Beegol

THE END OF THE TRUCK-ROLL:

Can Machine Learning diagnostics improve the Broadband experience and reduce operating costs?

Beegol Whitepaper

Contents

1 ABSTRACT

Providing a high Quality of Experience (QoE) to the broadband end user requires every component in the network to work well, whether it is the in-home WiFi network, the gateway, the local access network, or the transport network. Along the way Internet Service Providers (ISPs) started to develop tools for identifying and locating network problems, but they continue to be challenged to get a holistic view of the health of all the network components needed to provide a great QoE. The industry is in great need of a platform that allows operators to identify, triage and locate network problems by simultaneously looking at every component. The platform must be able to collect any data, anytime, anywhere, from in-home up to the transport network, and request additional data on demand. In this white paper, we present several use cases developed for a large ISP in Brazil and provide a path towards addressing these challenges using telemetry and machine learning.

2 INTRODUCTION

Highly reliable broadband service is critical, especially in the current remote work and home-school environments. Connection problems such as performance degradation and outages quickly erode customer satisfaction and increase the ISP's operational expenses. When a customer has a problem they often call into a Customer Care Center, which may follow up with a truck roll. In many cases it is hard to locate the root source of a broadband problem quickly and accurately, resulting in unproductive calls, unnecessary repair visits, and increased costs for the operator.

Network diagnostics and healing have become even more challenging because of the highly complex eco-system for multimedia services: the application provisioning required to stream a movie involves the in-home WiFi, the local access broadband, the transport network, and the application CDN, among other elements. If any of these components are not working correctly, the user QoE degrades, and the customer likely calls the ISP to complain. Current network diagnostics primarily focus on specific network problems within specific components and do not provide a holistic view into the end-to-end ISP network. There is a real need for an **end-to-end data-driven machine learning platform capable of automatically identifying and locating network problems and solving them remotely whenever possible.**

In this paper, we explain several use cases that highlight deficiencies in the current diagnostic technologies, with a focus on those that are harder to diagnose or require data collection flexibility. For clarity, we will use the following terminology:

GATEWAY:
customer premise equipment (CPE) including modem, modem with a WiFi router (usually called integrated gateway) such as MTA for DOCSIS or ONT for GPON

END USER DEVICE:
device connected via WiFi/Ethernet to the Gateway, including computers, mobile phones, TV's, and similar devices.

LOCAL ACCESS BROADBAND NETWORK:
connection between the CMTS up to the Gateway, including Fiber nodes and the CMTS for DOCSIS, or between the OLT up to the Gateway for GPON.

TRANSPORT NETWORK:
network upstream of the CMTS or OLT, including the IP routers.

3 PROBLEM STATEMENT

The current broadband ecosystem consists of several components, all of which are key to delivering a good QoE. These include integrated end user devices, gateways, local broadband access (DOCSIS or GPON), the transport network, and additional key components such as DNS (Weifan Jiang 2021), CDN, DHCP, and game servers. All of these must work properly for the user to have a QoE sufficient enough to generate high levels of satisfaction. Although some of the components may not belong to the ISP (e.g., the user can have their own router, game servers can be hosted outside the ISP network, DNS may belong to the application service, etc.), the user ultimately holds the ISP responsible for internet service.

Given such a complex environment, it is hard to identify and locate problems quickly and accurately. When problems arise, they can happen in any part of the network from the physical layer up to the application layer. Problems with similar symptoms can have different root causes. Transient problems may not be visible during the diagnostic phase which can lead to ISPs encountering unproductive repair visits and increased costs.

The typical ISP has tools that assess each network component individually. This approach takes periodic 'snap-shots' of each component, and cannot offer an integrated, end-to-end view with real-time data. For example, Proactive Network Management (PNM) tools only diagnose physical layer problems, WiFi platforms manage the in-home WiFi service, Network Operating Center (NOC) tools can only manage transport networks, and Speed test tools only show certain elements of a user's Quality of Service (QoS). While these independent tools work for individual pieces of the network, they do not provide a comprehensive view, which presents several challenges:

- Most of these tools take 'snap-shots' at a predetermined frequency. If the data collection frequency is too high, they can overflow the network management system with data. If it is too low, they might not get enough data to locate transient and intermittent problems.
- Since they are independent tools, the snap-shots may be taken at different times, so the ISP does not have a complete view of what is going on at any single point in time.
- It is usually not possible to collect additional, on-demand, real-time data needed to complete a diagnostic analysis.
- Data collection is not event driven. Latency, for example, can be due to congestion or may just be a transient behavior. To determine the cause, we need to collect real-time, granular data precisely when the event – congestion – occurs.
- It may not be possible to add new data collection mechanisms when needed. Some of the tools allow changing collection frequency, but adding a new mechanism requires a firmware upgrade. For example, an upgrade is needed to enable laser degradation diagnostics on a PON device to determine product lifetime and service requirements.

4 USE CASES

Below are several use cases from a large ISP in Brazil with more than 10 million gateways (including DOCSIS and GPON). Each case demonstrates deficiencies with current diagnostic tools used by this ISP to monitor their network.

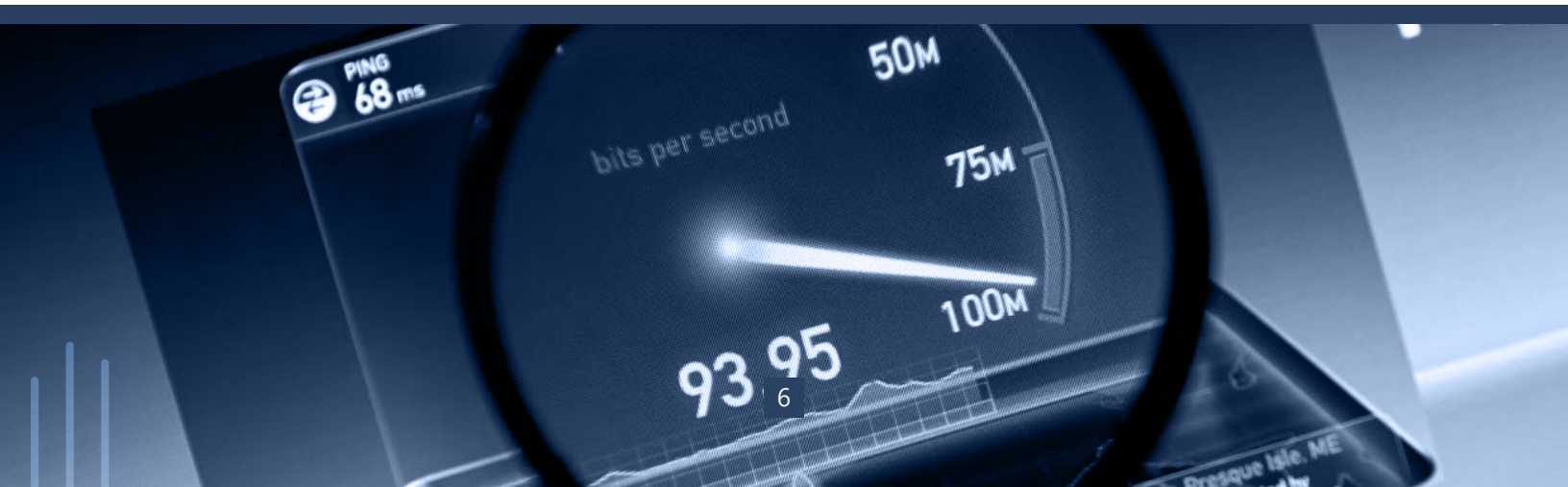
4.1 Speed test QoS and Proactive Network Maintenance (PNM)

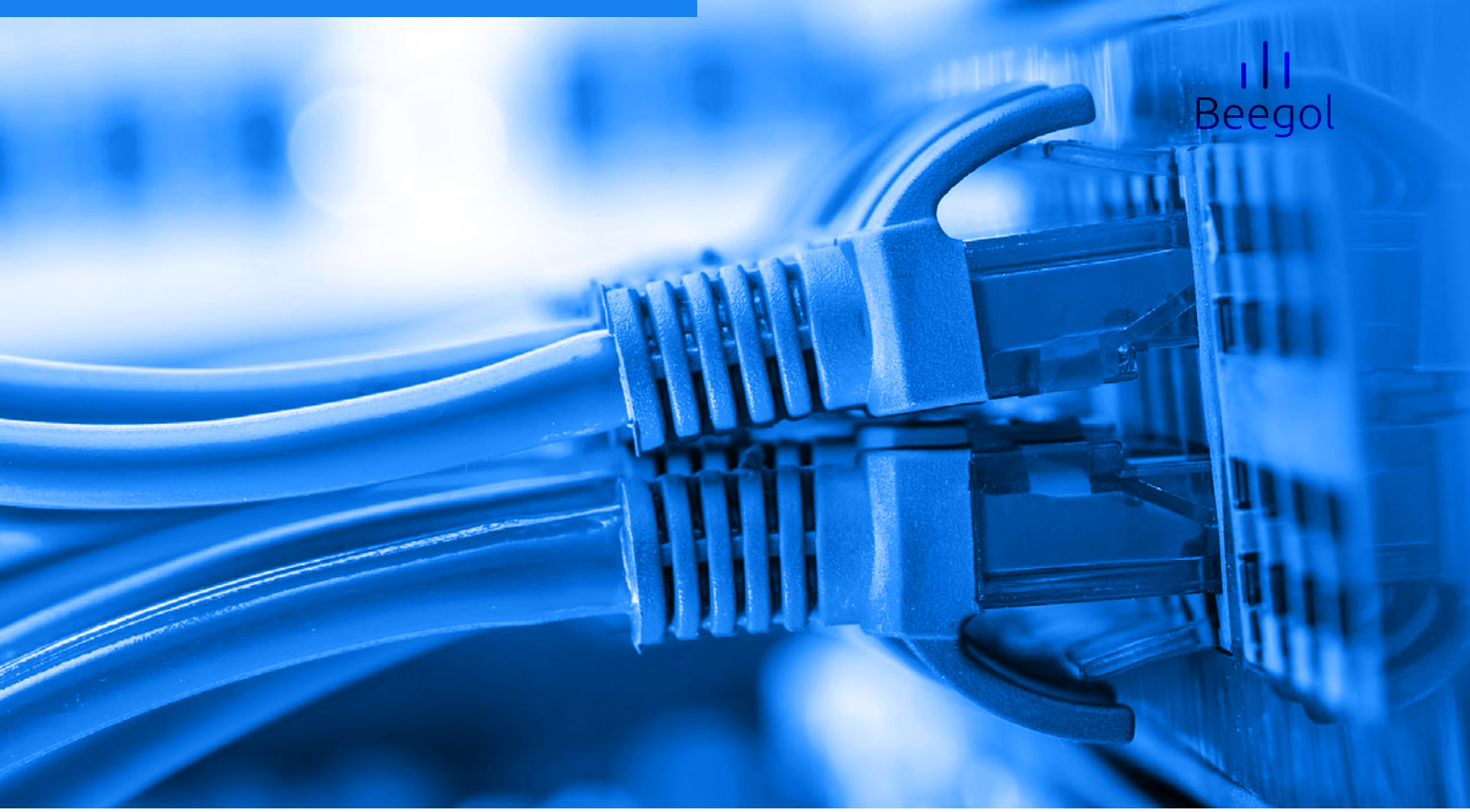
The Speed test gives the ISP the QoS of the end user and can precisely diagnose a broadband degradation. Current Speed test tools can have several pitfalls depending on indvidual implementation:

- They run from an end user device which includes the WiFi layer performance, i.e., they do not reflect the true broadband performance. Few ISPs have Speed test that run directly from the modem.
- They only take measurements against a fixed end point in the network. It is not possible to test latency, for example, against different end points in the network.
- They cannot run at different time frequencies, i.e., a Speed test run every two hours for a broadband with degradation, yet only three times a day for the rest of the network.
- They cannot run a real-time Speed test when user calls the Call Center.

Ideally, Speed tests should run directly from the gateway to infer broadband quality without WiFi interference. It should check the client's bandwidth usage before running so it won't disturb user QoS. It should confirm that only one device is running at a time on each node/splitter to avoid node saturation The test should measure downstream throughput, upstream throughput, latency, jitter and packet loss, against different end-points including local access and off-net. The ISP should be able to run the test at regular yet different time frequencies depending on the status: every two hour on gateways with access problems, and every eight hours on all other devices. The Call Center should also be able to run a real-time Speed test and look at historic broadband measurements.

If the test compares every single user in the network with a single metric, providing a clear view of the average user QoS by node, geography, city, or operation, ISPs can prioritize users with the most severe degradation problem. A mobile application that shows the Speed test series to the end user will further enhance transparency, increase subscriber loyalty, and avoid technical calls.





PNM tools can identify and locate cable access network problems (Labs, Best Practices and Guidelines, PNM Best Practices: HFC Networks (DOCSIS 3.0) 2016) (Labs 2020). The CableMon paper by Jiyao et al (Jiyao Hu 2020) analyzed the performance of existing PNM tools, finding that

- There are too many false positives when an ISP uses the standard PNM standard for diagnosis. For example, if an ISP uses the recommended Main Tap to Radio (MTR) threshold, more than 24% of cable modems would require repair.
- Traditional PNM tools cannot identify all problems, especially intermittent problems. If the PNM collects data at limited time intervals, it may not be able to identify intermittent issues.

A PNM tool by itself can generate low results with respect to the total number of problems identified and many false positives. In the study, CableMon used technical Customer Care calls to label if a client had a local access broadband problem or not. Whenever the customer called for a technical problem, they assumed there was a local access problem. The goal was to estimate the PNM parameters which maximize the match between the technical call rates and the PNM positive diagnostic. It was able to detect 81.9% of the anomalies which lead to the generation of a customer service ticket. In addition, CableMon predicted 23% of network-related trouble problems. It shows that to improve accuracy and precision, PNM tools must therefore be augmented with a "ground truth" user QoS indicator that can accurately state if the user has a broadband degradation problem or not. Ideally, a **PNM tool should be integrated with a Speed test to improve accuracy and precision and reduce unproductive truck rolls.**

Additionally, intermittent issues can only be captured by looking at PNM data on a very granular timeline. PNM data needs to be collected at different time intervals to accurately identify problems: each half hour, every minute, or even every second.

4.2 Measuring latency for estimating end user performance

A recent technical report by the Broadband Internet Advisory Board Group (BITAG) ((BITAG) 2022) explained that **a high QoE requires not only high throughput but also low latency.** It also stated that latency must be measured from the end-user and under working load conditions – known as working latency. They measured latency at regular time periods, as well as working latency within a given two-minute time-window during workload. They concluded that in order to measure latency that affects user performance, the tool must:

- measure latency from the user viewpoint but without any WiFi interference thus, measuring broadband latency from the gateway;
- measure latency against different endpoints in the network, for example, local access latency, off-net latency, key interconnection points latency;
- measure latency within different time-granularities, for example from 30 minutes to every second. Fine time granularity latency measures are key for identifying congestion; and
- measure latency driven by events (i.e., suspected congestion).

Most latency tests do not consider all of these measurements, and therefore cannot accurately measure end user performance.

4.3 Detecting Congestion on the Transport Network

In a large ISP, call rates suddenly increased by 30% within a large city. Operations looked at potential root causes but found none associated with the local access. Without a congestion detection algorithm, the ISP could not identify potential problems in the transport network.

An effective congestion detection algorithm should comply with the following methodology (David Clark 2014):

1. **Measure latency for each user on different network endpoints, e.g., local access and off-net measure every 10 minutes. When latency exceeds a minimum threshold of 100 milliseconds, the measurement frequency increases to every second;**
2. **Identify the common routes with high latency at similar periods on every day of the week;**
3. **Run several traceroutes to identify and validate the IPs of the high latency routes; and**
4. **Ping individual routes frequently to validate high latency and congestion at specific dates and times.**

Following this framework provides the ISP with all the routes with high latency, as well as the users affected by it daily. Figure 1 is an example that shows the congested routes identified by the algorithm created for a large ISP in a given week . The red boxes indicate that latency was above the minimum threshold on these routes during the measured time. Without an algorithm to detect congestion, the Network Operations Center (NOC) has a view of routes with high utilization but cannot identify which users are affected by it and cannot see second by second congestion.

trace_route	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
10.11.128.1|10.11.128.1|191.179.64.10|189.3.85.1|200.244.18.249|	3	3	2	2	1	2	2	1	0	2	3	2	0	1	1	1	0	2	0	1	1	4	3	3
192.168.198.1|10.10.0.1|179.233.128.1|189.87.140.33|200.244.211.215|	1	2	1	1	3	0	0	1	2	0	3	0	2	2	2	1	2	3	3	2	2	0	0	1
192.168.198.1|10.51.80.1|201.17.32.66|201.17.0.70|201.17.34.146|	1	2	1	2	0	0	1	0	1	2	1	0	2	1	2	0	1	3	3	5	4	3	3	3
10.10.0.1|10.10.0.1|187.180.224.1|189.87.155.141|200.244.19.150|	2	1	1	1	2	1	2	2	1	1	2	0	0	0	1	3	4	1	2	3	2	0	0	1
10.11.128.1|10.11.128.1|191.179.64.10|189.3.85.1|200.244.18.251|	1	2	2	2	1	1	3	2	3	0	0	1	0	0	2	0	1	1	0	0	1	2	0	0
192.168.198.1|10.11.128.1|191.179.64.10|189.3.85.1|200.244.18.249|	0	1	3	0	2	0	1	2	1	1	3	0	0	1	0	1	3	1	0	2	3	0	0	0
192.168.198.1|10.35.128.1|187.122.188.1|200.167.43.21|200.244.216.56|	1	1	0	1	2	2	1	3	2	1	2	2	0	0	0	1	0	0	0	2	0	1	1	1
10.31.128.1|10.31.128.1|201.17.39.226|201.17.34.213|201.17.34.209|	1	0	0	2	2	1	2	4	1	1	0	1	1	1	3	1	1	3	1	1	0	1	2	0
10.10.0.1|10.10.0.1|179.233.128.1|189.87.140.65|200.244.211.211|	2	0	2	0	0	1	1	1	0	2	0	1	1	1	2	4	1	0	1	2	0	2	0	1
192.168.198.1|10.10.0.1|187.180.224.1|189.87.155.141|200.244.19.150|	0	0	0	2	1	0	1	1	3	1	1	2	2	1	1	0	1	1	1	2	0	1	0	0
192.168.198.1|179.211.128.1|187.122.188.10|201.73.3.101|200.244.216.85|	0	0	0	1	0	2	0	3	0	0	2	1	1	0	0	0	2	0	0	0	0	0	0	0
177.195.120.1|177.195.120.1|187.122.188.1|189.87.140.89|200.244.211.211|	1	0	0	1	0	1	0	1	1	2	2	4	0	0	0	2	1	1	0	1	1	1	1	1
10.13.128.1|10.13.128.1|179.233.128.2|201.73.3.45|200.244.216.85|	0	1	1	1	1	0	2	2	0	0	3	1	0	0	2	0	0	1	1	1	1	1	0	1
192.168.198.1|10.35.128.1|187.122.188.1|201.73.3.97|200.244.213.156|	1	0	0	2	1	1	2	1	0	1	0	1	0	1	0	2	1	1	0	0	0	1	1	0
192.168.198.1|10.11.0.1|179.233.96.2|189.52.252.169|200.244.19.140|	1	1	0	1	1	0	2	2	1	1	1	0	0	0	1	0	0	2	1	1	0	1	1	0
192.168.198.1|10.13.128.1|179.233.128.2|201.73.3.41|200.244.216.85|	0	0	2	0	2	1	1	0	1	1	1	0	1	0	0	1	0	2	0	1	0	0	0	0
192.168.198.1|100.81.64.1|100.81.64.1|201.17.33.130|201.17.0.70|	1	1	0	1	0	2	1	1	1	0	2	1	2	1	0	0	1	0	1	0	1	0	1	0
192.168.198.1|10.35.128.1|187.122.188.1|189.87.138.25|200.244.211.201|	0	1	2	1	0	0	2	0	1	0	0	1	0	0	1	0	0	0	0	0	1	1	0	2
192.168.198.1|179.211.128.1|187.122.188.10|189.87.140.1|200.244.211.215|	0	0	0	1	0	1	2	1	1	1	1	2	1	0	0	1	0	0	0	0	0	1	2	0

Figure 1 – Congested routes

Additionally, some of the problems are very hard to detect, such as:

- Problems with a router's firmware upgrade which caused the load balance to work incorrectly, increasing latency exponentially.
- Heavy congestion on an interconnection with another large ISP (ex. A bottleneck with the other ISP can affect latency of users accessing CDNs (Amogh Dhamdhere 2018) on the primary ISP.)
- Lack of utilization optimization of some intra-city metro network routes.

Figure 2 shows the latency reduction on a key route after a congestion diagnostic and ISP action plan. The latency decreased from almost 100 milliseconds to approximately 20 milliseconds on this key route.

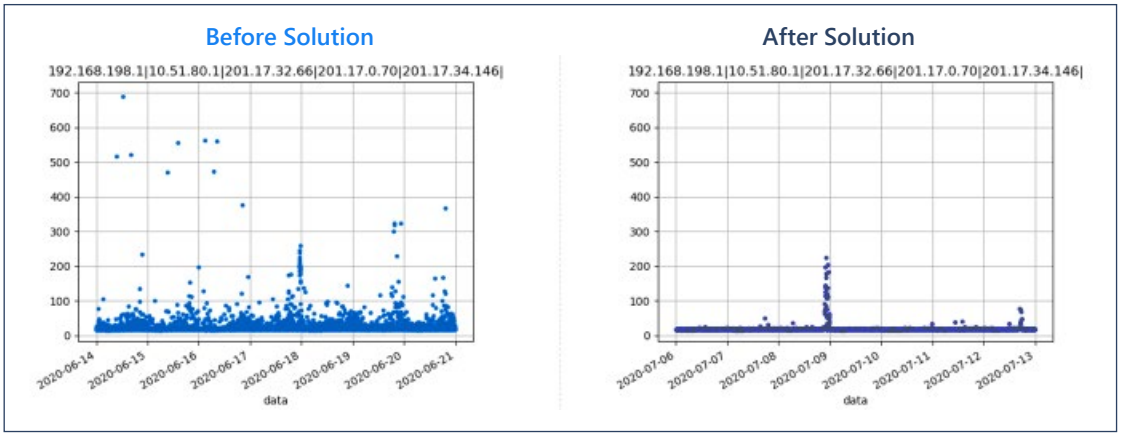


Figure 2 – Latency before and after congestion detection and solution for a single route

4.4 Health Check DNS and CDN

Another use case concerns the **ability to analyze Domain Name Server (DNS) and Content Delivery Network (CDN) performance,** called DNS and CDN health care tools. The ISPs usually track overall DNS and CDN performance, but current tools do not show individual performance from a end-user viewpoint. It is also helpful if the tool can identify the CDN that each user is accessing. For example, two neighbors may experience very different QoS for the same video application. A CDN check can show that one neighbor is accessing the local CDN, while the other one is accessing an overseas CDN because the application assigned an incorrect CDN address based on its IP address. This type of problem is extremely hard to detect since its root cause belongs to the video application and not the ISP, but the user blames poor performance on the ISP.

These tools measure the availability and latency of each component for every single user. It consolidates the results by geography to display overall network performance, and generates alarms when the CDN or DNS present degradation problems. Figure 3 shows the average response time (in milliseconds) by each type of DNS domain and by the node. When a CDN or DNS presents a problem, the ISP is informed immediately and can alert the affected users, which can prevent customer care calls.

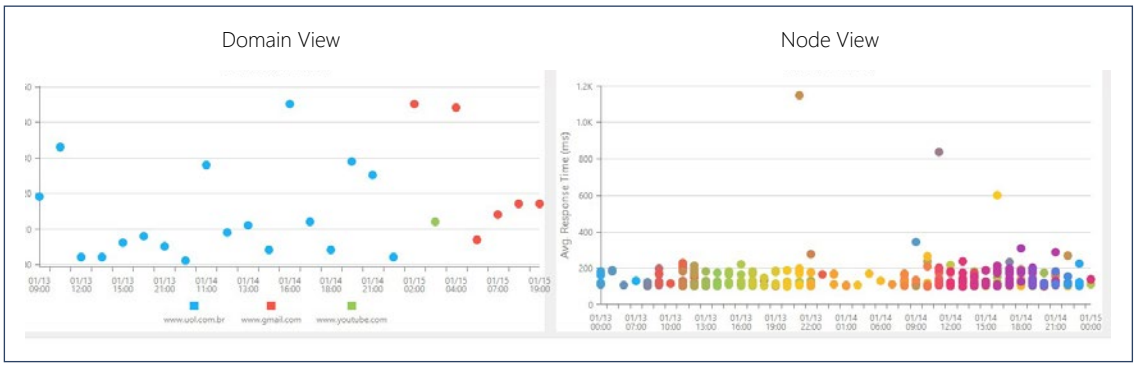


Figure 3 – DNS response time

4.5 Analyzing User In-Home Independent Network

The last use case concerns a WiFi module that checks the quality of the channel, channel interference, device fingerprint, and all the standard WiFi operations. One of the hardest problems to solve occurs when the user has their own WiFi router, degrading the QoE because of frequent disconnection from the modem. The user typically resolves the issue by turning off and on both the gateway and router, believing the broadband connection was the root cause of the frequent outages. It is difficult for the end user, however, to distinguish if the problem belongs to the ISP or to their own router. A set of tools that maps the in-home network, even when the router belongs to the user, is valuable in this scenario.

5 THE BEEGOL SOLUTION AND RDK

Broadband service continues to be the growth engine for global operators, and customers have their choice of service providers. This means that operators must have a highly reliable network to keep and attract new customers. Network superiority requires real-time, end-to-end network visibility, the ability to quickly troubleshoot issues based on real-time data from any aspect of the network, and the ability to use modern machine learning tools to automate solving issues before they impact end users.

Beegol offers a telemetry-based machine learning platform that can probe all broadband service components in the end-to-end network and identify the root cause of problems. Since the agent is integrated into the gateway, it can measure real user QoS, it can reach every network element, and it can simultaneously look at all network layers from the physical layer to the application layer. The telemetry aspect was designed to be programmable, so that data collection frequency is flexible, data is captured in real-time, and can even be event driven. Beegol's agent can also execute commands for self-healing tasks. There is an interaction between the machine learning module and the agent that requests additional data if required to triage a problem and provide an accurate diagnostic. Any command available in the modem can also be executed by the Beegol platform to support the self-healing and diagnostic capabilities. For example, Beegol can combine ARP tables with WiFi logs and WiFi data to identify a WiFi disconnection problem.

The Beegol solution offers operators a new set of tools to help manage existing DOCSIS/GPON/DSL networks, as well as future networks. This proprietary solution addresses the use cases previously noted, and can provide operators with the following key attributes:

1. Real time end-to-end network visibility, including a simultaneous view of end-user WiFi devices, WiFi router/extenders (ISP or user), gateways (DOCSIS, GPON, DSL), access/transport network, and additional key components such as DNS, CDN, DHCP, and Game Servers.
2. Data collection from any network component at any time.
3. Operator owns and manages data.
4. Customizable solution based on operator needs and capabilities, including UI dashboards for tech ops personnel and automated seal healing tasks.
5. Pre-integrated with RDK.
6. The self-healing and e-care module executes simple commands to solve some of the problems automatically, generates or blocks unnecessary trouble-tickets, and sends e-care messages to the end-user through the ISP's APIs. These real-time machine learning automation capabilities can quickly solve issues, increase network reliability, and reduce operating costs.
7. Increased network reliability allows operators to provide best-in-class broadband service, leading to greater customer satisfaction, brand loyalty, willingness to pay for service, and lower customer churn.

Figure 4, below, shows an overview of the Beegol platform.

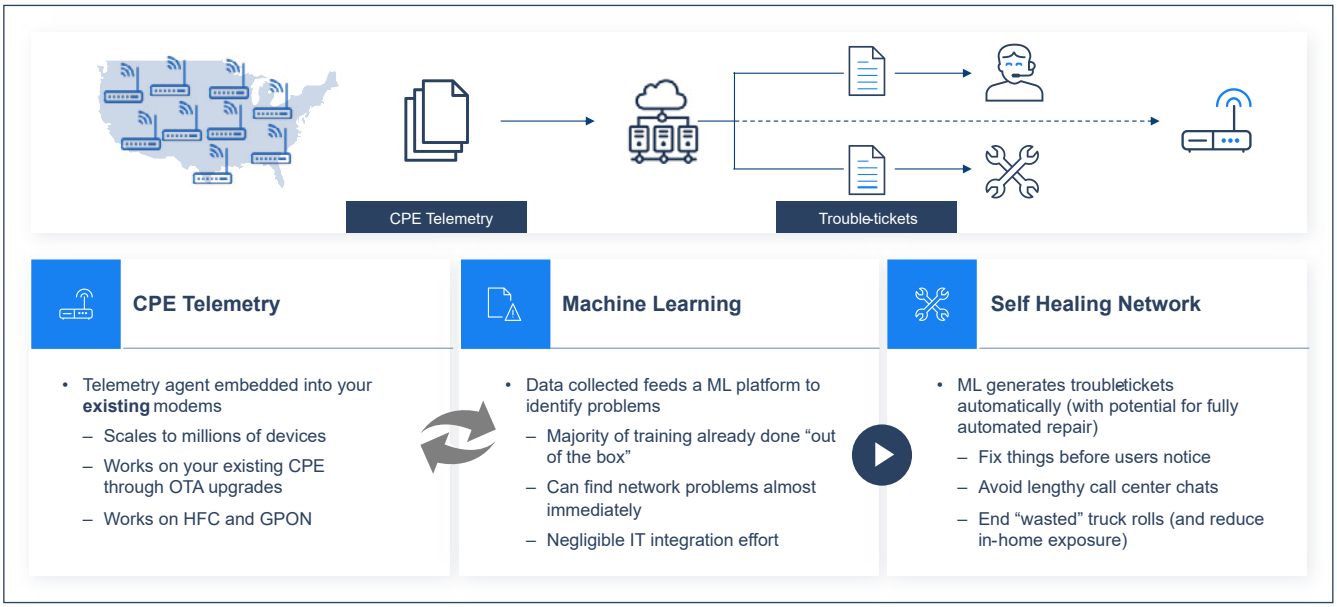
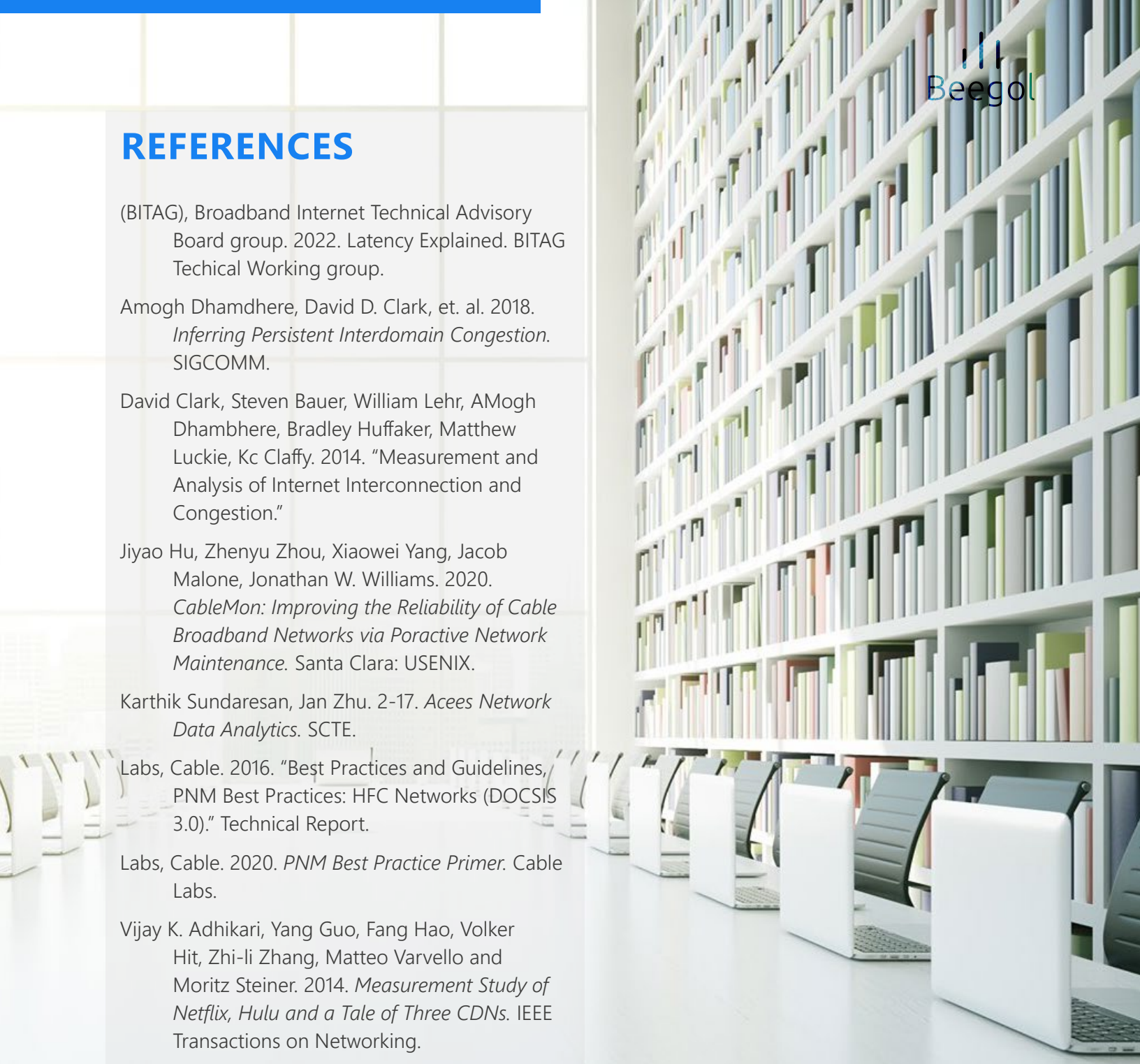


Figure 3 – Overview of Beegol's platform

Beegol's platform is based on a telemetry-command agent which is fully integrated into RDK. RDK is an open-source software platform for the connected home currently deployed in more than 80 million devices worldwide that standardizes core functions used in broadband devices, set-top boxes, and IoT. RDK Broadband supports DOCSIS, GPON, and DSL platforms. Beegol's integration into RDK provides a plug-and-play platform that can be integrated with Call Center and operation's trouble tickets systems and a customizable dashboard, providing key advantages for the ISP:

- Plug and play for every RDK enabled gateway.
- The ISP retains control of all data and the collection processes.
- Hardware agnostic, i.e., it can run seamlessly on any RDK device.
- Technology agnostic, i.e., it can run on DOCSIS, GPON, vDSL and future platforms
- Ability to send new data collection scripts to the gateway without a firmware upgrade.
- Enhances the telemetry systems currently available to the RDK ISP community.

In order to retain and attract customers, todays ISP's need to ensure superior QoE while also controlling operating expenses. This can be achieved by investing in a holistic platform that can look at all network layers and service components in parallel and in real-time, to identify and locate problems accurately and fast. The platform should be able to collect the data needed to diagnose a given problem, and to request more data in real-time if needed to increase accuracy. It should also be able to run self-healing commands in the gateway and run new data collection scripts added on the fly. A machine learning diagnostic platform that addresses all of these issues can drastically improve the Broadband experience and reduce operating costs.



REFERENCES

(BITAG), Broadband Internet Technical Advisory Board group. 2022. Latency Explained. BITAG Techical Working group.

Amogh Dhamdhere, David D. Clark, et. al. 2018. *Inferring Persistent Interdomain Congestion.* SIGCOMM.

David Clark, Steven Bauer, William Lehr, AMogh Dhambhere, Bradley Huffaker, Matthew Luckie, Kc Claffy. 2014. "Measurement and Analysis of Internet Interconnection and Congestion."

Jiyao Hu, Zhenyu Zhou, Xiaowei Yang, Jacob Malone, Jonathan W. Williams. 2020. *CableMon: Improving the Reliability of Cable Broadband Networks via Poractive Network Maintenance.* Santa Clara: USENIX.

Karthik Sundaresan, Jan Zhu. 2-17. *Acees Network Data Analytics.* SCTE.

Labs, Cable. 2016. "Best Practices and Guidelines, PNM Best Practices: HFC Networks (DOCSIS 3.0)." Technical Report.

Labs, Cable. 2020. *PNM Best Practice Primer.* Cable Labs.

Vijay K. Adhikari, Yang Guo, Fang Hao, Volker Hit, Zhi-li Zhang, Matteo Varvello and Moritz Steiner. 2014. *Measurement Study of Netflix, Hulu and a Tale of Three CDNs.* IEEE Transactions on Networking.

Weifan Jiang, Yunfan Zhang, Tao Luo, Ethan Katz-Bassett, Thomas Koch, Matt Calder. 2021. *Towards Identifying Networks with Internet Clients Using Publica Data.* IMC.

Beegol

São Paulo

Rua do Rocio, 199 cj 51
Vila Olímpia
São Paulo, SP
04552-000
Brazil

Los Angeles

10250 Constellation Blvd
Suite 100
Los Angeles, CA
90067
USA