Multicast Orchestration with Arista Media Control Service Feature Updates

High Performance Cognitive Orchestration for Multicast Workflows

Unveiling the latest feature updates in Arista's Media Control Service (MCS) through this whitepaper, we highlight significant enhancements to its capabilities. The recent updates introduce three key functionalities: maintenance mode, the ability to orchestrate seamlessly across multiple zones, and the capability to perform load sharing. MCS, built on the solid foundation of EOS, now incorporates these cutting-edge features to elevate its cognitive orchestration for multicast flows. These updates play a pivotal role in fortifying high-bandwidth critical broadcast workflows by providing maintenance flexibility, enabling efficient orchestration across diverse zones, and facilitating load sharing for optimized performance. As a Network Control Solution, MCS continues to complement the Broadcast industry's transition to IP, offering an enhanced layer of security, performance, and determinism for real-time on-prem and remote production workflows.

Overview

As increasing scale and performance requirements drive the trend toward orchestration gains momentum across various systems, it poses a challenge for existing orchestration technology to evolve in tandem. To meet this challenge, selecting an architecture capable of adapting and growing to meet new demands becomes crucial. Arista switches, grounded in cloud networking principles emphasizing high performance, streaming telemetry, and analytics, present broadcasters with dependable, cost-effective, and highly scalable platforms to construct and extend their dynamically evolving workflows. Leveraging Arista's top-quality hardware and the robust self-healing EOS software,multicast flow orchestration, with MCS, guarantees that on-air broadcast operations are directed and safeguarded across workflows.

MCS Architecture and Workflow

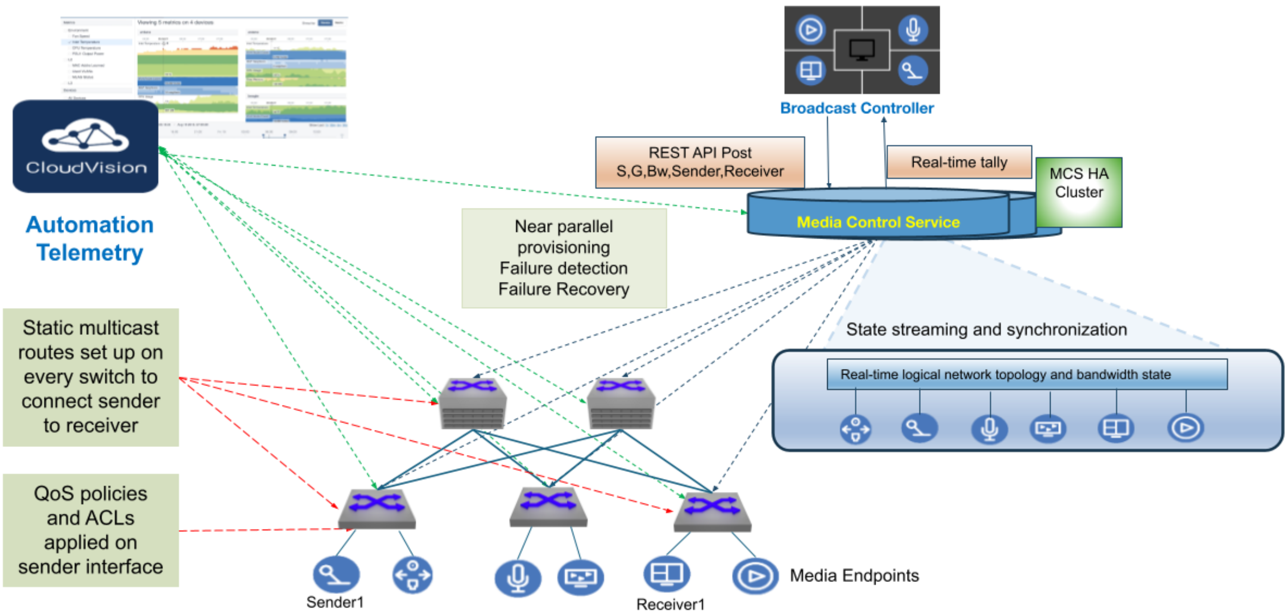

Figure 1: MCS Architecture

Here we present a comprehensive framework that introduces a layer of abstraction atop a Media over IP network, offering a groundbreaking solution for the evolving landscape of broadcast control. This innovative approach involves MCS deployed as a CloudVision Exchange(CVX) cluster to provide High Availability and Data redundancy. The Arista Extensible Operating System (EOS) is virtualized as a CVX instance, providing a centralized point of access for multicast orchestration. This, coupled with continuous synchronization and aggregation of real-time network state, covers topology, capacity, and multicast information.

The Broadcast Controller seamlessly interacts with our Media Control Service (MCS), utilizing MCS to provision multicast senders to receivers. MCS, in turn, identifies the best available path with bandwidth availability, configuring Quality of Service (QoS) policies solely on the ingress interface of the First Hop Router (FHR) and static multicast routes across each hop to reach the receiver. Additionally policing is also added per multicast flow, thereby providing a fine grained approach to bandwidth management.

In scenarios where endpoints are connected to switchports, an IGMP snooping querier is configured for the VLAN to prevent undesired multicast flooding within the VLAN. The application of QoS exclusively on the FHR's ingress port is designed for TCAM optimization and resource conservation. Furthermore, MCS can also be utilized on switches with limited TCAM scale by utilizing the knob to disable policer creation if desired. MCS continues to manage bandwidth to prevent oversubscription even if policer creation is disabled.

While the Broadcast Controller maintains its traditional control matrix view and remains topology-agnostic, it informs MCS of all sender details, including Multicast IP, Source IP, Incoming Interface, Switch Mac address, and Bandwidth. The API integration between MCS and the Broadcast Controller is facilitated through secure HTTP[S] APIs, enabling the acquisition of network and topology information and the configuration of multicast flows. Additionally, notification channels allow the Broadcast Controller to subscribe to real-time telemetry updates and multicast flow configuration status updates. Importantly, our approach eliminates reliance on dynamic protocols in pathfinding, opting for a straightforward configuration of static multicast routes. All configurations, including QoS and multicast configurations programmed by MCS, are not written to switch running-configuration so it will not be loaded as part of startup-configuration. This is essential to prevent any disruptions caused by outdated or absent configurations during switch reboots and upgrades, thereby improving the usability and maintainability of multicast state at scale through MCS's state synchronization mechanism in the broadcast workflow.

The crucial aspect of MCS lies in the real-time feedback provided in the form of a tally, ensuring the Broadcast Operator is informed of the success or failure of each take. This tally not only fills a crucial gap left by IGMP but also serves as a tool for identifying any changes that affect MCS programmed flows including unexpected link/switch failures causing impacted flows and scheduled maintenance mode on devices. The entire state is efficiently streamed in real-time to our CloudVision Portal (CVP) for in-depth monitoring and troubleshooting.

New Features

Maintenance Mode for multi-spine deployments

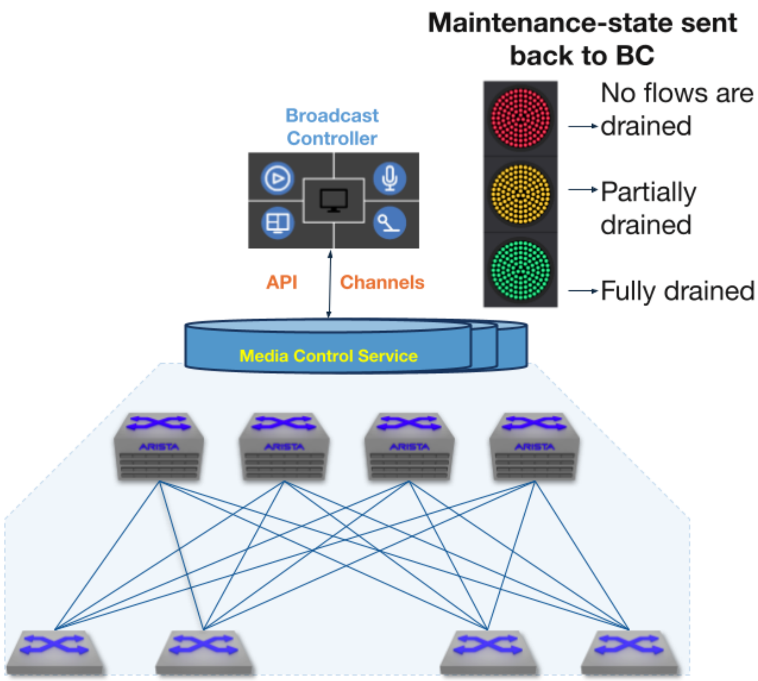

Figure 2: Maintenance Mode

This feature allows for the assignment of MCS-controlled devices to maintenance mode, enabling the redirection of flows from that switch to another switch based on operator preferences. This feature becomes particularly advantageous in scenarios where multiple paths exist between the sender and receiver.

The Broadcast Controller initiates a request to MCS to designate a specific device to maintenance mode. MCS, in response, marks the device as under maintenance mode and reveals the receivers affected by this status. Upon the Broadcast Controller's request, MCS proceeds to reprovision the receivers that have been marked with the maintenance flag, utilizing an alternative path, provided bandwidth is available. In instances where bandwidth is unavailable, MCS will not move the flows from the maintenance device. A switch that has been assigned to be under Maintenance cannot be included in the path for any new flows provisioned by MCS. Once the maintenance is concluded and the devices are reinstated, they become available for new flows. This flexible API-driven process enhances operational control and efficiency enabling broadcasters to keep on top of deploying new EOS with features, bug fixes, or security patches. As an integral component of this workflow, MCS also issues a notification regarding the maintenance state of the device, represented in a traffic light pattern. In this pattern, the Red state indicates that no flows have been diverted from the maintenance device, the Amber state indicates that some flows have been partially redirected, and the Green state signifies that all flows have successfully been transferred from the maintenance device.

Load Sharing of Programmed Flows

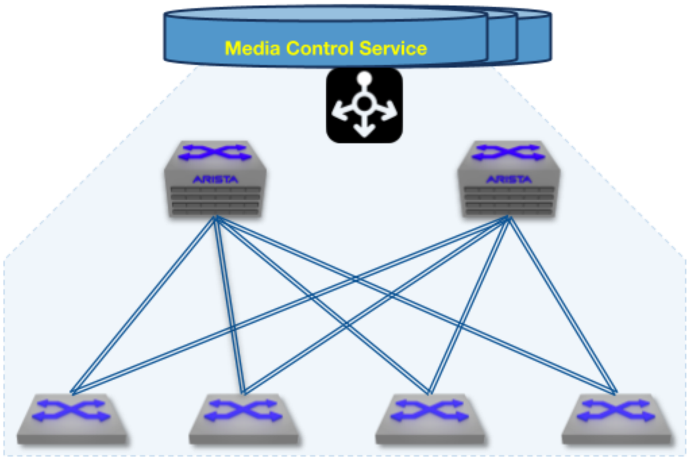

Figure 3: Load Balancing

In diverse broadcast workflows, the distribution of load in the IP network can vary, prompting the need for adaptable solutions. To cater to these preferences, MCS offers the flexibility of operating in either default flow programming mode or load balancer-enabled mode.

Load balancing of programmed flows is a critical aspect, and MCS facilitates this across multiple dimensions, including multiple links and multiple spines. In the absence of load balancing, the default flow programming selects a spine randomly based on available bandwidth, with the first available bandwidth link chosen to exhaust. Subsequent flows landing on that spine continue to utilize the initially selected link before moving on to the next available one.

With load balancing activated, MCS employs the Consistent Hashing algorithm, calculating a hash based on the switch MAC address and interface ID to select a spine with the maximum total available bandwidth. The application of Consistent Hashing is instrumental in preserving route consistency across multiple links and spines, allowing MCS systems to scale up or down seamlessly without disrupting existing workflows. Additionally, within each spine, MCS calculates a hash based on Source (S), Group (G), and switch MAC addresses to determine the optimal link. This not only ensures efficient load distribution but also provides the inherent benefits of Consistent Hashing, such as maintaining a deterministic mapping between flow and links, minimizing reconfiguration, and enabling effective scaling in deployments. Incorporating Consistent Hashing into load balancing enhances the scalability, resilience, and overall performance of broadcast workflows supported by MCS.

Multi-zone multicast orchestration

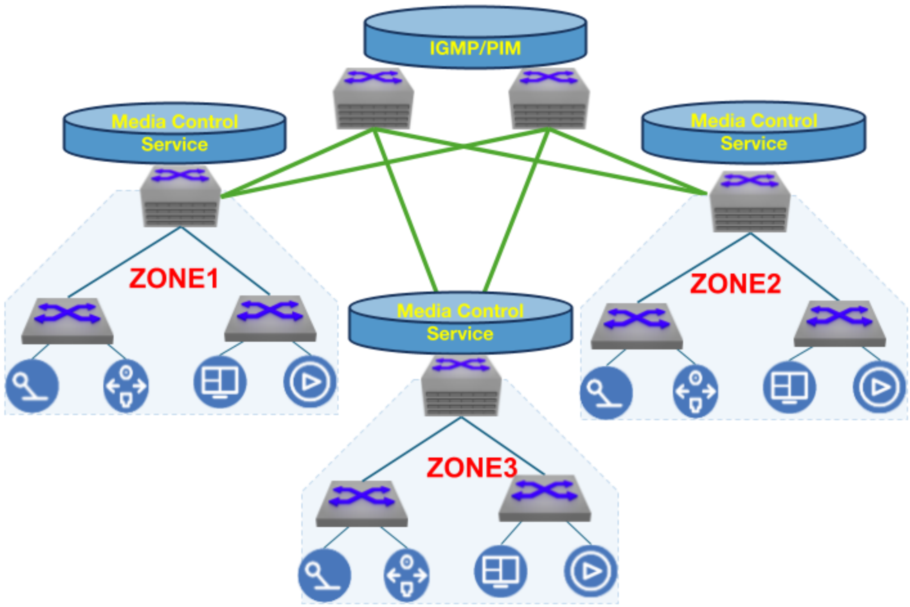

Figure 4: Multi-Zone Multicast Orchestration

In the modern landscape of broadcasting and media production, the adoption of multiple media zones by broadcasters and media houses is driven by various compelling reasons. As we transition towards increasingly interconnected and collaborative media workflows, the ability to seamlessly connect these zones proves to be strategically advantageous. One pivotal factor is resilience, where the segmentation of media production facilities into distinct zones mitigates the impact of maintenance, failures, or mis-configurations. This compartmentalization ensures that disruptions in one zone do not cascade across the entire system, thereby enhancing overall operational robustness. Furthermore, the integration of media flows across multiple existing networks is pivotal for maximizing flexibility. By enabling interoperability and multicast flow sharing between media zones, broadcasters can optimize their workflows and enhance the efficiency of media asset utilization.

The zones are connected by different connectivity types, MCS <-> MCS, MCS <-> IGMP/PIM. We take advantage of IIF Aware IGMP Host Proxy and MCS Boundary feature to facilitate these different connectivity types.

IIF Aware IGMP Host Proxy

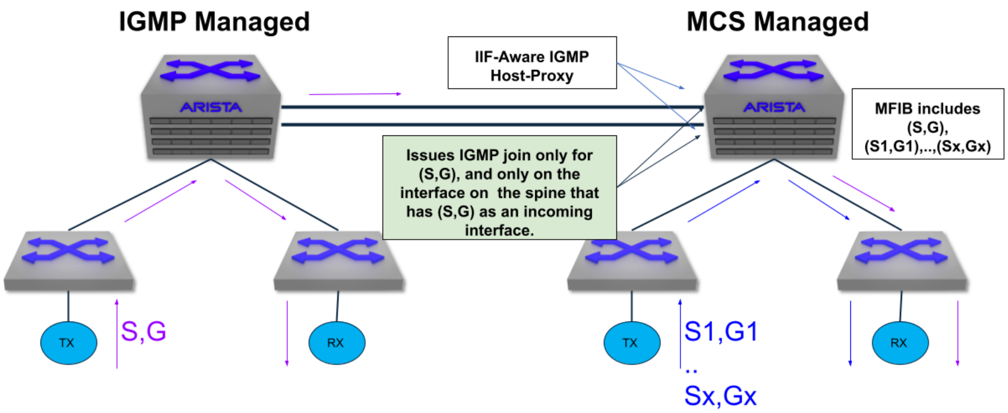

Figure 5 : IIF Aware IGMP Host Proxy

To establish a connection between MCS and a dynamically controlled multicast zone, it is essential to implement IGMP Host Proxy on MCS edge devices.The existing behavior of the IGMP host proxy is to send IGMP Joins on all IGMP host proxy enabled interfaces for all entries in the MFIB. This is very inefficient from both a cpu and bandwidth perspective.

Certain broadcast workflows require the capability to direct traffic from the IGMP zone to MCS zone, considering factors such as bandwidth availability and load balancing. The MCS zone is responsible for providing a list of interfaces that a multicast flow may traverse upon a route request. Once an interface from the list is selected, an IGMP Join will be triggered by the IGMP Host Proxy configuration for that specific multicast flow. Following the received configuration, IGMP Host Proxy will be enabled on the specified interfaces. Within this subset of IGMP Host Proxy-enabled interfaces, IGMP joins will exclusively be transmitted on the interface identified as the IIF for the corresponding (S, G) pair.

To meet these requirements, an enhancement was made for the existing IGMP Host Proxy feature. This enhancement involves introducing a novel IIF-aware mode for IGMP Host Proxy, ensuring that IGMP joins for a particular (S, G) are sent solely on the interface established as the IIF for that specific (S, G). A mechanism will be incorporated to facilitate the activation and configuration of IGMP Host Proxy on individual interfaces, specifically those received from MCS. The existing method of enabling and configuring IGMP Host Proxy is on an interface or a group of interfaces. The control for the IIF-aware mode will be at a per-interface level, presently under the control of MCS. By default, all interfaces enabled by MCS for IGMP Host Proxy will operate in IIF-aware mode, meaning that the IGMP Host Proxy agent will enable IGMP Host Proxy on those interfaces based on the list received from MCS.

MCS boundaries

As we aim for MCS to perform load-balancing across a set of interfaces connecting multicast zones, it is imperative to define a boundary with the above-mentioned interfaces connecting the zones. This list is defined as a set of interfaces and switches. For example, Boundary1 = Sw1: Eth1, Sw1: Eth2, Sw1: Eth3, while Boundary 2 = Sw2: Eth1, Sw2: Eth2, Sw3: Eth1, S3: Eth2, thus possibly spanning across multiple switches. The boundary can be bidirectional so the interfaces can act both as senders and receivers. The designated list serves the purpose of load-balancing flows, allowing the reception and transmission of flows from and to a group of ports. In both scenarios, the Broadcast Controller is responsible for publishing this list to MCS. The Broadcast Controller will be notified through MCS's notification channels regarding the interface picked by MCS in either load-balancing direction for each individual S,G.

MCS provides GET and notification APIs to provide telemetry on defined boundaries and associated interfaces. MCS also exposes the selected interface in either direction. In the event of link failure, MCS allows the Broadcast Operator to retain his choice in order of recovery. MCS informs the Broadcast Controller about affected receivers, and it is expected that upon receiving a new request to reprogram the receiver, MCS will choose a new member from the list. In a list involving multiple switches, a device reboot prompts the Broadcast Controller to select any or all flows for repair, initiating another load-balancing operation within the list. For flows that are not repaired, when the device comes back online, the flows will be mapped as they were prior to the reboot.

MCS to MCS Zone

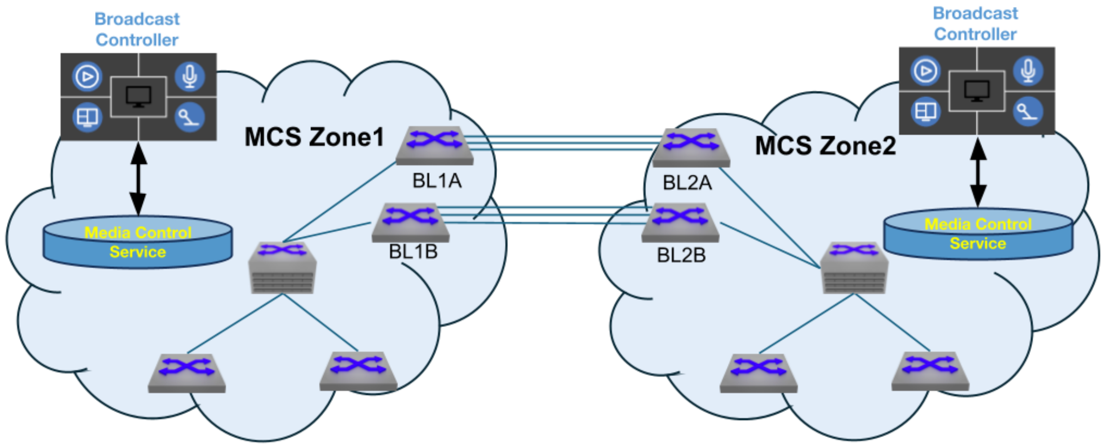

Figure 6 : MCS<->MCS

The depicted configuration illustrates Border Leafs, but these could also be Spine to SuperSpine connections. Multiple links interconnect the systems, and it is imperative to distribute the load across these links. BL1a/b and BL2a/b are subject to control by MCS1 and MCS2, respectively.

When a flow is needed from MCS Zone 1 to MCS Zone 2, Broadcast Controller 1 directs MCS1 to transmit flow X through the MCS2 Interconnect, selecting one of the links (Link1, Link2, or Link3 on BL1a or 1b). MCS1 designates a member from the group connecting MCS Zone 1 to MCS Zone 2 and reports this outgoing choice (Link3) to Broadcast Controller 1. Broadcast Controller 1 then informs Broadcast Controller 2 about this selected member. Subsequently, Broadcast Controller 2 instructs MCS2 to route flow X from Link3 (BL1b) to the required destination. Load-balancing responsibilities with MCS are assigned as follows: MCS1 manages load-balancing on the egress from MCS Zone 1, and MCS2 takes charge of load-balancing on the egress from MCS Zone 2. In case of insufficient bandwidth, the request will be unsuccessful.

MCS to/from IGMP/PIM

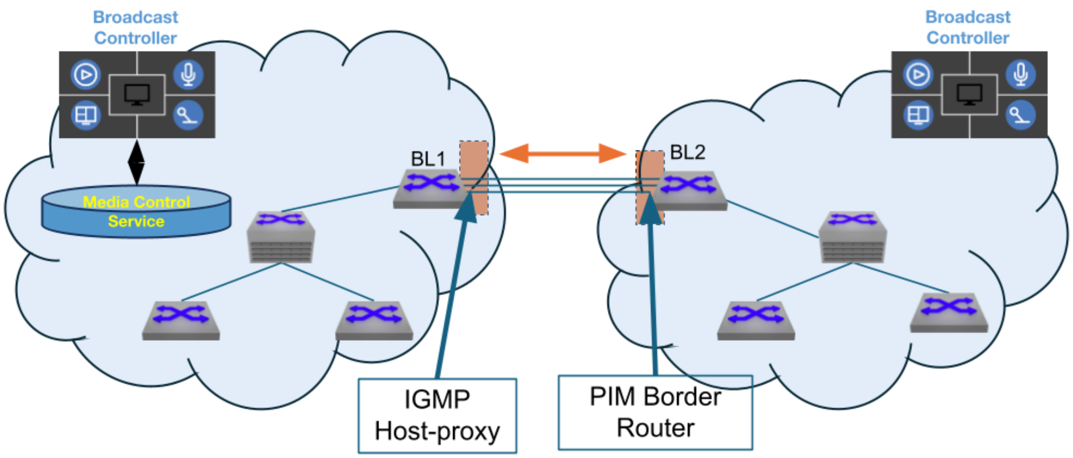

Figure 7: MCS<->IGMP

There is a need to facilitate the sharing of media flows between MCS and the IGMP zones. While Border Leafs are illustrated, these could alternatively be Spine to SuperSpine connections. The system incorporates multiple links, necessitating load balancing across them.

BL1 operates under the control of MCS1, while BL2 is integrated into the dynamic IGMP/PIM zone. The IGMP/PIM border leaf features a PIM Border Router configuration on its interconnect members, with no PIM interaction extending from the IGMP/PIM zone to the MCS zone. Multicast flows received from the MCS zone at the IGMP/PIM Zone are locally registered, enabling Broadcast Controller 2 to route them.

In a scenario requiring a flow from MCS Zone to IGMP/PIM Zone, Broadcast Controller 1 instructs MCS1 to send flow X to the IGMP/PIM Interconnect (Link1, 2, 3). MCS1 selects a member connecting the two zones, reports the outgoing interface to Broadcast Controller1 (Link3).

Broadcast Controller 2 does not need knowledge of the receiving member, as the IGMP/PIM system can identify the flow registration on the appropriate member and route it normally. MCS1 handles load-balancing on the egress from MCS Zone 1, and if bandwidth is insufficient, the request fails.

Similarly, for a flow required from IGMP/PIM Zone to MCS Zone, Broadcast Controller 1 directs MCS1 to receive flow X from the IGMP/PIM zone interconnect. MCS1 chooses a member connecting the zones, reports the incoming choice to Broadcast Controller 1 (Link3). The IGMP Host-Proxy on the MCS interconnect ports draws the flow from the IGMP/PIM Zone, and MCS1 manages load-balancing on the ingress to MCS Zone 1. If bandwidth is unavailable, the request fails.

MCS to/from IGMP/PIM SuperSpine to/from MCS

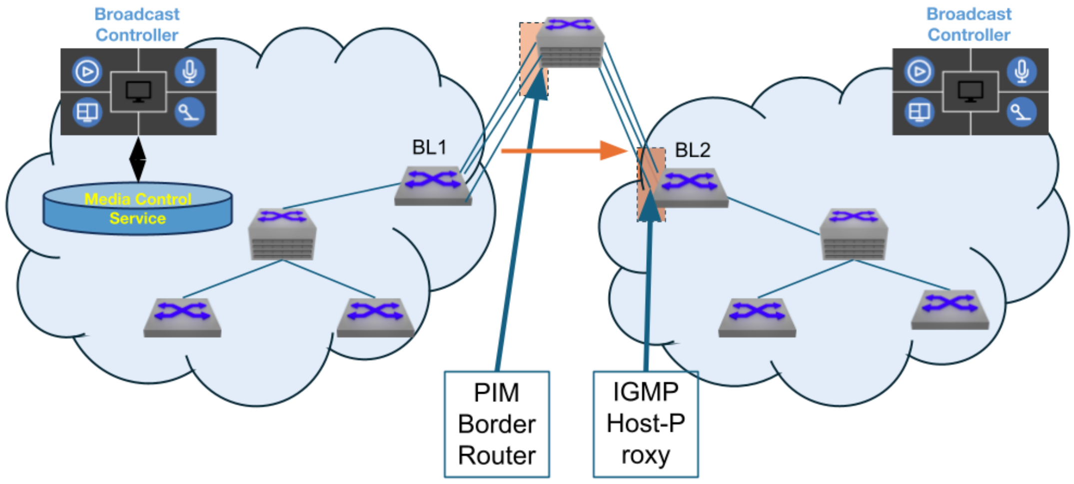

Figure 8: MCS<->IGMP Super Spine <-> MCS

Here the objective is to facilitate the sharing of media resources between MCS zones connected through SuperSpine[s]. Although Border Leafs are depicted, there is flexibility for them to be replaced with a direct connection from the SuperSpine to the MCSx Spine, rendering BL1/2 optional. Multiple links interconnect systems, necessitating the need for load balancing across these links. The SuperSpine can be connected directly to the MCSx Spine, bypassing BL1/2. To establish a flow from MCS Zone 1 to MCS Zone 2 through the SuperSpine, Broadcast Controller 1 directs MCS1 to transmit flow X to the SuperSpine Interconnect (Link1, 2, or 3). MCS1 selects a member connecting MCS Zone 1 to the SuperSpine and reports this choice to Broadcast Controller 1 (Link3). That S,G becomes locally routable on the SuperSpine as a result of the Multicast Border Router feature being enabled on those BL facing interfaces.. Load-balancing responsibilities on the egress from MCS Zone 1 are managed by MCS1, and if bandwidth is insufficient, the request fails.

In the reverse direction, when MCS Zone 2 pulls the flow from the SuperSpine, Broadcast Controller 2 instructs MCS2 to route Flow X from one of the links (Link 1, 2, or 3) to the required destination.The SuperSpine system recognizes the source flow X registered on the appropriate incoming member and can route it normally. Similarly, the SuperSpine system receives the IGMP Joins via IGMP-Host-Proxy and traffic is naturally forwarded to the MCS2 Zone.. Load-balancing on the egress from the IGMP/PIM SuperSpine is managed by MCS2, and if bandwidth is unavailable, the request fails. This symmetric process applies to the reverse flow direction as well.

Conclusion

Here, we have unveiled the latest feature updates in Arista's Media Control Service (MCS), showcasing three pivotal enhancements: maintenance mode, cross-zone orchestration, and load sharing. These updates fortify MCS's cognitive orchestration for multicast flows, providing maintenance flexibility, efficient orchestration across diverse zones, and optimized load sharing. Arista plays a key role in enabling a resilient, agile, and scalable architecture for broadcasters seeking to transition and expand their workflows. The demand for a solution that is both resilient and adaptable, ensuring the secure and deterministic transfer of multicast flows, is met by Arista's MCS. As broadcast workflows become more complex, there is an increasing need for a solution that can reliably and predictably handle multicast flows across any workflow and scale. MCS powers Multicast Flow Orchestration across diverse broadcast workflows and ensures on-air broadcast operations are directed and safeguarded across evolving workflows, aligning with the dynamic needs of the industry.Arista's MCS stands as a forerunner in the broadcast industry, offering a sophisticated solution that aligns seamlessly with the industry's dynamic landscape and growing demands.

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office
1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062