Global Knowledge®

Expert Reference Series of White Papers

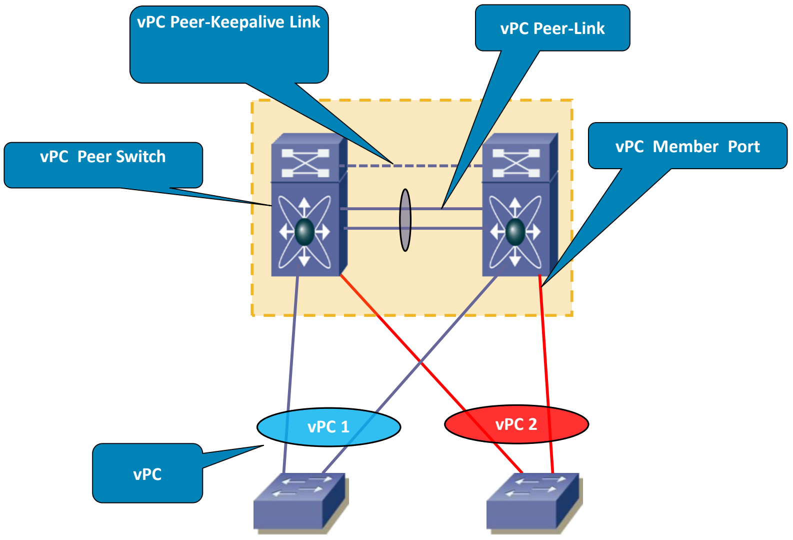
Multicast Implementation with Virtual PortChannels and FabricPath

Multicast Implementation with Virtual PortChannels and FabricPath

Carol Kavalla, CCSI, CCNP, CCDP

Introduction

The assumption is that the reader has a working understanding of multicast, virtual PortChannels (vPCs), and FabricPath. As a reminder, both vPCs and FabricPath allow all links to forward data. The limitations of Spanning Tree Protocol blocked ports are avoided and both vPC and FabricPath provide loop avoidance.



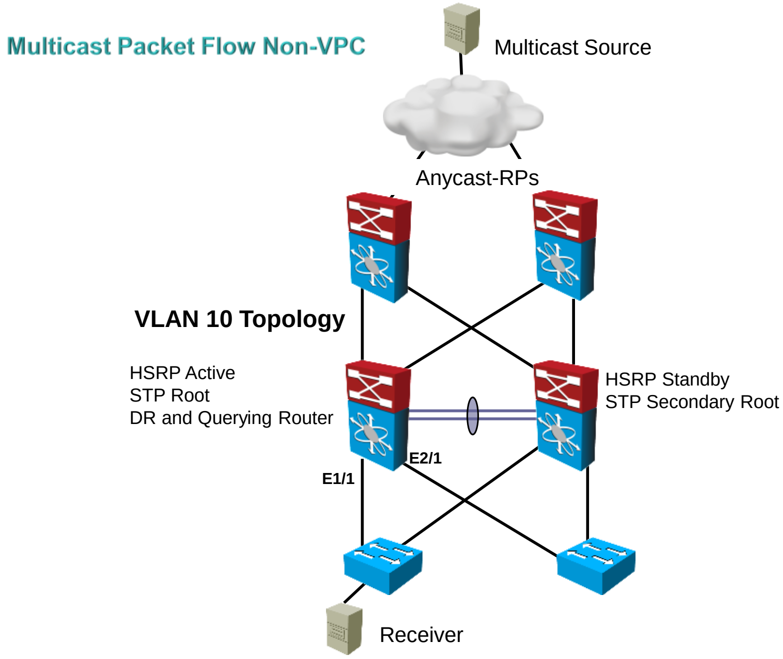
First let's do a quick review of the function and benefits of vPCs.

vPCs are an extension of port channels. Port channel technology does not allow members of a port group to terminate on more than one device. vPCs virtualize two Nexus switches to become one logical switch.

This provides two major benefits: the use of all available uplink bandwidth with loop avoidance and high availability for the downstream switches (as shown in the above illustration). vPCs are widely implemented on Cisco Nexus switches in a Data Center environment.

Cisco Nexus software supports PIM-SM (Protocol Independent Multicast—Sparse Mode) only. This paper looks at how vPC manages multicast traffic. The NXOS synchronizes the multicast forwarding state on both of the vPC peers. Similarly, the group information learned via IGMP Snooping is shared between vPC peers as well. The PIM process in vPC mode ensures that only one of the vPC peers is actively forwarding multicast data downstream to receivers.

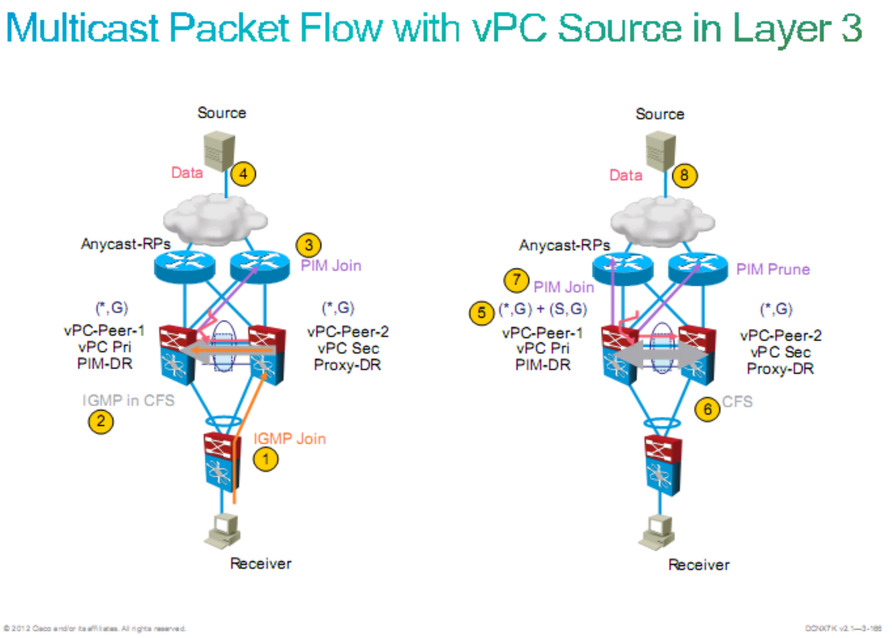
Multicast Packet Flow without vPC



In the diagram above, the receiver resides in VLAN 10.

1. The receiver sends an Internet Group Message Protocol (IGMP) Join toward the access switch. The Access Switch creates a Layer 2 IGMP snooping entry and forwards the IGMP Membership Report to the DR.
2. The DR creates a Layer 3 *,G entry in its PIM routing table. Interface E1/1 will be the interface in the outgoing interface list (OIF) for the *,G entry
3. The DR creates a *,G PIM Join/Prune message and sends it toward the RP
4. Assuming the RP is already receiving multicast data, the RP forwards data traffic down the shared tree toward the DR
5. The DR forwards the traffic out the interface in its OIF, E1/1
6. As the first incoming multicast packet causes the default threshold of 0 to be exceeded, the DR switches over to the Shortest Path Tree (SPT) and sends a PIM S,G join up the tree toward the Source
7. Multicast data now flows down the SPT to the DR and ultimately to the receiver

Multicast Packet Flow with vPC Source in Layer 3

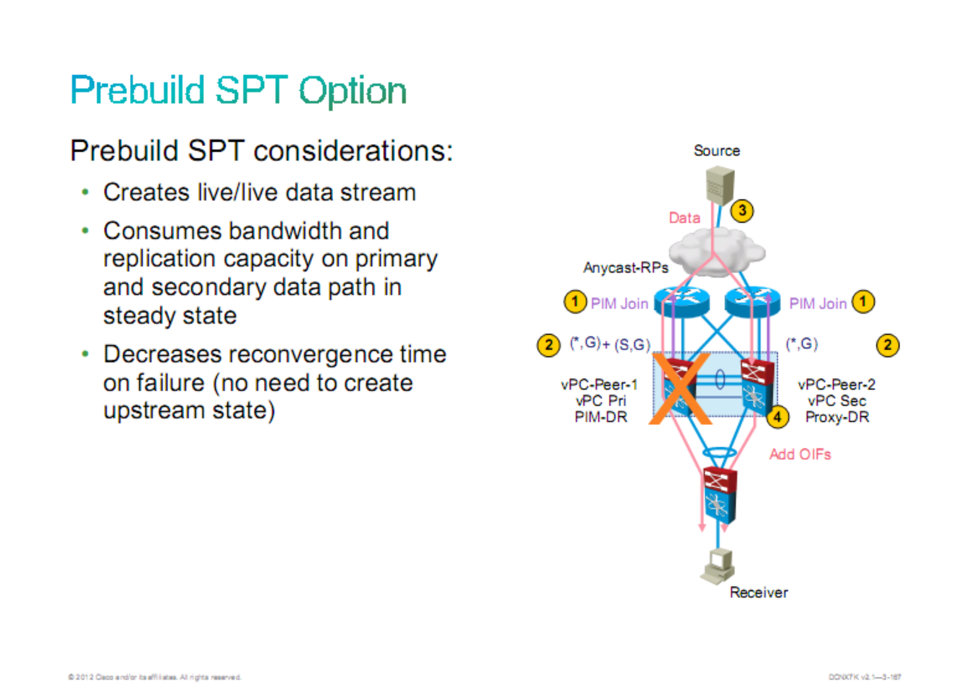


Multicast Packet Flow with vPC Source in Layer 3

DCNX7K v2.1—3-188

1. The receiver sends an IGMP Join toward its access switch. The link to Peer-2 is chosen as the upstream interface. Peer-2 creates a Layer 2 IGMP snooping entry as well as a Layer 3 *,G entry in its PIM routing table. The vPC will be the interface in the outgoing interface list (OIF) for the *,G entry
2. Peer-2 sends an IGMP packet encapsulated in Cisco Fabric Services (CFS) to Peer-1
3. Peer-1 (note it is the DR) sends a PIM *,G join toward the RP to join the shared tree
4. Peer-2 is the vPC proxy DR and is not responsible for sending PIM Joins
5. Assuming the RP is already receiving multicast data, the RP forwards data traffic down the shared tree toward Peer-1
6. Peer-1 forwards the multicast traffic to Peer-2 on the peer-link and the traffic is dropped by Peer-2
7. As the first incoming multicast packet causes the default threshold of 0 to be exceeded, Peer-1 switches over to the Shortest Path Tree (SPT)
8. The peer switches negotiate via CFS for who will have the forwarder role for the SPT
9. In the above example, Peer-1 is elected the forwarder for the SPT and forwards the S,G PIM join toward the source.
10. Multicast data now flows down the SPT to Peer-1

The NXOS has an option for configuring a prebuilt SPT on the vPC peer that was not chosen as the SPT forwarder. This decreases the convergence time as the non-forwarding vPC peer already has created the SPT state toward the source. The possible disadvantage to consider is that your network will have a live-live data stream that will use up bandwidth and replication capacity on the secondary SPT.



Prebuild SPT Option

Prebuild SPT considerations:

- Creates live/live data stream
- Consumes bandwidth and replication capacity on primary and secondary data path in steady state
- Decreases reconvergence time on failure (no need to create upstream state)

DCNX7K v2.1—3-167

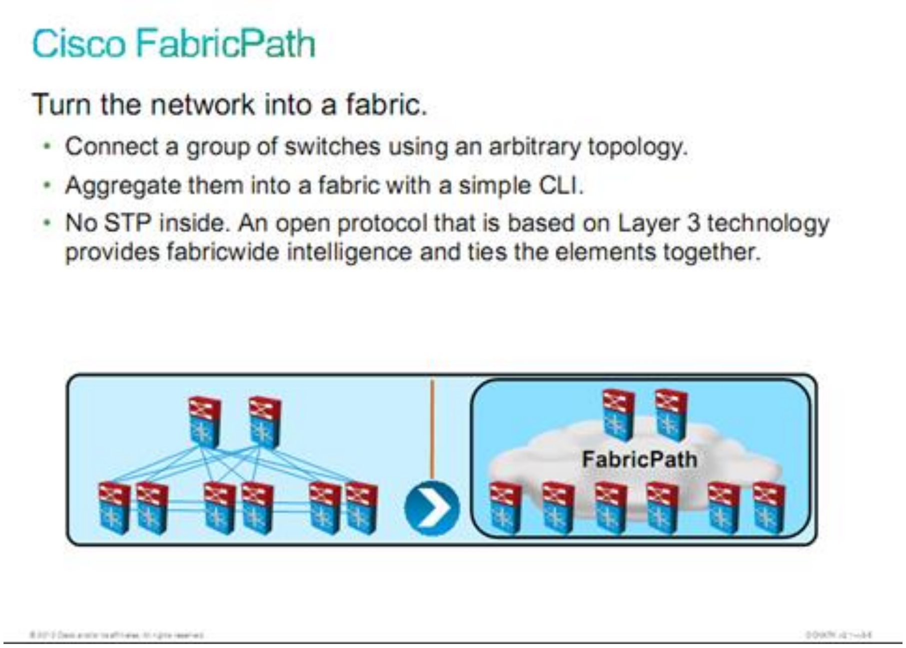
1. Both Peer-1 (forwarder) and Peer-2 (non-forwarder) join the SPT for new sources
2. Both Peer-1 and Peer-2 create the S,G state in their PIM routing tables
3. Only (in this case) Peer-1 adds interfaces to its OIF
4. In our example Peer-1 fails; Peer-2, the new forwarder, adds the vPC to its OIF
5. As Peer-2 is already receiving the multicast data down the SPT, there is no need for it to build a SPT toward the source

Implementing Multicast with FabricPath

A quick review of the function and benefits of FabricPath:

FabricPath provides Layer 2 routing, which allows multipath forwarding at Layer 2. As with vPCs, all links are in the forwarding state. A couple of differences between FabricPath and vPCs are:

- Spanning Tree Protocol is disabled where FabricPath has been enabled
- FabricPath scales to a larger size as more than two switches are supported in a FabricPath topology



Cisco FabricPath

Turn the network into a fabric.

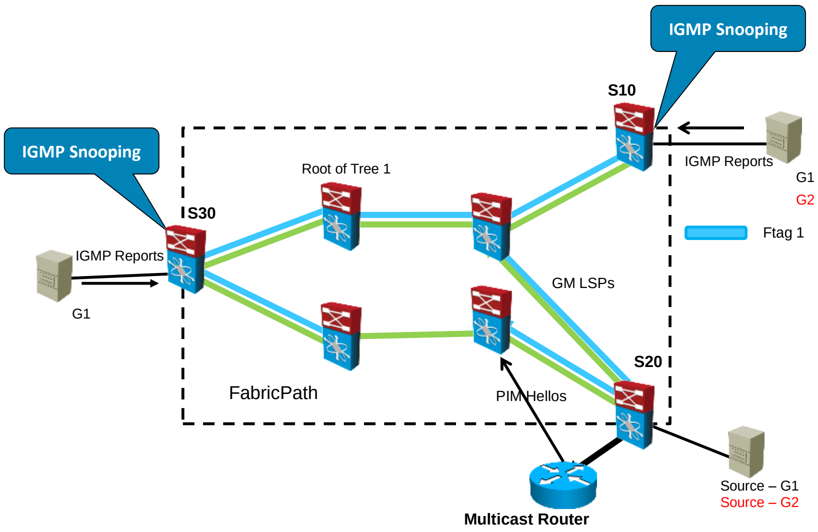
- Connect a group of switches using an arbitrary topology.
- Aggregate them into a fabric with a simple CLI.
- No STP inside. An open protocol that is based on Layer 3 technology provides fabricwide intelligence and ties the elements together.

[illegible]

IGMP Snooping with FabricPath

Below is a simple diagram with Multicast and one multi-destination topology, Topology 1. There are two multi-destination trees, topology 1 and topology 2.

With multicast routing in a FabricPath environment, traditional IGMP snooping on the Classic Ethernet (CE) edge ports is combined with Group Management Protocol Link State Packets (GMP-LSPs) inside FabricPath to advertise the edge switches interest in a multicast group.
In this example, S10 and S30 have received IGMP joins for Group1 and will flood GMP-LSPs into the Fabric. The



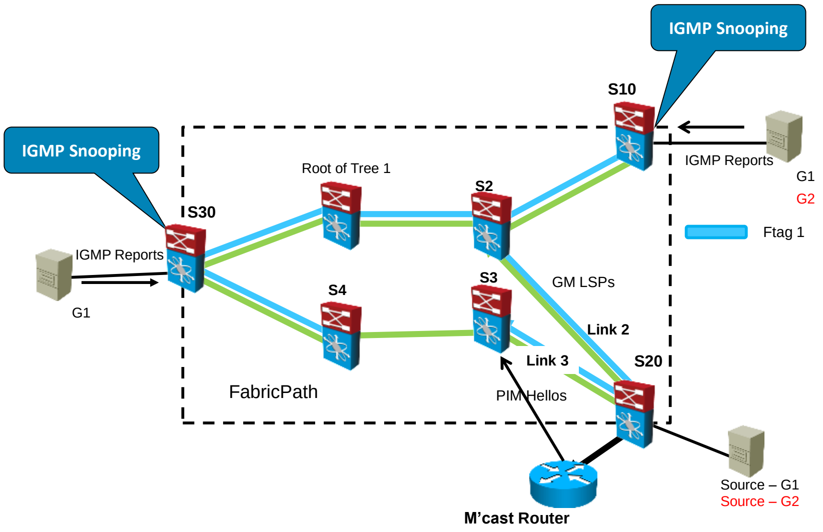
Multicast tree has been pruned between the two bottom switches within the FabricPath Core, ensuring no multicast loops.

1. IGMP snooping learns of interested receivers on FabricPath edge switches
2. Local membership tracked on CE ports based on receiving IGMP reports/leaves
3. Group membership is advertised in FabricPath IS-IS using GM-LSPs

GM-LSPs

1. Group Membership Link State Packets (LSPs) contain multicast forwarding information
2. Build Layer 2 multicast forwarding state for FabricPath core ports
3. IGMP snooping state created only at FabricPath edge switches
4. GM-LSPs are flooded to other FabricPath switches to advertise the edge switches that need
5. a particular multicast group

Multicast Data Plane with FabricPath



1. When the source sends multicast data to S20, S20 does a lookup in its mac address-table and sees that S10 and S30 want multicast Group 2 data.
2. It also sees in the mac address-table that it has two paths to get to those switches, L2 or L3.
3. S20 will forward the multicast data out both ports
4. S3 will drop the multicast data as the tree between itself and S4 has been pruned
5. S2 will also do a mac address lookup and see that the multicast data should be forwarded to S10 and also forwarded toward the Root of Tree 1
6. The Root of tree 1 does its own mac address-table lookup, and sees that the multicast data needs to be forwarded to S30
7. Lastly, S30 does the mac address-table lookup and sees that a local host wants the G1 multicast data and forwards the multicast out its CE port toward the receiver
8.

Conclusion

Multicast Sparse Mode and its derivatives are supported in the Nexus OS. As shown in this paper, it has been implemented in the Nexus platform to provide optimum performance in both virtual PortChannel and FabricPath environments.

Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge through training.

DCUFI - Implementing Cisco Data Center Unified Fabric v5.0

DCUFD - Designing Cisco Data Center Unified Fabric v5.0

DCNX5K - Implementing the Cisco Nexus 5000 and 2000 v2.0

DCNX7K - Configuring Cisco Nexus 7000 Switches v3.0

ICMI - Implementing A Cisco Multicast Infrastructure

Visit **www.globalknowledge.com** or call **1-800-COURSES (1-800-268-7737)** to speak with a Global Knowledge training advisor.

About the Author

Carol Kavalla has been teaching about Cisco technology and its products for more than 14 years. She owns a consulting company and specializes in routing, switching, and data center implementations.