TOMORROW starts here.
Cisco Dynamic Fabric Automation Architecture

Miroslav Brzek, Systems Engineer
mibrzek@cisco.com
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
The Data Center Fabric Journey

STP

MC-LAG
vPC

L2MP
FabricPath

DFA

MAN/WAN
Cisco Dynamic Fabric Automation Architecture

Innovative Building Blocks

Fabric Management

Workload Automation

Optimized Network

Virtual Fabrics
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
    - Control Plane
    - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Cisco Dynamic Fabric Automation

Scale, Resiliency and Efficiency

Advantages

- Any subnet, anywhere, rapidly
- Reduced Failure Domains
- Extensible Scale & Resiliency
- Profile Controlled Configuration

✦ Full Bi-Sectional Bandwidth (N Spines)
✦ Any/All Leaf Distributed Default Gateways
✦ Any/All Subnets on Any Leaf

Network Config Profile

Network Services Policies

# show port-profile name WebProfile

<table>
<thead>
<tr>
<th>port-profile WebServer-PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>description:</td>
</tr>
<tr>
<td>status: enabled</td>
</tr>
<tr>
<td>system vlans</td>
</tr>
<tr>
<td>port-group: WebServers</td>
</tr>
<tr>
<td>config attributes:</td>
</tr>
<tr>
<td>switchport mode access</td>
</tr>
<tr>
<td>switchport access vlan</td>
</tr>
<tr>
<td>no shutdown</td>
</tr>
<tr>
<td>security-profile Protected-Web-Srv</td>
</tr>
</tbody>
</table>

evaluated config attributes:

<table>
<thead>
<tr>
<th>switchport mode access</th>
</tr>
</thead>
<tbody>
<tr>
<td>switchport access vlan</td>
</tr>
<tr>
<td>no shutdown</td>
</tr>
</tbody>
</table>

assigned interfaces:

Veth10
Cisco Dynamic Fabric Automation

Flexible Topologies

- Traditional Aggregation/Access
- Folded CLOS
- Three Tiers / Fat Tree
- Full Mesh
Cisco Dynamic Fabric Automation Architecture

**Device Roles**

Virtual Machines
Physical Machines
FEXes
3rd Party Switches
UCS FIs
Blade Switches
Storage

Firewalls
Load Balancers
3rd Party Appliances

Routers
Switches
3rd Party Devices

- Spine
- Leaf
- Border Leaf
- Services Leaf

Virtual Leaf*

*Virtual Leaf: N1KV/OVS being a “light” participant on the control plane protocol (supporting VDP)

**Note:** the different Leaf roles are logical and not physical. The same Leaf Switch could perform all three functions (Regular, Services and Border Leaf)
Cisco Dynamic Fabric Automation

Fabric Properties

- High Bi-Sectional Bandwidth
- Wide ECMP: Unicast or Multicast
- Uniform Reachability, Deterministic Latency
- High Redundancy: Node/Link Failure
- Line rate, low latency, for all traffic

Fabric properties applicable to all topologies
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Fabric Control Plane

IS-IS as Fabric Control Plane

**IS-IS for fabric link state distribution**
- Fabric node reachability for overlay encapsulation
- Building multi-destination trees for multicast and broadcast traffic
- Quick reaction to fabric link/node failure (Layer-2 BFD)
- Enhanced for mesh topologies

**Fabric Control Protocol doesn't distribute**
- Host Routes
- Host originated control traffic
- Server subnet information
Fabric Control Plane

Host and Subnet Route Distribution

- Host Route Distribution decoupled from the Fabric link state protocol
- Use MP-BGP on the leaf nodes to distribute internal host/subnet routes and external reachability information
- MP-BGP enhancements to carry up to 1.2 Million routes and reduce convergence time

**Note:** Route-Reflectors deployed for scaling purposes
Fabric Control Plane

Host Originated Protocols Containment

- ARP, ND, IGMP, DHCP originated on servers are terminated on Leaf nodes
- Contain floods and failure domains, distribute control packet processing
- Terminate PIM, OSPF, eBGP from external networks on Border Leafs
Fabric Control Plane

Host Detection and Deletion

- In order to advertise host reachability information, a leaf must first discover locally connected devices.

- Detection of local hosts
  - Based on VDP* or ARP/ND/DHCP
  - *VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41

- Detection of remote hosts
  - Received MP-BGP notifications

Note: Discovered IP address information from ARP/ND-Table get redistributed into MP-BGP Control Plane for End-Host reachability.
Fabric Control Plane

Host Detection and Deletion (Detail on Data Plane Trigger)

- Data packet from Server reaches Leaf
- Leaf detects new MAC learn event
- VLAN detected based on:
  - IEEE 802.1q tag used between Server and Leaf
  - VLAN configured on Leaf port
- Based on learned VLAN and Leaf local configuration parameters, logical configuration get pulled, instantiated and applied on Leaf

```
interface eth1/1
switchport mode trunk
switchport trunk native vlan 10

interface eth1/32
switchport mode trunk

switch# show mac-address-table

<table>
<thead>
<tr>
<th>VLAN</th>
<th>MAC Address</th>
<th>Type</th>
<th>Age</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0018.b967.3cd0</td>
<td>dynamic</td>
<td>10</td>
<td>Eth1/1</td>
</tr>
<tr>
<td>11</td>
<td>001c.b05a.5380</td>
<td>dynamic</td>
<td>200</td>
<td>Eth1/32</td>
</tr>
</tbody>
</table>
```

ARP/ND/DHCP
Fabric Control Plane

Host Detection and Deletion (Detail on VDP* Trigger)

- VDP* session gets established between virtual Leaf and physical Leaf
- Segment ID (VNI) gets sent from virtual Leaf based on Virtual Machine configuration
- Physical Leaf responds with next available VLAN defined in Pool (system fabric dynamic-vlans)
- Based on learned Segment ID (VNI), logical configuration get pulled, instantiated and applied on Leaf

```
switch# show evb host
```

<table>
<thead>
<tr>
<th>Host Name</th>
<th>VNI</th>
<th>Vlan</th>
<th>BD</th>
<th>Mac-address</th>
<th>IP-Address</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server</td>
<td>31000</td>
<td>3000</td>
<td>3000</td>
<td>0050.56ac.1f71</td>
<td>192.168.131.103</td>
<td>Eth1/21</td>
</tr>
</tbody>
</table>

*N1KV/OVS

VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Optimized Network

Distributed Gateway at the Leaf

- Any Subnet anywhere => Any Leaf can instantiate ANY Subnet
  - All Leafs share gateway IP and MAC for a Subnet (No HSRP)
  - ARPs are terminated on Leaf, No Flooding beyond Leaf

- Facilitates VM Mobility, workload distribution, arbitrary clustering

- Seamless Layer-2 or Layer-3 communication between physical hosts and virtual machines
Optimized Networking

Distributed Gateway Mode

- Distributed Gateway exists on all Leafs where VLAN/Segment-ID is active
- There are different Forwarding Modes for the Distributed Gateway:
  - Proxy-Gateway (Enhanced Forwarding)
    - Leverages local proxy-ARP
    - Intra and Inter-Subnet forwarding based on Routing
    - Contain floods and failure domains to the Leaf
  - Anycast-Gateway (Traditional Forwarding)
    - Intra-Subnet forwarding based on Bridging
    - Data-plane based conversational learning for endpoints MAC addresses
    - ARP is flooded across the fabric
Optimized Networking

IP Forwarding within the Same Subnet

1. H1 sends an ARP request for H3 – 10.10.10.30
2. The ARP request is intercepted at Leaf1 (L1) and punted to the Supervisor
3. Assuming a valid route to H3 does exist in the Unicast RIB, Leaf1 (L1) sends the ARP reply with G_MAC so that H1 can build its ARP cache

**Note:** the ARP request is NOT flooded across the Fabric nor out of other local interfaces belonging to the same Layer-2 domain
Optimized Networking

IP Forwarding within the Same Subnet (2)

4. H1 generates a data packet with G_MAC as destination MAC

5. Leaf1 (L1) receives the packet and performs Layer-3 lookup for the destination

6. Leaf1 (L1) adds the Layer-2 and the FabricPath headers and forwards the FabricPath frame across the Fabric, picking one of the equal cost paths available via the multiple Spines

7. Leaf4 (L4) receives the packet, strips off the FabricPath header and performs Layer-3 lookup and forwarding toward H3
Optimized Networking

IP Forwarding Across Different Subnet

1. H1 sends an ARP request for Default Gateway – 10.10.10.1
2. The ARP request is intercepted at the Leaf1 (L1) and punted to the Supervisor
3. Leaf1 (L1) acts as a regular Default Gateway and sends ARP reply with G_MAC
Optimized Networking

IP Forwarding Across Different Subnet (2)

4. H1 generates a data packet destined to H2 IP with G_MAC as destination MAC

5. Leaf1 (L1) receives the packet and performs Layer-3 lookup for the destination

6. If valid routing information for H2 is available in the unicast routing table, Leaf1 (L1) adds the Layer-2 and the FabricPath headers and forwards the FabricPath frame across the Fabric, picking one of the equal cost paths available via the multiple Spines

7. Leaf4 (L4) receives the packet, strips off the FabricPath header and performs Layer-3 lookup and forwarding toward H2
Optimized Networking

Introducing Layer-3 Conversational Learning

- Use of /32 host routes may lead to scaling issues if all the routes are installed in the Hardware tables of all Leaf nodes
  - Layer-3 conversational learning is introduced to alleviate this concern
  - Disabled by default -> all host routes are programmed in the Hardware

- With Layer-3 conversational learning, host routes for remote endpoints will be programmed into the Hardware FIB (from the Software RIB) upon detection of an active conversation from a local endpoint

Default Behavior (No Layer-3 Conversational Learning)

After Enabling Layer-3 Conversational Learning
# Encapsulation and Forwarding

## What about Encapsulation?

<table>
<thead>
<tr>
<th>Feature</th>
<th>DFA + FabricPath</th>
<th>DFA + VXLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-Host Reachability and Distribution</td>
<td>Multi-Protocol BGP</td>
<td>Multi-Protocol BGP</td>
</tr>
<tr>
<td>End-Host Detection</td>
<td>ARP/ND &amp; VDP</td>
<td>ARP/ND &amp; VDP</td>
</tr>
<tr>
<td>Fabric Multicast Control-Protocol</td>
<td>FabricPath IS-IS</td>
<td>PIM Bi-Dir</td>
</tr>
<tr>
<td>Fabric Topology Control-Plane</td>
<td>FabricPath IS-IS</td>
<td>Layer-3 IS-IS</td>
</tr>
<tr>
<td>Leaf IP/MAC binding distribution within Fabric</td>
<td>FabricPath IS-IS</td>
<td>Multi-Protocol BGP</td>
</tr>
</tbody>
</table>
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Cisco Dynamic Fabric Automation

Virtual Fabrics for Public or Private Cloud Environments

**Advantages**

- Any workload, any Virtual Fabric, rapidly
- Scalable Secure Virtual Fabrics
- Virtual Fabric Tenant Visibility
- Routing/Switching Segmentation
Virtual Fabrics

Introducing Segment-ID Support

- Traditionally VLAN space is expressed over 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a Data Center to 4096 VLANs
Virtual Fabrics

Introducing Segment-ID Support

- Traditionally VLAN space is expressed over 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a Data Center to 4096 VLANs

- DFA leverages a double 802.1Q tag for a total address space of 24 bits
  - Support of ~16M segments

- Segment-ID is hardware-based innovation offered by Leaf and Spine nodes part of the DFA Fabric
Virtual Fabrics

802.1Q Tagged Traffic to Segment-ID Mapping

- Segment-IDs are utilized for providing isolation at Layer-2 and Layer-3 across the Integrated Fabric.
- 802.1Q tagged frames received at the Leaf nodes from edge devices must be mapped to specific Segments.

The VLAN-Segment mapping can be performed on a leaf device level:
- VLANs become locally significant on the leaf node and 1:1 mapped to a Segment-ID.

- Segment-IDs are globally significant, VLAN IDs are locally significant.
Virtual Fabrics

How are Segment-IDs Utilized?

- Each IP Subnet, defined at the edge of the Fabric is associated to a Layer-2 domain, which is represented by a Segment-ID.
- A Segment-ID will be used to uniquely identify a VRF within the Fabric.
- Multiple Layer-2 domains can be defined for a given Tenant and are mapped to a Layer-3 VRF.
Virtual Fabrics

Fabric Routed Flows

1. H1 sends a packet to H3 traffic between the Server and the Leaf is tagged with a local VLAN-ID 123

2. Layer-3 lookup is performed by Leaf1 (L1) in the context of the BLUE VRF

3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet into the fabric. The Segment-ID identifying the BLUE VRF is added inside the Layer-2 header

4. Leaf4 (L4) receives the frame and associates it to the BLUE VRF by looking at the Segment-ID value. It then sends it to H3 using a local VLAN-ID 123

Note: this behavior applies to all Fabric routed flows (intra-subnet or inter-subnet)
Virtual Fabrics

Fabric Routed Flows

1. H1 sends a packet to H3 traffic between the Server and the Leaf is tagged with a local VLAN-ID 123

2. Layer-3 lookup is performed by Leaf1 (L1) in the context of the BLUE VRF

3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet into the fabric. The Segment-ID identifying the BLUE VRF is added inside the Layer-2 header

4. Leaf4 (L4) receives the frame and associates it to the BLUE VRF by looking at the Segment-ID value. It then sends it to H3 using a local VLAN-ID 555

Note: this behavior applies to all Fabric routed flows (intra-subnet or inter-subnet)
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Advantages

- Device Auto-Configuration
- Cabling Plan Consistency Check
- Automated Network Provisioning
- Common point of fabric access
- Tenant, Virtual Fabric & Host Visibility
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Workload Automation & Open Environment

Advantages

- Any workload, Anywhere, Anytime
- Open Integration: Orchestration
- Automated Scalable Provisioning
- Workload aware fabric
Workload Automation & Open Environment

Compute & Storage Orchestration

Network & Services Orchestration

Compute/Virtualization Controllers

Fabric Management

e.g. UCS Manager, vCenter, SCVMM, etc.

Here is the Glue!

*VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41
Workload Automation

Leveraging VDP for Leaf Auto-Configuration

1. Create Logical Org Network
2. Communicate Org Network to Fabric

Fabric Management
Segment-ID, IP information (GWY, Mask, Org, etc.)
Orchestrator
Network Services Controller
Segment Information Download (Push)

Create Logical Org Network
Communicate Org Network to Fabric
Leveraging VDP for Leaf Auto-Configuration (2)

1. Create Logical Org Network
2. Communicate Org Network to Fabric
3. Query the DCNM Network DB (Segment-ID as key)
4. Instantiate Blue Network
5. Segment-ID from the Virtual-Switch

Fabric Management

Orchestrator

Network Services Controller

Tenants

N1KV/OVS

New VM gets created in Blue Network
Workload Automation

Leveraging VDP for Leaf Auto-Configuration (3)

1. Create Logical Org Network
2. Communicate Org Network to Fabric
3. Instantiate Blue Network
4. New VM gets created in Blue Network
5. VDP information exchange

Fabric Management

Orchestrator

Network Services Controller

Tenants

Orchestrator

Related Components:
- N1KV/OVS
- VLAN-ID to the Virtual-Switch
- SVI, VRF Creation
- Configuration Download (Pull)
Workload Automation

Leveraging VDP for Leaf Auto-Configuration (4)

Leaf receives 802.1q tagged frames and associates them to the Segment-ID

1. Create Logical Org Network
2. Communicate Org Network to Fabric
3. New VM gets created in Blue Network
4. VDP information exchange
5. Instantiate Blue Network
Agenda

- DFA Overview
- Optimized Networking
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support
Cisco Dynamic Fabric Automation

Platform Support

Compute & Storage Orchestration

UCS Director

vmware

openstack

Cisco Prime DCNM

Cisco Prime NSC
Cisco Dynamic Fabric Automation

Platform Support

Compute & Storage Orchestration

- UCS Director
  - VMware
  - OpenStack

- Cisco Prime DCNM
- Cisco Prime NSC

- Nexus 7x00 (F2/F2e/F3)
- Nexus 6000
- Nexus 5500
- Nexus 5600
- Nexus 6000
- Nexus 2000 (FEX)
- Nexus 1000V/OVS
- Nexus 7x00 (F3)

- VXLAN
- FabricPath

- NX-OS 7.1 (roadmap)
- Shipping
Cisco Dynamic Fabric Automation

Enabling the Unified Fabric with Automation, Optimization and Simplification using

- Orchestration Integration with UCS Director, Openstack & vCD for Workload Automation
- Optimized Networking for Extensible Resiliency and for Workload distribution and VM Mobility
- Secure Virtual Fabrics
- CPoM for Multi-Device Configuration, Automated Network Provisioning and Network & Host Visibility
Prosíme, ohodnoťte tuto přednášku

Děkujeme