other WiFi-consuming devices, and as bounded by their respective silicon footprints.

Therefore a need existed, to assess the overall health of an in-home or in-enterprise WiFi service from the perspective of the devices connected to it, to proactively "see" and welcome them, from a provisioning perspective. (Along the lines of, "Hi, you're new here, welcome, let me get you set up for the best you can do.") Simultaneously, a visual and notification-ready means was developed, to alert technicians and customers of the same, and via the tangents of the Timeline portal.

Behind the scenes, and for a household, the machine-level data available from RDK devices can be applied to an internal "collective pain index." When someone is high on the CPI, corrective action is necessarily maximal. The CPI under development within the RDK sources machine data from five types of devices, shown in Table 2, and examines the normal and pain thresholds for those devices, as a substandard derivation, per device. Such data is enormously useful in deriving maximal and cumulative device-triggered pain in a home, and goes a long way in proactively correcting our own issues that inadvertently impact our customers.

Table 2: Client device types used for WiFi experience telemetry, monitoring and analysis

Broadband Device	Laptop, smartphone, tablet
Media Player	Set-top
Media Source Device	Camera
IoT Device	Thermostat, lightbulb, sensor(s)
Access Point	WiFi extender POD using WiFi-connected backhaul

To summarize, forthcoming versions of the WFHI will include per-device telemetry information, in part to keep customers informed of the realities of their WiFi-connected device capabilities, and in other part to ensure that those per-device needs are being met. This mitigates current customer pain points, like having to tell someone that their IoT thermostat or connected webcam is too far from the gateway. By aggregating the per-device pain thresholds into the larger WFHI, we attain a more realistic and granular view, that can be harvested internally for ML/AI support engines, and, ultimately, by customers, to see current device state.

On the Timeline side, those RDK-sourced CX-els about WiFi and gateway device health are ingested, processed, and correlated with downstream impacts, like call-in rates (CIRs) and truck rolls (TRs). Interested care and field representatives can be proactively alerted, and impacted customers proactively messaged – we see a problem, we're working on it, thanks for your patience – before any CIRs or TRs. In many cases, service calls can be obviated with a notification of suggested action; already, this has saved millions in non-rolled trucks, and earned positive NPS points with customers.

6. Use Case 3: Digital Set-top Health

A third correlation between RDK-derived telemetry data and the AI engines at work to improve field operations and customer experience is the proactive analysis of error conditions with set-top boxes and the resultant video experience. Comcast currently supports millions of RDK-based digital set-tops throughout its U.S. footprint, with the potential to extend the footprint into Europe via its Sky acquisition (and via the addition of additional access network types discussed separately in this SCTE workshop) in the future.

From a machine learning and AI perspective, that aggregate footprint is a prodigious producer of “big data.” A telemetry team within Comcast, responsible for applying telemetry and analytics to set-top data, relates that it gathers upwards of 113 Terabytes of data per day of technical telemetry data, which is equivalent to roughly 4.5 billion daily events, ranging from single key presses to channel tunes to closed captioning and/or secondary language activation.

The data also enables the team to proactively optimize the video experience, such as through stack state analysis. For instance, RDK data, known internally as “XRE data,” for “Cross-platform Runtime Environment,” can represent the last 30 pages a customer saw, before they executed a tune. If a design goal is to perpetually make it easier and more intuitive for consumers to find and consume the content they desire, then the stack state should be optimized to get smaller: A person went through fewer screens to get to a tune event. Telemetry data is a reliable, accurate, and fast means to empirically identify whether or not a feature or change works, in terms of the customer experience.

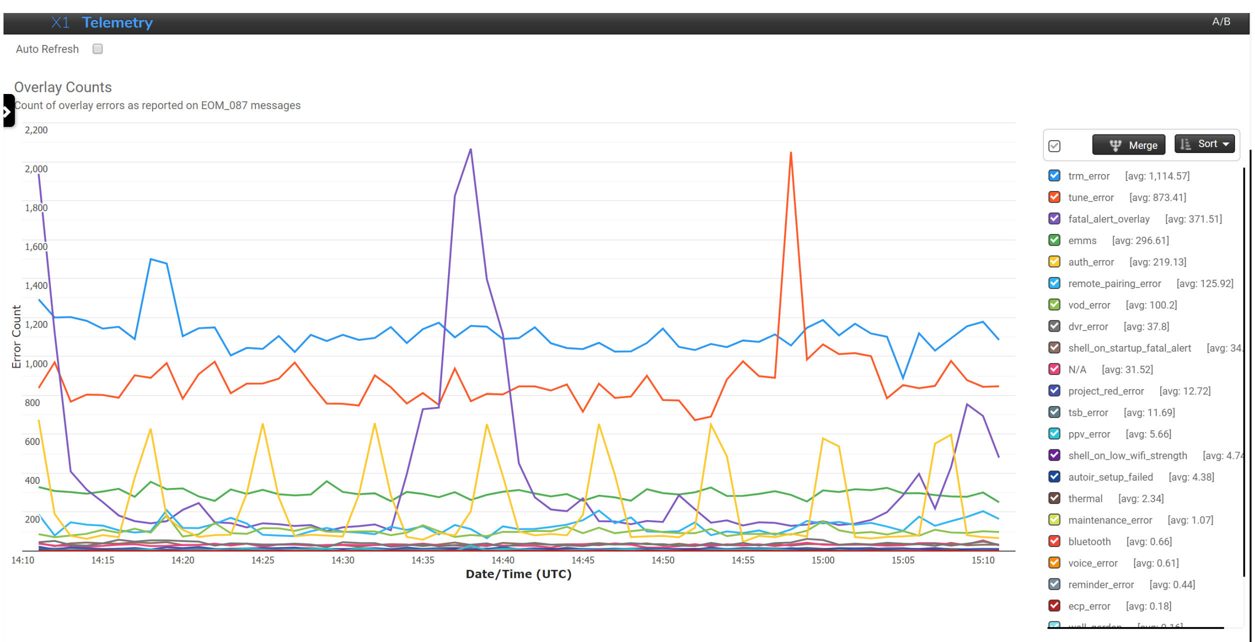

Figure 7: A depiction of STB-related telemetry data

From a field technician perspective, XRE error gathering and analysis has already resulted in the identification and correction of set-top tuning anomalies in some in-home networks that use MoCA (Multimedia over Coax). Last year, the XRE telemetry team began seeing an appreciable increase in XRE-10007 QAM tune failures on two types of RDK-based set-tops. When the devices were in IP-only mode, they tuned successfully, and at a latency similar to a QAM-only or hybrid QAM-IP devices. But if the tune request traveled over a MoCA path to consume a QAM channel, the success rate dropped to the 89% range. The RDK code was consequently optimized to immediately tune via IP, if a MoCA path failed a QAM tune. The action produced video tuning events that succeeded 99.2% of the time.

From a Timeline perspective, XRE data represents a regular flow of data, flowing into the data mart that sits behind the Timeline platform and that can provide context to related internal care systems, like Einstein360, Comcast's agent customer management tool. As a direct result, care agents are proactively alerted to XRE events when speaking with or otherwise helping a specific customer. Timeline orchestrates the event data and makes it available to Einstein360's ITG (Interactive Troubleshooting Guide) function. The ITG guides the agent through the resolution of the problem. It's a useful and fast way to more quickly understand a situation, from problem to resolution, simplifying the conversation. It essentially lets care agents know, with high likelihood, why a customer is in touch, and how to fix the problem.

Conclusion

Data analytics are an invaluable tool to proactictively optimize device and network health. The analytical data derived from RDK-based set-tops and broadband gateways/routers is vital to the service provider's quest to make the customer experience its best product. This paper outlines three use cases, fueled by RDK-sourced data markers, and spanning broadband gateways, WiFi, and video, via set-top boxes. In all cases, the telemetry data available is vast and informative, capable of device and network diagnostics not previously applicable.

When populated into Comcast's internal Timeline platform, these data inform a linear view of each customer's experiences, whether related to signal quality, care-related incidents, and proactive assistance. The Timeline platform began as an internal, employee-generated platform, initially shown at Comcast's thrice-annual "Lab Week" program, and because involved care employees sought a way to keep in touch with customers they'd been helping, to assure that an incident was satisfactorialy resolved.

The landscape of "big data" is showing tangible progress in the use of machine-level (ML) data and artificial intelligence (AI) to proactively optimize the network. Service providers using gateways, digital set-tops and related cloud components that are based on the Reference Design Kit (RDK) are reaping beneficial returns. This paper highlighted multiple use cases involving RDK-derived data to provide context about, predict and fix problems before they impact customers, recovery with alacrity when problems occur, and apply anomaly detection to resolve "edge cases" and related challenges that occur across mixed cloud environments.

Abbreviations

AI	Artificial Intelligence
API	Application Program Interface
BPS	Bits Per Second
CIR	Call in Rate
CMTS	Cable Modem Termination System
CPI	Customer Pain Index
CX	Customer Experience
CXEL	Customer Experience marker
DOCSIS	Data Over Cable Service Interface Specification
ITG	Interactive Troubleshooting Guide
FEC	Forward Error Correction
HAL	Hardware Abstraction Layer
HFC	Hybrid Fiber-Coax
HD	high definition
Hz	hertz
IP	Internet Protocol
ISBE	International Society of Broadband Experts
KPI	Key Performance Index
LAN	Local Area Network
ML	Machine Learning

MPEG	Moving Pictures Expert Group
QAM	Quadrature Amplitude Modulation
RDK	Reference Design Kit
SoC	System on Chip
SSID	Service Set Identifier
SCTE	Society of Cable Telecommunications Engineers
TR	Truck Roll
UX	User Experience
XRE	Cross Platform Runtime Environment
WFHI	WiFi Happiness Index