SCTE CABLE-TEC EXPO 22
SEPTEMBER 19-22 • PHILADELPHIA, PA

2022 Fall Technical Forum
SCTE • CABLELABS • NCTA

A Roadmap for Cable Access Reliability

A Technical Paper prepared for SCTE by

Jason Rupe
Principal Architect
CableLabs
858 Coal Creek Cir, Louisville, CO 80027
303-661-3332
j.rupe@cablelabs.com

Ron Hranac
rhranac@aol.com

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

1. Introduction

As the speed of access networks increases and become less of a bottleneck to providing quality service, customers turn their concerns toward reliability. But they refer to service reliability, not network reliability. Still, network reliability is a key component of a reliable service. So, what is a cable operator to do?

As the cable industry turns more attention toward reliability, we have the opportunity to lead. The reliability engineering discipline is many decades old, and has a lot of tools, knowledge, and practices that we can start from, along with our own cable industry history of successful reliability engineering. Now, service usage is different, expectations are higher, networks are built and services are provided in new and different ways, and the technology we use today is rapidly evolving. The way we assure reliability has to be different too.

This paper provides a roadmap for addressing network and service reliability for the cable industry. Instead of a complete answer, it is a roadmap for the work ahead. There are many routes to take depending on where and how far the service provider wants or needs to go. CableLabs® and the Society of Cable Telecommunications Engineers (SCTE) can provide the van pool for part of the journey, and there will be vendors at the rest stops to help, but the journey is for the operators to take.[1]

2. Definitions

There are many sources for finding definitions of reliability and the many related terms. A few simple ones are offered here, and hopefully explain away some of the sources of confusion. Unfortunately, some of these terms have use in marketing, engineering, and non-technical contexts with different meanings. Even in an engineering use, there are often assumptions being made that make it difficult to know just what is being defined and under what context. As we apply a little focus on these definitions for our specific purpose, consider that these definitions also are the desirable properties of networks and services.

- *Reliability* as a word by itself is ambiguous, context dependent, and can mean a lot of different things depending on the situation. Consider first the perspective of the user of the word, and the context they use it under.
    - Customer – Whatever it is, it must work as I want it when I want it, without repair action on my part, so that the system is invisible to me when I use the service – this is service or use case reliability.
    - Provider – Sometimes a service provider uses this word to mean availability, suggest a lower repair rate, infer fewer customer calls, or other operational costs – this is operations reliability, or network reliability [1].
    - Academic – A more precise definition of reliability is the probability that something functions as intended up to time t>T given it works at time T=0. This is the reliability at time t. Note the reliability function is a decreasing function over time. Note also this says nothing about networks and services, which are repairable.

Availability is better suited for repairable items, though reliability is still relevant.

- *Availability* is the long-term percentage of time that a repairable system works. In other words, availability is the ratio of time that a service, device, or network is available for use over the total time, usually expressed as a percentage of the total time. Equally, it can be expressed as the

[1] Get it? A Roadmap for Cable Access Reliability (CAR). Did you expect a General Path Solution (GPS)?

probability that a repairable system works at some far future time. As such, availability considers the uptime, downtime, and repair time issues of the system or network as a whole. It can't tell you whether failures are frequent or infrequent, or repairs are lengthy or fast, but tells you the proportion of time that something can be counted on to work.

- *Maintainability* is the ease with which something can be maintained. Often this feature is determined by maintenance time estimates, sometimes through time and motion studies. Maintainability can include repair, but does include planned maintenance.
- *Repairability* refers to how easily and quickly a component or system can be repaired, or the property of being repairable. This term focuses on repair instead of planned maintenance, though the distinction is not always clear.
- *Survivability* is the ability of a system or network to operate under attack, and provide service in the presence of failures. Parts can fail, but the system or network still functions and provides service.
- *Resiliency* refers to failure recovery and fault tolerance, and the ability to provide service under degradation, over a broad range of demands. Degradation exists, but service functions.

Note that survivability and resiliency are related, but different in that the former refers to surviving a partial failure such as a lost link in a mesh network, while the latter refers to functioning under degradation such as ingress interference in a DOCSIS® network.

- *Performability* is the convolution of the performance function and the probability function of the system. It's a complicated concept, but let's think of it like this: 90% of the time, my bike works great; but 9.9% of the time, the tires are low and it is hard to pedal the bike; and the rest of the time, the bike is in the shop. If I consider that the bike with low tires performs at 50% while the fully functional bike is performing at 100%, in this simple example, the performability overall is $(0.9 * 1) + (0.099 * 0.5) + (0.001 * 0.0) = 0.9495$, which is less than 95%, even though availability is 99.9%. Think of performability as a state probability weighted performance measurement. Then realize that the probability function of the possible states, translated to the probability function of the possible performance levels, is a better measure of the experience than simple availability. If you replace the performance states with a continuous performance function, the concept still works though the math gets more complicated. But keep in mind that a single number representing performability is not as useful as the full function representing probability of performance.

Referring to the customer's definition of reliability, see that all these factors contribute to a user's perception of service or use case reliability. The unreliability of a service can be impacted by a number of performance measures as they relate to the usage or use cases associated with the service. Users are all unique, but they reveal their preferences through their product choices and willingness to pay for features; this information translates well to their perception of service friction[2] and thus reliability.

3. Goals for cable network and service reliability

In support of our industry's 10G Platform goals, operators have a lot of well-informed tasks to accomplish. Categorically, some of these tasks include:

[2] The concept of service friction in this context is new; we use it here to represent any impedance a customer experiences from using a service as intended in the desired manner.

- Assure service reliability primarily, which requires network and system reliability, availability, maintainability, and appropriate resiliency and survivability. But it also requires reliable processes, procedures, management, and more.
- Build a foundation of understanding, linking customer experience to system and network events, so operations, design, and upgrades all provide the best service possible.
- Design reliable network and service solutions, with degrees of freedom to manage service reliability, that are also reliable in executability, obtainability, etc.
- Select reliable, repairable solutions and components for given deployments.
- Create and maintain fault management that is reliable, inexpensive, and maintainable. That includes proactive network maintenance (PNM), which identifies and can be used to fix faults before customers are impacted.
- Develop operations tools that are inexpensive, reliable, understandable, and useful for proactive and reactive maintenance.
- Build intelligence to enable micro-financial decisions for preventive maintenance, technology replacement, resiliency, operations planning, etc.

3.1. Measure to Manage

Service and network reliability require well defined measures of performance that can enable management for effective results. This requires well understood service performance measures, network performance measures, and operations performance measures. When a customer experiences any service friction, that should be reflected in a key performance indicator. Aligning the measures to the customer experience is most important. It is not acceptable to answer a customer complaint with an "everything looks fine" because that suggests either you are blind to an important aspect of service, your measurements are insufficient, or your customer is wrong. The latter option is not a helpful assumption. The other two tell you improvement in your operations is needed. High levels of "no trouble found" point to the need for improvement as much as repeat trouble tickets do.

Doing all of this well requires knowledge of the failure modes, effects on networks and service, and criticality of the failure modes. A useful tool for capturing and referencing this knowledge is a failure modes, effects and criticality analysis (FMECA). Considering fault management in networks and complex systems, faults should be included with failures.

To obtain and maintain this knowledge, effective collection of components and failure modes is required. This enables analysts to determine useful corporate knowledge including what manufacturer or lot of components are not performing to specifications, what parts are wearing out, which failure modes must be addressed quickly to defend service, etc.

The most convenient example with direct application to our cable industry happens to be in this year's Cable-Tec Expo. See [2] for this year's Fall Technical Forum paper on applying FMECA to cable faults. Also see the Appendix of this paper for an example with explanation. This work is based on expert knowledge and is generalized for hybrid fiber/coax (HFC) networks.

But operator specific knowledge is necessary to support reliable services, so that problems specific to certain plant designs, aging or degradation, or even poorly performing components (hardware or software) can be found and addressed.

To use our cable industry's strength of sharing knowledge and energy toward common goals, we could develop standard methods for coding repair tickets to capture failure mode and component details so

operators can fully benefit from this knowledge, and apply it to assure service. But the implementation and use of the result will still be operator specific.

The industry could also benefit by standardizing how service and network reliability are measured. Fortunately, we're well on our way in an effort through the CableLabs PNM Working Group, which is sharing the output with several SCTE Working Groups, too.

But the work is just beginning, and the industry can benefit much by continuing the effort further. We should work to specify standard ways for measuring service and network reliability including

- the measurement definitions,
- how they relate to service and network reliability,
- how to track statistics and interpret them, and
- how to set control limits, perhaps setting specification limits, too.

See the Appendix for a starting framework that could serve as the foundation. But it is only a start. As you will see in the rest of this paper, we need equivalent, supportive measurements from all aspects of network operations to fully support service reliability.

3.2. Setting Service Level Agreements

Based on existing service performance information, service level agreements (SLAs) for high end customers can be set with confidence, and even rebates can be offered at net profit. When new technology is involved, models of the resulting performance may be needed, and appropriate SLAs should be set based on the network providing the service. Fortunately, simple mathematical models are often sufficient for setting and designing services for SLAs.

SLAs should be based on customer use cases but translated to service and network measures of performance. Define the service missions and translate the measurements defined to the customer use cases. For example, consider the use case of watching a movie through video streaming, including pausing a few times, requiring several functions to work when needed for the duration; what is the resulting experience, and how does it vary by customer or network condition or resource utilization?

The SLAs must be set rationally, so that they are achievable, and demonstrable. Achievability can be validated through a model, and the model fed with field data when available. Demonstrability can be achieved through data collection and translation to the customer experience. The translation again can be achieved through a use case model. For example, the movie use case just mentioned requires high availability from the network and supporting systems, and reliable performance of the network and functions for the duration of the use. If the network availability is 99.99% (equally 0.9999), and the probability of successfully delivering the movie and needed functions for the two-hour duration is 0.99999, then the overall probability of success for that mission is approximately $0.9999 * 0.99999 = 0.99989$. If a user has this use case once a week, then the probability of not experiencing a failed attempt to watch a movie in a year is approximately $0.99989^{52} = 0.9943$; there is a good chance (0.0057) that quite a few customers (more than two in a thousand) will not be able to watch a movie at least once a year, even with these seemingly high reliability and availability targets!

3.3. Service Assurance

Low friction, high reliability service is delivered through reliable networks and systems supported by reliable, efficient operations. From a network perspective, reactive, proactive, and predictive maintenance all play a role, along with fault management.

- *Reactive*: Fast restoration first then repair, prioritized by severity of impact.
- *Proactive*: Timely repair, cost efficient, prioritized by severity and opportunity, afforded due to resiliency, with no restoration needed. Also, proactive maintenance can be thought of as fault management, as it is a mechanism to manage faults before they become failures that must be reactively addressed.
- *Predictive*: Planned maintenance to address degradation before service is impacted in any way. Predictive maintenance occurs before a fault impacts network or service performance, so it can happen ahead of proactive maintenance. For example, detecting a trend in early degradation of a particular component type can lead an operator to predictively replace those components based on useful life prediction. Prognostics and Health Management is an emerging field of research which addresses this need. But predictive maintenance can also follow from proactive maintenance, such as when additional damage is observed in the proactive repair, leading to further maintenance planning. Well planned maintenance can minimize operations costs.

Standard methods for coding repair tickets to capture failure mode and component details for service and network assurance, as mentioned previously, would help operators gain full benefit from that knowledge for superior service assurance.

Note that reliable operations can play a most important role when customer facing, because operations usually faces the customer in response to service friction. The first touch point for a customer when they experience friction is usually the call center; today that is supplemented with a software application. Behind these touch points resides all the network operations tools and back-office systems, all of which are a part of the service provided, and must reliably reduce friction for that customer. A poor experience is a failure in service, so must be addressed through rapid reactive repair. Likewise, service can be proactively and predictively repaired, too, through early detection of risk (security, privacy, fault, and failure), and continuous improvement of systems and processes.

See Figure 1 for a depiction of the various types of repair cycles which complement and assure effective operations. Note that predictive management of services includes planning and engineering, including information technology (IT), functions that engineer reliability into the solutions that deliver service, as well as predictive maintenance to replace failed systems well before they have a chance to degrade other parts or impact network functions, far ahead of impacting service. But if you wait or don't detect the problems that become faults and failures, then you can still stay ahead of service impact through proactive management, which includes fault management and PNM, plus other forms of proactivity. But if you wait further, service is impacted because the faults and failures are felt by the customer through their service experience. Reactive management requires fast, and often expensive, restoration and repair; but sometimes the repair is not as fast as everyone wants because other resources, processes and systems are reactively taxed. Spare parts supply chains may extend the restoration and repair time, as might technician availability. When service is impacted, severity should determine the restoration priorities, and repair to follow that. Note that, with proactive and predictive maintenance, restoration is not necessary. Reactive repair requires more work, higher stress, higher cost, and results in less customer happiness.

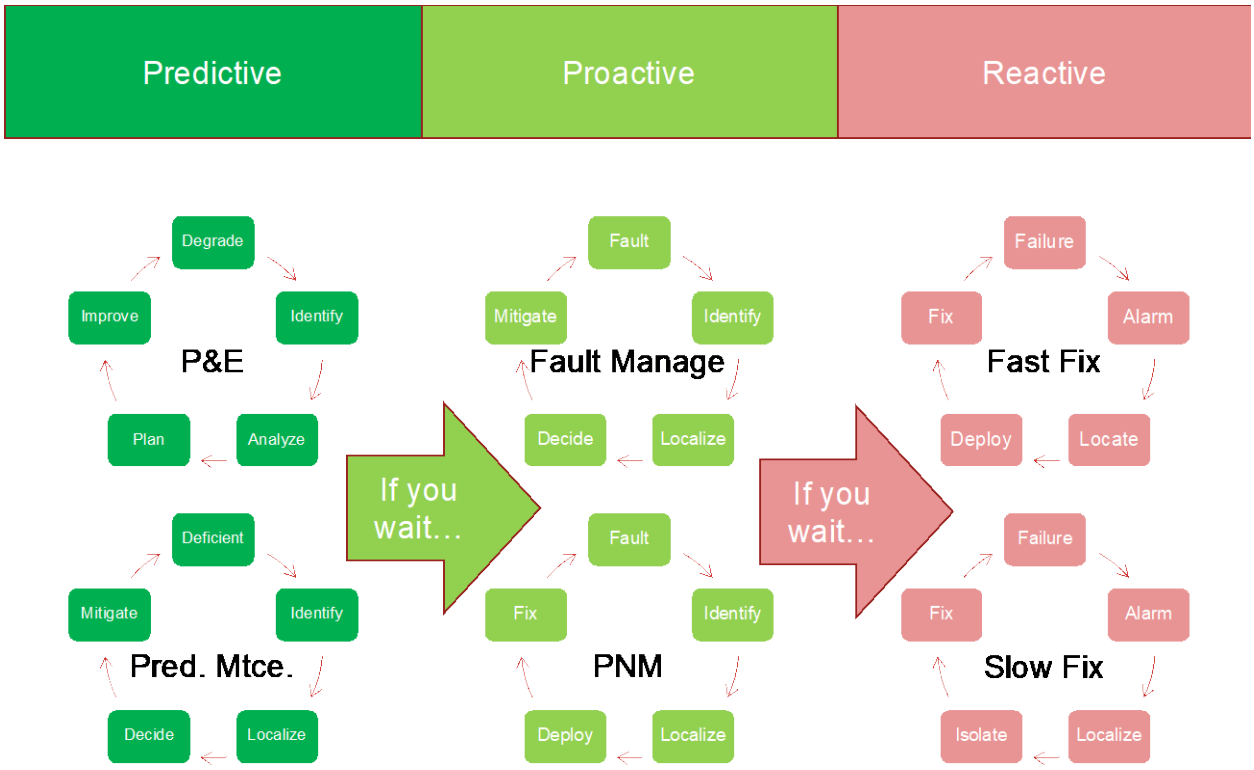

Figure 1 – Various types of repair cycles coexist in operations.

3.4. Fault Management

Knowing how a network can fail, the faults that lead to failures, and how those faults and failures are revealed in network performance can let an operator automate fault management and operate using PNM. Identifying the faults that impact service and where they come from is an important first step. How they relate to failures is important, too. An important goal in fault management is to automate as much of the fault identification, localization, and isolation as possible. And to do so reliably, which includes low false positive and very low false negative occurrences.

Based on event frequency, effect on service, and ability to test or monitor, set the policy based on established goals. Faults that are automatically mitigated can be ignored by repair technicians, but may need to be monitored by systems if they are indicators of other problems. Faults that require intervention can be handled with the appropriate repair cycle. Efficient fault management, like repair, requires an effective way to translate telemetry into action, such as the ProOps framework available from CableLabs [3], [4], [5] which provides a framework to observe (collect telemetry and information), orient (add context, assess the information, and potentially collect more information to assess), decide (translate information into faults and failures, then identify and localize faults and failures), and act (take appropriate action based on the assessment and information, with consideration of resources, priorities, schedules, etc.).

3.5. Repair and Supply Chain Optimization

With a strong handle on the priorities and planning for maintenance, operators can optimize their repair operations in many ways. Planning repairs on a longer schedule allows optimization of the travel time and distance required with maintenance, avoids unproductive technician time, and minimizes outage impact on service.

In addition, spare parts can be optimized to reduce held inventory and assure spare parts are on hand when and where needed to never adversely impact service, and never require expensive expedited shipping of parts. Critical parts necessary to correct critical failures should be readily available. Well-designed spare parts inventories can be created with lowest cost and appropriate spare parts availability. There are many applicable mathematical models available that can help operators set optimal inventory levels and policies for given targets of delivery time and probability of shortage.

3.6. Removing Degraded or Poor Quality Components

Technicians who deal with the plant all day know that some components wear out sooner than others, and some have specific faults in their design or manufacture that results in failure modes that emerge earlier or uniquely to these components. Sometimes environment has a strong influence on the early emergence of these failure modes. Temperature cycling, humidity, exposure to water, and even dry climates can impact network components differently. But even in controlled environments, poorly designed, selected, or built components can exhibit early failures which need to be addressed predictively. Early warnings from a few components can foretell the emergence of failures in the rest. As a result, tracking failure events by component type, manufacturer, age, location, and other factors can allow the operator to predict early issues and address them with predictive maintenance programs, instead of waiting for one-by-one replacement at failure. It is far cheaper to replace soon-to-fail components while doing other maintenance, to save on truck rolls and unproductive time. If such a program is required earlier than expected, a vendor management issue may need to follow, including perhaps warranty assisted replacement.

3.7. Vendor and Contract Management

Once an operator sets their goals for service, and can articulate how the network and its components translate to meeting those goals, they can align their contracts toward the goals, and even manage vendors to meet their contribution to the goals.

Component and system testing assures functionality, which reduces friction in the user experience. Testing for design and features is well established in our industry. Testing for basic features and functionality is a necessary foundation. Testing for capabilities necessary to provide specific services and features is important, too. Because long duration testing of hardware-software integrated systems is not feasible in most cases, it is important to test software well, life test hardware, and design-in system health monitoring and management capabilities for what can't be assured otherwise. Measuring early and useful life performance of components and system parts allows prediction of problems and validation of vendor performance.

3.8. Network Design

Networks should be built with performance goals in mind, and that performance should include reliability concerns as well. Doing this requires modeling of network behavior, including protection and restoration and resiliency mechanisms, for hardware, software, systems, and even people.

Network architecture will dictate allocations of service and network reliability to provide a given level of service, based on the measurements specified. How much friction-degradation and/or downtime can be allowed at, say, an optical backbone link as compared to the cable modem termination system (CMTS), cable modem (CM), or access network? Operators who are targeting service and network reliability will be collecting information and modeling to assure good decisions get made at the point of system and network design. This purposeful design enables management of network sections and knowing where to focus resources for network and service health.

Network operational cost-benefit modeling is an important component of this ability. Start with a framework for modeling the needed tradeoffs and making decisions around improvements.

This work can apply to operations design. Should a technician be sent to fix a proactive problem today, or should we wait a week in case there are more issues that can be solved with the same truck roll? If we are not sure whether a particular fault is caused by a failure mode in the home or in the yard, which technician type should be sent to keep costs lowest, and have the best chance of fixing the problem the first time?

This work can apply to decisions about the customer, too. For example, it may be worth modeling the impact of an uninterruptable power supply (UPS) in the home, or conduct a cost-benefit analysis of providing long term evolution (LTE) backup in the gateway.

But most obviously, architecture choice applies to the network decisions too. Should the operator consider media access control (MAC) manager redundancy architectures, or is a single hardware solution good enough if software and/or state are maintained redundantly? Should nodes be daisy chained in a particular deployment scenario, or is an optical ring truly necessary for the level of service we need to provide?

All these decisions need to be made with data and analysis considered, not just a gut feel, or a first-cost-driven approach. For some examples which come from our own world and are simple to use, see [6], [7].

3.9. Technology Tradeoffs and Lifecycle Management

Operators and vendors both need to benchmark the performance and reliability of existing deployed technology. This allows us all to set goals for future technology based on needed improvements or stability of reliability, availability, maintainability, survivability, and performance. Operators can model the comparison in deployed areas against the goals set by the company, and then enforce the component performance to assure goals are met as the new technology is deployed. Some high-level steps to follow:

- Benchmark existing technology
- Set goals for architectures as deployed
- Set goals and requirements for components
- Deploy and measure performance

Network components wear out. Replacing versus repairing is a decision that should consider costs, useful life, and impact to service.

At some point, an entire system or network may need to be replaced, because it has been used to the end of its useful life. This limit happens when the network or system can no longer meets its intended function in a reasonable way, or the requirements of the system or network have shifted so it can no longer meet

the current set of necessary use cases.[3] See [8] for an appropriate model and treatment of the problem. Planning for wear out is important for budgeting, operations planning, supply chain management, and more.

4. History is Our Foundation

The cable access network community has given attention to reliability for decades, with considerable success. Aside from the various papers mentioned in the previous section, there are several other noteworthy works worth mention, study, and utilization.

In the late 1980s the cable industry began upgrading its networks from all-coax tree-and-branch to what is today known as HFC. Around the same time, the industry became interested in network reliability. Operators, equipment vendors, and others worked together to determine just how reliable cable networks really were and what it would take to improve their reliability. Of particular interest was whether cable networks could meet the old Bellcore "four nines" availability spec. More on that in a moment.

The topic of network reliability and availability is introduced in the context of cable networks in [10]. In chapter 20 of that book, the topics of benchmarking, definitions, calculations, redundancy, and network analysis are all discussed.

In 1992, CableLabs and several cable operators organized an Outage Reduction Task Force to "address the issues that stem from cable system outages." The task force studied and reported on key topics relating to reliability in the cable industry [11]. CableLabs published "Outage Reduction" as a summary of the task force's work, with chapters covering seven major topics in a large three-ring binder:

- Customer expectations, detection, and tracking
- Reliability modeling of cable TV systems
- Plant powering in cable TV systems
- Outside plant and headend protection
- Service restoration
- Cable TV system power supplies
- Power grid interconnection optimization

"Outage Reduction" was accompanied by a computer diskette with a Lotus 1-2-3 based reliability model. In addition to the published document and reliability model, CableLabs conducted half-day training workshops for member companies on the subject matter in the document's first four chapters.[4] Among the many recommendations in "Outage Reduction" was a critical threshold of no more than two outages in a three-month period (0.6 outages per month per subscriber) be a target for operators to achieve and maintain.

While the aforementioned guidance was considered suitable at the time for an entertainment model, any movement to telephony and data services required a higher performance threshold – hence the interest in the Bellcore four nines (99.99%) Standard Application Grade availability spec [12]. That parameter translates to no more than 53 minutes of outage time per year. Studies and analyses in the 1990s

[3] Arguably, DOCSIS technology was born out of the need to meet the new set of use cases that the current network technology could not; but the network could be augmented to allow it to meet the new use cases, reusing coax.
[4] The first four chapters of "Outage Reduction" were also published in the December 1992 through March 1993 issues of *Communications Technology* magazine.

confirmed that cable networks could meet four nines, assuming certain network architecture design criteria, device and component cascade limits, backup power and redundancy, and so forth.

While much has changed since the cable industry's earlier work in reliability and availability, some of the methods and knowledge collected form a useful foundation for today. Now that we are in a DOCSIS access network world, some of that work should be revisited.

Alberto Campos [13] in 2011 presented a paper that laid another foundation for evaluating the quality of experience (QoE). He tied performance metrics that impact QoE to the events that operators experience in the network, and the reliability of several of these features. He identified a large number of factors that contribute to the customer experience, and highlighted the importance of key elements by proposing a service availability metric. This proposed approach gathered in one place the many factors that influence service reliability and quality, plus it provided a convenient way to pull it all together into a single quantity for management. With some updating, a useful standard or operational practice could be created; with additional tailoring, operators can have a strong foundation of measurements to manage with.

Thankfully, SCTE has a new working group on Network and Service Reliability which should be the right place to tackle the new challenges, building on the foundations noted in this paper, and the papers and resources referenced by these works.

5. Conclusion

If you are an operator, you probably have been thinking while reading this paper that you already are doing all these things. You may have even participated in some of the noted foundational work. But there are at least two questions each of us should ask:

- Are we designing and executing these activities toward improved service and reliable networks and services? and
- Are we maintaining our reliability management and knowledge with changes to service, customer demand, technology changes, competition, and factors outside our control?

Operations survive by being cost focused. But that focus should be a long-term focus. And when it is, designing your operations and services toward appropriate reliability goals is your friend, and serves as the lenses for you to keep your eye on that long-term focus of managing cost as well as revenue and the drivers of both.

Once you can answer the two previous questions, you are ready to join us at SCTE's Network Operations Subcommittee Working Group 8 (NOS WG8): the Network and Service Reliability working group. See you there!

6. Acknowledgements

The authors wish to thank the numerous people who have contributed to the development of this work, especially including the hard working operator and vendor members participating in the FMECA work.

7. Appendix

7.1. Failure Modes, Effects, and Criticality

FMECA is a proven methodology for analyzing a system, process, or network for ways it can fail, determining the effects of failure, and assessing the criticality of each failure modes. The applications of this method are broad, but generally allow for appropriate design of technology to meet the requirements. An existing deployed solution is often a source of information when conducting an FMECA, either for augmenting the existing solution with improved operations, telemetry, fault management, etc.; or for designing the next generation solution for optimal performance.

A sub-team from the PNM Working Group at CableLabs has been working for many months on an FMECA that focuses on physical layer failures from the headend out to the customer, the access network. A sample of that is provided in Figure 2.

Figure 2 – A sample of the draft FMECA currently being built.

In the figure, see hardline, connector, and part of the adapter failure modes; these components are part of the outside plant subsystem of the cable access system. Component and subsystem effects are described under the sub-effect heading, where we include several degradation causes and detectable impairment types. Under the heading of network effect, we indicate the effect each failure mode can have on the network from accelerating degradation, through signal impedance and capacity loss, to network separation. The service impact is indicated under service effect, and depicted in greater detail in Figure 3.

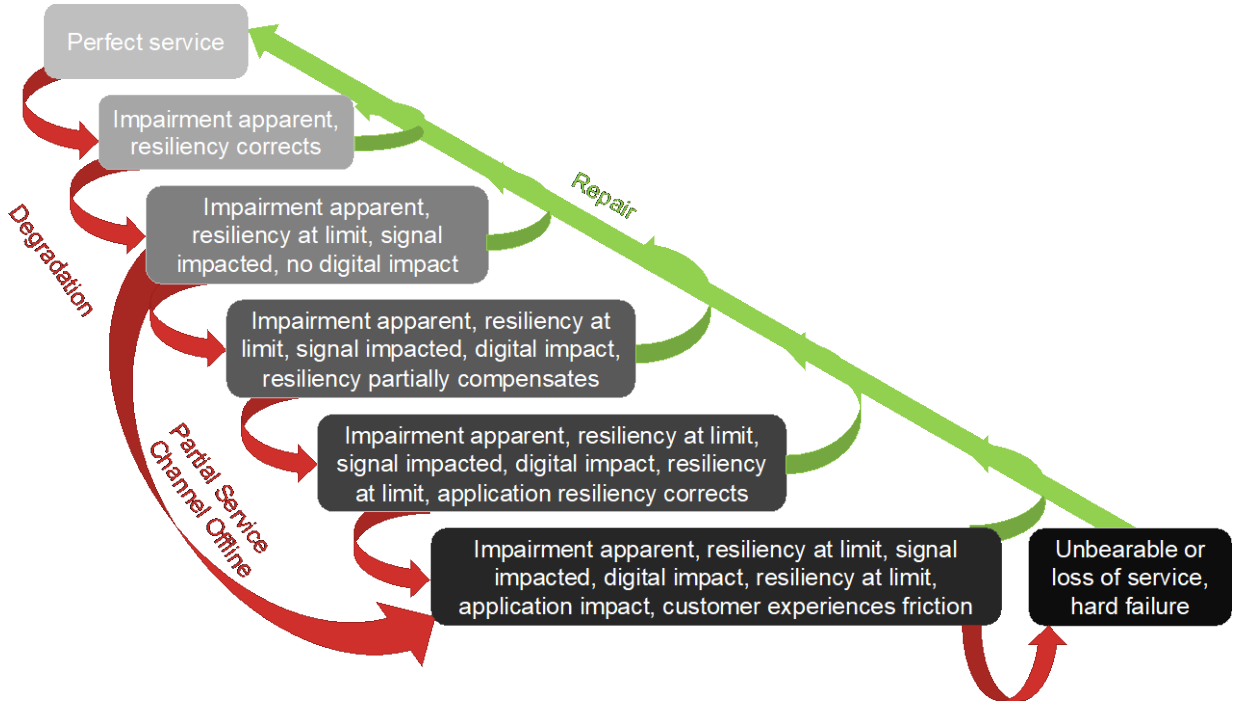

Figure 3 – A depiction of how network performance states relate, from perfect function to untolerable or failed service.

The FMECA here is focused on the physical network as it supports the mission of customer service. But some failure modes can be detected early, and some may accelerate degradation which eventually impacts service.

Because some of the failure modes can have no immediate effect on service, the FMECA documents some effects that impact the network and its components as well. PNM has identified and cataloged several signal impairments that when not too severe do not impact service, but can impact network RF bandwidth or at least foretell of future service issues.

As this work continues, we should be able to show how the repair actions relate to the failure modes, and thereby find new opportunities for improving fault management (identification, localization, and removal) in the access network.

7.2. Service Reliability

7.2.1. Reliability of a service is availability

Reliability is the probability that a system or component is working at a future time when needed. Replace component or system with product, and the intent is close but not accurate.

A service can experience downtime or degradation, but is repairable. Reliability is a non-increasing function which does not describe a repairable thing like a service. What we really want to consider then is availability, which is the probability of a service being in a functional state at some future time. And usually that future time is not defined, so we usually mean a long-term steady state of the system.

Put another way, managing service reliability requires an availability measure of the service.

7.2.2. Reliability of a service is really performance

But a service is a complicated mix of use cases and capabilities, any one or more of which could be available at different times. The service is not a single thing necessarily, and depending on how a customer wants to use it, may or may not work as intended. And if not as intended, is there value in it working in an alternate way or a degraded way? Think in terms of an email that does not go through right away, or a streaming video that takes a few seconds to buffer; does the customer notice or care, or not?

To address this issue, the concept of performability was developed many decades ago. It is a functional convolution of the performance probability distribution function with the value achieved at each performance level possible. While complicated, it does deliver a single measure of performance.

Put another way, service availability actually needs to be a performance measure of the service that includes each possible degraded performance level including complete failure, and the utility that a customer gets from the given performance level. Network reliability and performance both contribute to service availability, but they do not represent it.

7.2.3. Customers are a mystery

Note, however, that each customer is different, and the impact to their perception of performance levels varies by their situation, tolerance, emotions, and the value they put on aspects of the service. A CEO trying to close an important deal might value video conferencing much more than a student doing homework. Further, tolerance for a degraded condition might depend on the person's tolerance to previous outages, expectations of the overall quality of service, and other factors.

Put another way, the impact of service performance on individuals is highly variable and complex. The utility they get, and their overall tolerance of the experience being poor, are not easy to quantify. All an operator can do is provide the best level of service they can for the use cases known, at the price point customers are willing to pay for it.

7.2.4. Service performance

Because simple is an important goal when developing measurement systems, and recognizing that all services are a three-legged stool of cost, performance, and reliability, with cost being understood by the customer, our measure should be centered around performance and reliability, which as just described is really the aspects of performance that are delivered in each available state.

In other words, we can quantify the probability of the service delivering given levels of performance, which is what we can manage. We can seek to understand the customer, what they care about, what they are willing to pay for, and how they see competitive options. But first we must measure what we are providing in terms of service: the performance of that service as a probability space, not just an average, not an average and standard deviation, but as a probability function.

7.2.5. Measuring service performance

The task here is to identify the features of a service that describe the utility of the service to a customer. If latency is not important to, say, a webpage load, then latency measures are less important. But if the service is also being used for video conferences or video games, then latency matters, and a solid latency measurement is important to service reliability.

Examining the use cases and types of service we offer in our industry, a few basic performance measures are obvious.

1. *Goodput* – bits per second, throughput of the useful service-delivering bits on the network, assuming digital data delivery (not analog video, for example).

2. *Latency* – how long it takes to deliver each bit of data.

3. *Jitter* – packet delay variation, or how stable is the latency, and to a certain extent the goodput of the service.

4. *Packet Loss* – data that does not reach its destination.

5. *Availability State* – what is the state of performance of the service in terms of its capability?

Note that, at a packet or bit level, packet loss may be a considered measurement for either availability or as a factor for goodput or latency.

While necessary for measuring service reliability, these measures are not sufficiently described yet, and are not the end of the task for providing service.

Each performance measure statistic must be based on sufficiently detailed measurements to assure sufficient resolution of the differences in service performance levels, and measure all aspects of the performance measure. For example, measuring performance once a day at the same time every day is neither sufficient resolution nor unbiased. Measuring from the CM to the node is not an end-to-end measurement so does not represent the service experience. For understanding service reliability in sufficient depth, service providers have to design the measurement system thoughtfully, to meet their goals of continuous improvement and maintain a focus on service assurance.

But also knowing there is a problem is only the beginning; it takes more information to know the cause, locate it, and remove it from impacting service. That is the work of network operations, or network reliability, which is a key part of service assurance.

7.3. A proposed measurement framework for cable

Each service should have requirements in terms of required goodput, latency, jitter, packet loss for performance, and availability, if at all possible. Lacking a complete set of requirements, it is still incumbent on the provider to measure the service delivered. This section proposes a service availability measure based on telemetry that can be collected from the cable network.

Many of the measurements suggested here are a part of the proposed FCC 22.7, which was announced in late January of 2022. That proposal makes addressing this issue an urgent one, but also supports much of what is addressed in this document, the first draft of which formed in late 2021, with this version acknowledging what is known about the FCC proposal.

First, we need to consider several aspects of the service from a measurement feature point of view. Then we can treat each measurement in that framework.

7.3.1. Features

Several features of service reliability measurements were suggested earlier.

- Use of the measures – Depending on the various uses of the measure, the remaining features must be sufficient to address all needs.

- Bias in the measures – Because service usage varies by time of day, day of week, etc., any point measurement must be taken over a sufficiently large amount of time, with sufficient frequency, and of a sufficient sample size of the traffic to be measured.

- Resolution of the measures – The frequency of sampling must be sufficient to provide proper resolution. For example, an estimate of availability found by sampling daily will not provide good resolution for a highly available service for quite some time.

- Service level – The applications should provide the estimates when and where possible. But that is not always possible. So, service specific measures of performance at the end devices are close and sufficient for many uses. And because the operators do not manage the applications in many cases, but only the service classes as defined in DOCSIS, we should rely on these service classes first, and augment with application specific measurements when possible.

- Actual or surrogate – In some cases, we use special measurement packets to estimate actual service performance. But this method is known to be highly inaccurate and relies on a translation model that is not ideal. It is best to avoid this approach, and favor measurements on the actual traffic.

Each of these features need to be applied to each measurement. The measurements in the set are complimentary, so a complete set is needed.

7.3.2. Goodput

Throughput in terms of end user useful data is goodput. If a goodput measurement is not possible, then a throughput measurement by service type is a useful approximation because goodput can be estimated from this throughput by modeling for overhead.

When the data rate needed exceeds the capacity of the link, interface, or other component, packet queueing and congestion happen, unless discard is the only option. When the purchased data rate is not supported in the grants given by the CMTS to the CM, then applications experience latency. These understandings lead to secondary measurements for throughput.

In many cases, this measurement is used to guard against network congestion. In the access network, a simple network utilization may be sufficient. But when considering that there are customers who may be impacted by impairments and low signal-to-noise ratio (SNR), or high performing customers who also are high bandwidth users, individual CM-level throughput is important to estimate.

Recommendation: One measurement of network utilization, one measurement of bandwidth requests made and granted requests by CM, and one measurement of utilization by profile by CM. All separate for upstream and downstream.

7.3.3. Latency

DOCSIS has defined a latency measurement to support low latency DOCSIS (LLD). This measurement is taken at the CM supporting DOCSIS measurement, which is a subset of the actual experience, but a useful one nonetheless. Using this measurement for all service traffic is an excellent starting point.

Recommendation: Use the LLD latency measurement already defined for DOCSIS, and apply it to all service classes. Report by CM and service class. Augment with application-level sampling of packet delay when possible.

7.3.4. Jitter

Jitter, or packet delay variation, is the variation in the arrival of data. Some applications handle this factor through buffering, but not all applications can be made insensitive to jitter. A measurement of jitter for service types sensitive to it would be important to define. Jitter is strictly defined already, but there are alternatives that we could develop that would meet the needs for our industry.

The time between the arrival of packets would provide useful data for estimating a jitter-like measurement. A mechanism that provides the packet delay variation directly is useful if it is well defined, testable, and validate-able.

Recommendation: Use packet-level jitter measurements already defined in specifications. Augment with application-level sampling when possible.

7.3.5. Packet Loss

While packet loss is not permanent in reliable transmission protocols, thus would be reflected in terms of latency and jitter and goodput at the application layers, it is included here as it is a proposed measure in FCC 22.7.

Applications that rely on unreliable transmission protocols will not experience packet retransmission, so packet loss is an important problem and should be measured.

Applications that are latency-impacted may discard packets that are late, resulting in the same impact as packet loss. Therefore, some application consideration is important for packet loss measurement.

Forward error correction (FEC) statistics are included in DOCSIS and would be an important supportive measurement to include here, and for a DOCSIS reporting point of view would surely be more than sufficient as a measurement which can generate appropriate statistics.

However, we may need to report FEC statistics by service class or application to provide a useful measure of service reliability-availability.

Recommendation: Rely first on FEC statistics, particularly uncorrectable codeword errors. Each of these represents lost packets or data which require either application layer or protocol layer retransmission or re-requests. For reliable protocols, measure discarded packets and retransmissions. For unreliable protocols, measure lost packets. Augment with application-level sampling of packet loss when possible.

7.3.6. Availability

The overall availability of the network is an obvious, important component of service reliability. Network availability should consider cases where the user wants to use the service, but it is not available. Estimates can be obtained through polling logs from the CM, or polling state from the CMTS, or through ping-response approaches, or likely a combination.

Timeout statistics would be a useful contributor here, but there are known issues with timeouts being inaccurate as estimates of availability due to various contributing factors. However, it may serve as a

surrogate measure that could be translated into an availability estimate through a translation model, or as a contributor toward an estimate that incorporates logs and other traffic data.

Recommendation: Provide timeout statistics, augmented with logs from the CMTS and CM to estimate network availability. More detailed assessment is needed to develop the models here. Augment with application-level or device specific sampling when possible.

Overall recommendation: Measure or estimate the service experience; when insufficient, drill down toward the cause, and address the fault.

Abbreviations

CEO	chief executive officer
CM	cable modem
CMTS	cable modem termination system
DOCSIS	Data-Over-Cable Service Interface Specifications
FCC	Federal Communications Commission
FEC	forward error correction
FMECA	failure mode, effect, and criticality analysis
HFC	hybrid fiber/coax
IT	information technology
LLD	low latency DOCSIS
LTE	long term evolution
MAC	media access control
NOS WG8	[SCTE] Network Operations Subcommittee Working Group 8
PNM	proactive network maintenance
QoE	quality of experience
SCTE	Society of Cable Telecommunications Engineers
SLA	service level agreement
SNR	signal-to-noise ratio
UPS	uninterruptable power supply

Bibliography & References

[1] R. Hranac, “Service Availability,” Communications Technology, December 2007. Available at https://scte-cms-resource-storage.s3.amazonaws.com/07-12-01%20service%20availibilty.pdf

[2] M. Spaulding, L. Wolcott, J. Rupe, “Improving Operational Intelligence for Maintaining Cable Networks,” SCTE Expo 2022.

[3] J. Zhu, K. Sundaresan, J. Rupe, “Proactive Network Maintenance using Fast, Accurate Anomaly Localization and Classification on 1-D Data Series,” 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), 2020.

[4] J. Rupe, J. Zhu, “Kickstarting Proactive Network Maintenance with the Proactive Operations Platform and Example Application,” SCTE Expo 2019.

[5] J. Rupe, J. Zhu, "Comparison of RxMER Per Subcarier, Bit Loading, and Impairment Driven versus Meaurement Variability," SCTE Expo 2020.

[6] J. Rupe, "A General-Purpose Operations Cost Model to Support Proactive Network Maintenance," SCTE Expo 2019.

[7] N. Foroughi, J. Rupe, "Distributed Gain Architecture: Increased Performance, Decreased Power Draw," SCTE Expo 2020.

[8] J. Rupe, "Optimal Maintenance Modeling on Finite Time with Technology Replacement and Changing Repair Costs," Annual Reliability and Maintainability Symposium (RAMS), 2000.

[9] M. Spaulding, L. Wolcott, J. Rupe, "Improving Operational Intelligence for Maintaining Cable Networks," SCTE Expo 2022.

[10] W. Ciciora, J Farmer, D Large, M. Adams, "Modern Cable Television Technology: video, voice, and data communications," Elsevier, 1999, 2004.

[11] Outage Reduction Task Force, "Outage Reduction," CableLabs Technical Report 1992.

[12] Telcordia Technologies Generic Requirements, "Reliability and Quality Measurements for Telecommunications Systems (RQMS-Wireline)," GR-929-CORE, 2002.

[13] A. Campos, "Holistic Approach to Evaluating Quality of Experience," SCTE Cable-Tec Expo 2011.

