Designing Resilient DOCSIS Remote-PHY Solutions using High-Availability System Architectures

Daniel Lavender, Principal Architect, CableLabs
Karthik Sundaresan, Principal Architect, CableLabs

Abstract

High Availability (HA) has not been standardized as part of the CCAP and M-CMTS specification efforts and therefore implementation of these capabilities has been interpreted differently by each system implementer. This paper presents scenarios for applying High Availability (HA) systems concepts to the new DOCSIS Remote PHY architecture.

An overview of DOCSIS Remote PHY is presented to introduce the new architecture. Next, to set the context for high availability systems, models and tactics for generalized HA systems are presented. The document concludes by combining the concepts of Remote PHY and HA to suggest models and scenarios for creating High Availability (HA) systems supporting DOCSIS Remote PHY architectures.

DOCSIS REMOTE PHY (R-PHY) CONCEPTS

Traditional DOCSIS CMTS architectures are migrating to distributed architectures. The DOCSIS Remote PHY specifications describe an architecture that moves the PHY layer processing from the CCAP Core or CMTS Core to the edge of the network (e.g., node). This architecture model is generally referenced as Modular Headend Architecture version 2 (MHAv2). Figure 1 presents the high-level DOCSIS Remote PHY Architecture. The terms “Remote PHY” and “Distributed PHY” are synonyms and are used interchangeably throughout the documentation for MHAv2.

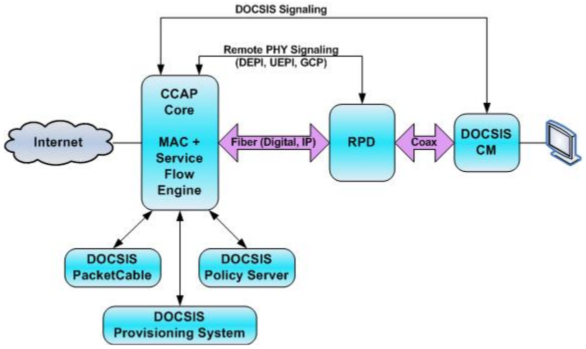


Figure 1 - DOCSIS Remote PHY Architecture
Source: R-PHY D03 specification

In the MHAv2 model, the PHY layer processing is moved from the CMTS/CCAP in the Headend to the node. MHAv2 allows a CMTS to support an IP-based digital HFC plant, meaning the Headend is connected to the Remote node via a digital fiber, a Layer 2 Ethernet link. In an IP-based digital HFC plant, the fiber portion utilizes a baseband network transmission technology such as Ethernet, EPON (Ethernet over Passive Optical Networks), GPON (Gigabit Passive Optical Network), or any layer 2 technology that supports a fiber-based layer 1. One of the common locations for a Remote PHY Device is the optical node device that is located at the junction of the fiber and coaxial plants.

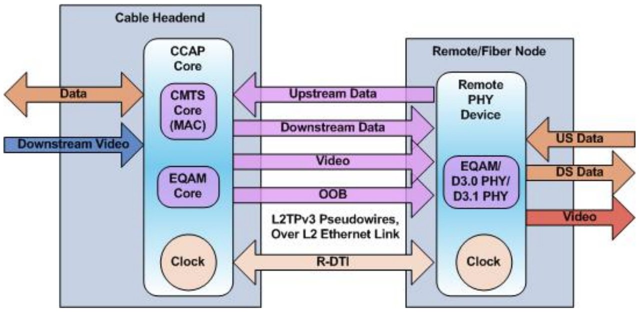


Figure 2 - R-PHY Architecture Components

The Remote PHY Device is connected to the CCAP MAC Layer using L2TPv3 (Layer 3) pseudowires, which tunnel the DOCSIS payload from the CCAP Core to the Remote PHY device. MHAv2 allows multiple CCAP Cores, Video Cores or OOB Cores to connect to a series of Remote PHY Devices.

The Remote PHY architecture uses a combination of pseudowires for Upstream External PHY Interface (UEPI), Downstream External PHY Interface (DEPI), Out of Band (OOB) data, and timing for signaling. A timing solution referred to as R-DTI is used to provide timing services for functions such as DOCSIS scheduling across the CCAP Core and Remote PHY Device (RPD).

R-DEPI, the Downstream External PHY Interface, is the downstream interface between the CCAP Core and the RPD. More specifically, it is an IP pseudowire between the MAC and PHY in an MHAv2 system that contains both a data path for DOCSIS frames, video packets, and OOB packets, as well as a control path for setting up, maintaining, and tearing down sessions.

R-UEPI, the Upstream External PHY Interface, is the upstream interface between the RPD and the CCAP Core. Like R-DEPI, it is an IP pseudowire between the PHY and MAC in an MHAv2 system that contains both a data path for DOCSIS frames, and a control path for setting up, maintaining, and tearing down sessions. The R-OOB specification outlines multiple approaches to passing OOB (out of band) signals for MPEG Video distribution, through a Remote PHY system.

Remote PHY architecture relies on centralized software in the core—because PHY has minimal complexity, it is well suited to be distributed to the edge and into the node.

Architectural Advantages

The architectural advantages are:

- Because Remote PHY is compatible with legacy HFC plants, it can be deployed incrementally into existing plant architectures.
- Moving the PHY layer to the edge brings full IP closer to the end user (the subscriber), thus reducing complexities. And transitioning to standard IP switching and routing architectures enables simple changes to be made dynamically to delivered services.
- This model enables increased capacity over the HFC network because of SNR gains in the digital optical L2 network. This is beneficial for DOCSIS 3.1 technology deployments and their higher order modulations, which increase the available bandwidth on the network.

Operational Advantages

The operational advantages are:

- Operators are facing a space and power crunch within their Headends; Remote PHY architecture reduces the power consumption and space requirements in the Headend by moving part of the CMTS to the node.
- The digital optics are easier to maintain and simplify plant maintenance for the operator.
- Remote PHY architecture makes it easier for the operator to improve services that matter to the customer.
- Capital expenditures can be spread out over multiple periods of the company's financial year.

The Remote PHY architecture keeps the simple elements (PHY) remote and the complex elements (MAC) centralized. This architecture scales to the needs of the operator. It allows an Integrated CMTS and Distributed-CMTS line cards to be present in the same CCAP chassis.

For an optical access network based on linear optics, the Remote PHY Device is located at the hub. For an optical access network based on digital optics, the RPD is located at the optical node.

The MHAv2 architecture permits RPDs to be managed by more than one CCAP Core. An RPD is controlled by exactly one "principal" CCAP Core and can communicate with zero or more "auxiliary" CCAP Cores, Video Cores, or OOB cores. An "auxiliary" core manages a subset of RPD resources, e.g., particular channels or RF ports. The principal and each auxiliary CCAP Cores establish their own GCP session and L2TPv3 control sessions with the RPD.

The term "CCAP Core" can refer to either the principal core or an auxiliary core. (Source: CM-SP-R-PHY-D03_150320, section 11.1). Redundant Remote-PHY Devices in the node with additional Remote-PHY failover capabilities in the CCAP Core is beneficial to MSO operations.

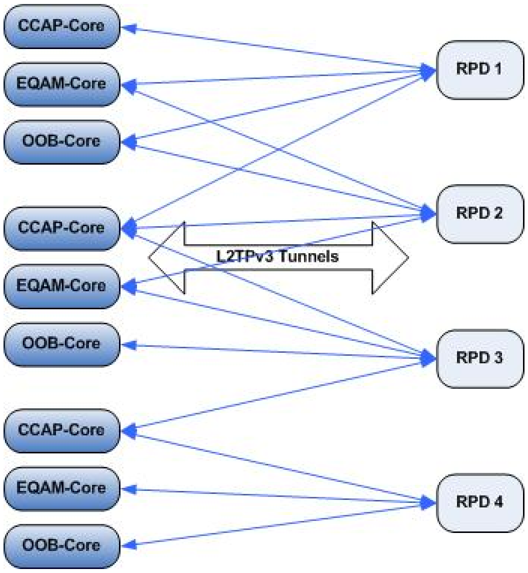


Figure 3 - R-PHY Architecture: Multiple RPDs Controlled by Multiple Cores

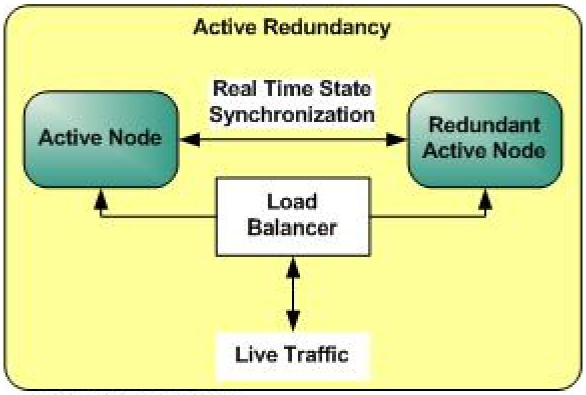
ARCHITECTURES FOR HIGH-AVAILABILITY SOFTWARE SYSTEMS

High Availability (HA) refers to the ability of resources in a computing system to remain available in the event of failures in the system. High availability systems are necessary as they enable business continuity and provide an expected and/or required level of availability to services provided to customers.

Engineering for HA Systems

From an engineering standpoint, key areas of focus to consider when designing high availability systems include:

- Eliminating single points of failure. This includes providing redundant components so that the failure of a primary component does not mean the entire system fails.
- Providing reliable failover between components. Failover between primary and redundant components requires cross-over connectivity and this can become a single point of failure if alternate paths are not included in the design. Four arrangements of redundant components are shown in the following figures (Active, Passive, N+1, Cold).
- Real-time performance monitoring. Detection of failing and failed components in near real time is critical for preventing failures and for reacting quickly to failures with automated recovery and/or manual intervention.

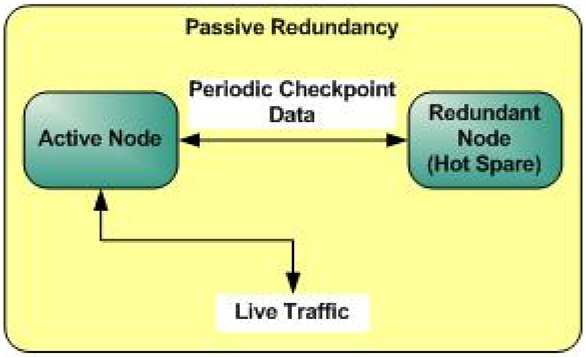


Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 4 - Active Redundancy Failover Pattern

When designing systems conforming to the Active Redundancy Failover Pattern (see Figure 4), all active and redundant components (nodes) receive and process all inputs in parallel. This results in the active and spare components having an identical state at all times. In the event of a failure or urgently needed downtime (for example, to replace a component), this can occur in milliseconds.

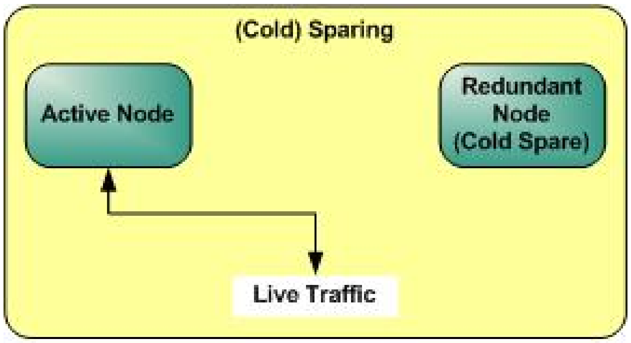
Additionally, both the active and redundant components can operate in an online configuration that provides increased capacity. In the event of failure of a component, the system can continue to operate, albeit with reduced capacity. Failover is usually invisible to the clients because they are connected to the server using virtual IP addresses managed by the load balancer.



Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 5 - Passive Redundancy Failover Pattern

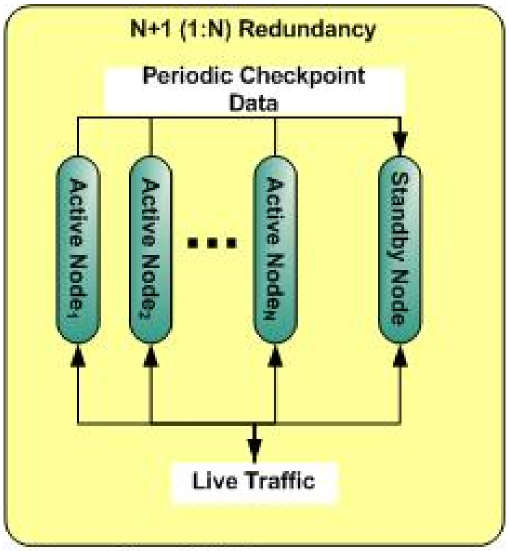
When designing systems conforming to the Passive Redundancy Failover Pattern (see Figure 5), one of the components receives and processes all inputs. The redundant component, known as a warm spare, receives periodic updates that synchronize it with the active component. In the event of failure, small amounts of data or transactions may be lost due to latency in the synchronization processes. Nevertheless, failover can occur quickly, generally within a few seconds. This pattern provides a more cost effective solution for environments that don't need near real time failover without any loss of data. Failover in this scenario is visible to the users, as they typically must reconnect to the server.



Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 6 - Cold Sparing Redundancy Failover Pattern

When designing systems that conform to the Cold Sparing Failover Redundancy Pattern (see Figure 6), one of the components receives and processes all inputs. The redundant component, known as a cold spare, is updated by applying periodic backups from the active node. These backups may be daily or more frequent, such as every five to ten minutes. In the event of failure, any required backups are applied to the cold spare, the processes on the cold spare are started, and the clients connect to the cold spare. When the active node is repaired, the process is reversed or alternatively, the original active node becomes the cold spare.



Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 7 - N+1 Redundancy Failover Pattern

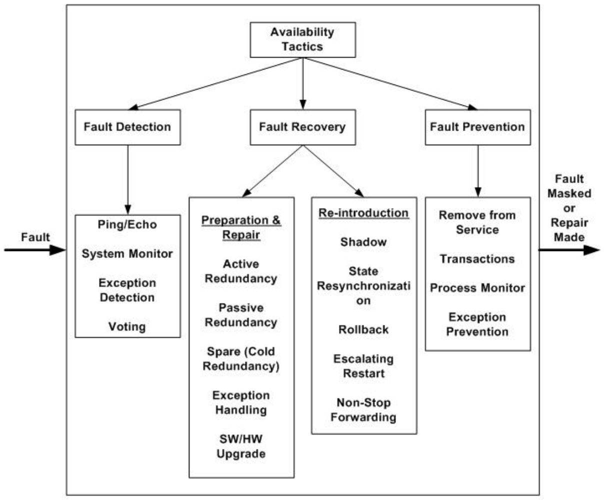
When designing systems that conform to the N+1 Redundancy Failover Pattern (see Figure 7), each of the 1–N active nodes operates as an independent device, receiving input, processing data, and providing output. Each of the active nodes periodically sends checkpoint messages to the standby node so it stays closely synchronized with the active nodes. If one of the active nodes fails, the standby node is immediately configured as the active node and begins processing as if it were the node it replaced. When the failed active node is repaired, it is brought online and the standby node resumes its function as the standby, first by re-synchronizing with the N active nodes.

Operational Considerations for HA Systems

Assuming that redundancy has been engineered into the system, the focus of the business continuity teams include:

- Fault Detection. This includes real time monitoring and oversight of the components with automated processes that detect and initiate failover.
- Fault Recovery. If a component fails, multiple paths may be followed depending on the type of failover (active, passive, N+1, cold spare) that was engineered into the system.
- Fault Prevention. This includes analyzing the monitoring data, constantly checking the system health parameters, or executing automated processes that prevent the occurrence of faults.

Figure 8 presents various approaches that operators can use for fault detection, fault recovery, and fault prevention.



Source: CarnegieMellon - http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 8 - SEI Availability Architecture Tactics

Figure 8 depicts the flow of a fault (on the left) through detection, recovery, and prevention, exiting (on the right) with the fault masked or a repair made.

The following sections present additional descriptive narrative for items that are of primary interest to operators thinking about using DOCSIS Remote PHY systems.

Availability Tactics – Fault Detection

Fault detection is focused on monitoring the health and performance of the system and reporting the health and any anomalies to a control system. From an architecture perspective, isolating fault detection activities from recovery and prevention activities increases independence of detection so it is not unduly influenced by the other activities. Three key aspects of fault detection are shown below:

- Ping/Echo. Pinging checks for a heartbeat in a system of component within a system. Periodically sending a message and receiving an event from a monitored system is a simple way to determine if a system is responding to input in the way that you expect. Lack of an echo from a ping request can be the first indicator that a system is experiencing difficulties.
- System Monitor. System monitoring is frequently implemented by installing an agent on a system or component that measures health-related operational parameters in real or near real time. If a measured parameter exceeds established thresholds, an exception or error message can be issued to a control system, which then takes recovery or preventative actions. Also, if the repair can be handled locally, the component can take immediate action (on its own) without direction from a controller.
- Voting. Generally used in real time systems that are in sensitive environments or performing mission critical functions. Frequently, three independent implementations of the same function vote on the next function or action and if two of the three voters agree, that action is taken. If the voters don't agree, then an exception can be raised and communicated to the control system.

Availability Tactics – Fault Recovery

Fault recovery is focused on getting the system back to a fully operational state. Four key fault recovery activities are described below:

- Preparation & Repair: Exception Handling. In general, every system fault is the result of a failure (or error) in some portion of the system. A popular approach for capturing the fault and reporting is through the concept of raising an exception. An exception captures an error event and provides additional details to a component or control system that are useful for triage and repair.
- Preparation & Repair: SW/HW Upgrade. This leverages the various failover models by upgrading the non-active system and failing over to the system with the new software/hardware. In the active-active scenario, one of the nodes is taken offline, upgraded and then brought back online.
- Re-introduction: State Resynchronization. This tactic periodically updates the state of the passive and cold spare nodes to the current state. By definition, the current state is the state at the time of a snapshot. Note, that as soon as a snapshot is taken it becomes stale, as the state of the active component is dynamically changing. However, the snapshot provides the ability to recover to a known state. The frequency of taking snapshots depends on the latency that is viewed as acceptable. The amount of latency that is deemed acceptable is often defined in a Service Level Agreement (SLA).
- Re-introduction: Escalating Restart. This tactic acknowledges the need to restart components in a known sequence. For example, a network node may need to be restarted prior to a related video playout server. Another example is restarting controllers before the child devices authenticate.

Availability Tactics – Fault Prevention

Two popular fault prevention activities are described below:

- Remove from Service. Periodically a system or component causes faults in other systems and components (that is to say, it has gone rogue). In this case, take the offending system offline, perform maintenance and repairs and then bring the system back online. Another possibility is to remove a system from service altogether.

- Exception Prevention. As the saying goes "An ounce of prevention is worth a pound of cure." This means that identifying, isolating, and fixing faults before they negatively impact system performance is always preferable to waiting for downtime and unhappy customers.

APPLYING HIGH AVAILABILITY DISTRIBUTED SYSTEMS MODELS FOR DOCSIS R-PHY RELIABILITY

Combining the concepts of DOCSIS Remote PHY and systems engineering for High Availability leads us to recommendations for creating highly available Remote PHY systems. Keys to fault tolerance in Remote PHY include building protective redundancy into the system and including fault detection, recovery, and prevention functionality as part of the core design and implementation.

These are the different EQAM redundancy types that can be used to define different quality of service (QoS). Note: "Loss of service" means loss of set-top box synchronization.

- Redundancy Type I - User transparent, no loss of service.
- Redundancy Type II - Not user transparent, no loss of service.
- Redundancy Type III - Not user transparent, with momentary loss of service.
- Other – (Not defined in the EQAM doc). Not user transparent, significant duration loss of service.

Cable operators must at least support Type III redundancy: "Not user transparent, momentary loss of service." (For details, refer to CM-SP-EQAM-VSI-I01-081107, section 11.1.2.) These QoS definitions are directly applicable to Remote PHY and are closely correlated to subscriber satisfaction.

The CCAP Technical Report, "High Reliability and Redundancy Capabilities" (CM-TR-CCAP-V03-120511, section 5.2.3), provides guidance on developing HA capabilities that address hardware and network redundancy.

Architecture and Failover Scenarios

EQAM Redundancy Quality of Service types can be directly mapped onto the primary failover architecture models (Active, Passive, N+1, and Cold) described earlier. Table 1 summarizes the mappings and includes one additional variation for each of the active and passive failover architectures. The following sections elaborate on the information presented in the table.

Table 1 - EQAM QoS Redundancy Types and Architecture/Failover Scenario Mapping

EQAM QoS Redundancy Type		Architecture and Failover Scenarios					
		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Type	**Description**	**Active-Active**	**Active-Passive**	**Active-Passive Remote**	**Parallel Path**	**N + 1 Redundancy**	**Cold Standby**
I	Transparent, no loss of service	Yes			Yes		
II	Not transparent, no loss of service	Yes			Yes		
III	Not transparent, momentary loss of service		Yes	Yes		Yes	
Other	Significant loss of service						Yes

Earlier, this document presented key architecture patterns that support high availability. These patterns are applied to Remote PHY in the following diagrams. The first five scenarios are preferred. Scenario 6 is included for completeness but is not recommended, as this scenario will result in significant loss of service during manual failover and restart. The scenarios are described below.

Active-Active Failover at the Node

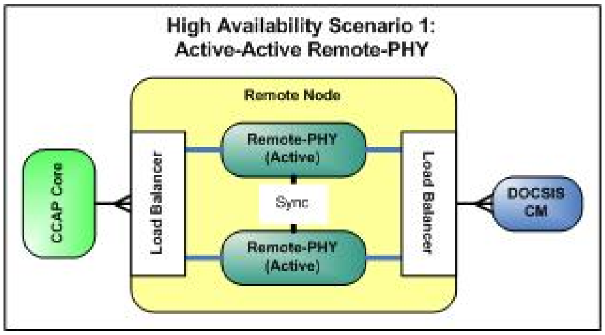


Figure 9 - Remote PHY Active-Active Failover Scenario

In the Active-Active failover scenario, both instances of Remote-PHY process signals and remain synchronized in real time using messaging between the devices. Load and virtualization is managed via the load balancers at both ends of the access point. When a failure occurs, all traffic is routed transparently through the remaining Active device by the load balancers. After the fault is repaired, the second device is brought back online and traffic is again routed through the device.

Active-Passive Failover at the Node

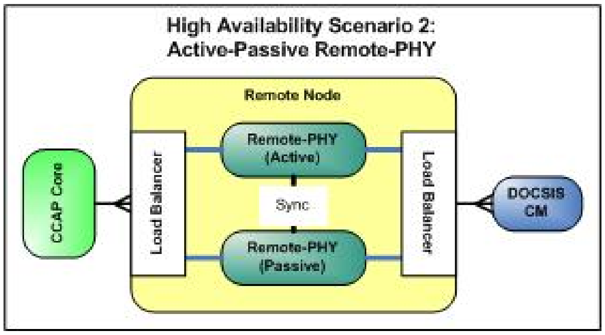


Figure 10 - Remote PHY Active-Passive Failover Scenario

The Active-Passive failover scenario routes all traffic through the Active Remote-PHY device via the load balancers at both ends of the access point. The Passive device is periodically synchronized with the Active device through shared messages. If the Active device fails, all traffic is routed transparently through the Passive device, which becomes the new Active device. After the fault is repaired, the device is brought back online and either becomes the active or the passive device.

Active Remote-PHY Failover at the Core

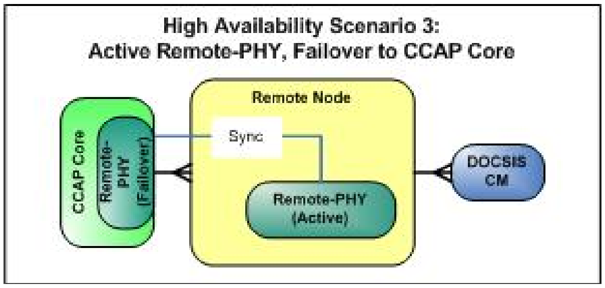


Figure 11 - Remote PHY Active-Remote Failover Scenario

The Active-Remote-PHY failover at the core scenario is a variation on the scenario described in Figure 10 (Active Passive Failover scenario). All traffic is routed through the Active Remote-PHY device. The Passive device is periodically synchronized with the Active device through shared messages. If the Active device fails, all traffic is routed transparently through the Passive device, which becomes the new Active device. After the fault is repaired, the device is brought back online and becomes the Active device.

Parallel Path Active-Active Failover at the Node

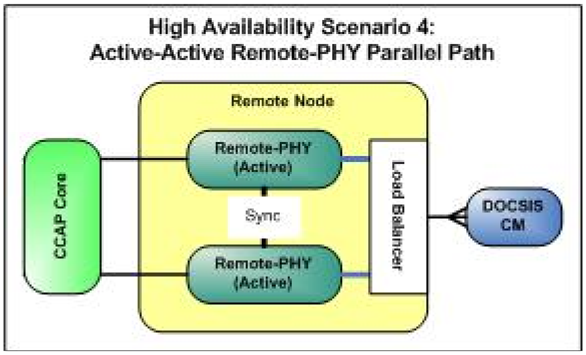


Figure 12 - Remote PHY Active-Active Parallel Path Failover Scenario

The Parallel Path Active-Active failover at the node scenario is a variation on the scenario described in Figure 9 (Active–Active Failover).

In this scenario, both instances of Remote-PHY process signals remain synchronized in real time through messages exchanged between the devices. Load and virtualization with the DOCSIS Cable Modems (CMs) is managed via the load balancers at the node. Between the node and the CCAP Core, link redundancy eliminates the need for a load balancer in the remote node. The CCAP Core in effect acts as the load balancer. If one of the devices fails, the Load Balancer routes all traffic transparently through the remaining Active device. After the fault is repaired, the second device is brought back online and traffic is again routed thru the device.

Active with N+1 Redundancy Failover at the Node

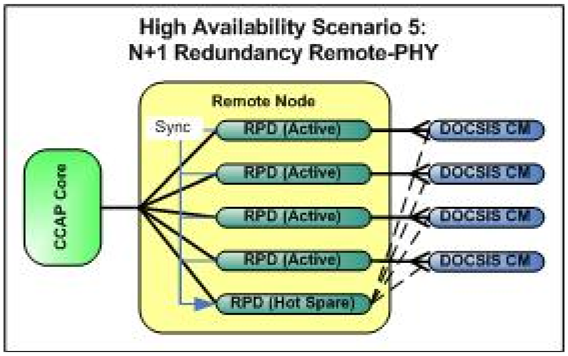


Figure 13 – Remote PHY N+1 Failover Scenario

The Active with N+1 Redundancy Failover scenario uses a hot standby spare (similar to the Active-Passive failover scenario) but instead of a one-to-one active to standby spare, this model has N active devices with a single spare available to all devices. In this example, N = 4 (plus the one spare). The Hot Spare (Passive) device is periodically synchronized with the Active devices through shared messages. If an Active device fails, all traffic is routed transparently through the Hot Spare device, which becomes the new Active device. After the fault is repaired, the device is brought back online and either become the active or the passive device.

Active-Cold Standby Failover at the Node

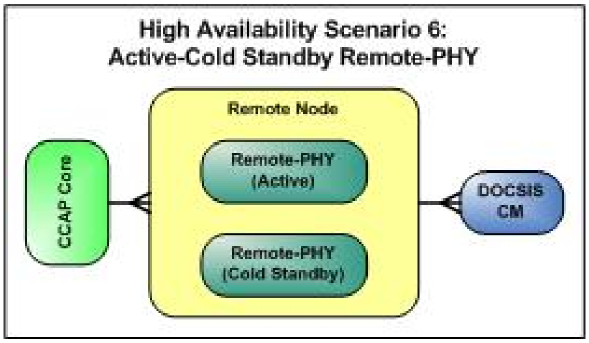

High Availability Scenario 6:
Active-Cold Standby Remote-PHY
Remote Node
Remote-PHY (Active)
Remote-PHY (Cold Standby)
CCAP Core
DOCSIS CM

Figure 14 - Remote PHY Active-Cold Failover Scenario

The Cold Standby scenario is included for completeness but is not recommended. This failover model should be avoided because it typically results in hours to days of recovery effort applying backups, configuring the devices and networks and bringing the standby environment online. This model is generally used with batch processes that can tolerate significant downtime. An example of a system that could use this model is a data warehouse that only generates reports sourcing datasets that are loaded on a monthly basis.

SUMMARY

High Availability is directly applicable to the DOCSIS Remote PHY architecture and should be a part of any operator's design. HA is always a balance between competing needs including cost, availability, and latency. The solution you choose will be determined by your various constraints. Active-Active architectures provide the lowest latency but N+1 redundancy is versatile and not terribly expensive to implement. Cold spare redundancy is the lowest cost solution, but has the highest latency and generally the worst availability. In all cases, the balance between cost, availability, and latency that best meets the business and customer needs is the best solution. Consider all the available options before you settle on a particular architecture.

In the communications services industry, N+1 redundancy is frequency used, especially for business critical application such as customer-facing services like streaming video and data. High Availability is especially important if your customers are business users.

High Availability is a rapidly evolving area and there are many opportunities to incorporate best practice recommendations into your HFC network.

REFERENCES

i. Chapman_John_PPT_MHAv2 SCTE (2013-08-30), John T. Chapman, Cisco Inc.

ii. Remote PHY Specification, CM-SP-R-PHY-D03-150320, Cable Television Laboratories, Inc.

iii. Realizing and Refining Architectural Tactics: Availability, Carnegie Mellon Software Engineering Institute, http://www.sei.cmu.edu/reports/09tr006.pdf

iv. Edge QAM Video Stream Interface Specification, CM-SP-EQAM-VSI-I01-081107, Cable Television Laboratories, Inc.

vi. CCAP Architecture Technical Report, CM-TR-CCAP-V03-120511, Cable Television Laboratories, Inc.