Cisco live!
What You Make Possible
Cisco Nexus 6000 Architecture

Sina Mirtorabi
Technical Marketing Engineer
Session Abstract

• **Session ID:** BRKARC-3453
• **Title:** Cisco Nexus 6000 Architecture
• **Abstract:** This session describes the architecture of the Nexus 6000, a revolutionary switch for the datacenter. The nexus 6000 provides high density 40 Gig ports with low latency, highly efficient unicast and multicast forwarding, and new troubleshooting tools.
Agenda

• Nexus 6000 overview
• Architecture detail
• SPAN
• Buffering and QoS
• Multicast
• Summary
Nexus 6000 Overview
New! Nexus 6000 Series

Cisco Nexus® 6000 Series

- Nexus 6004
  - 96 port 40G 4RU Switch

- Nexus 6001
  - 48 port 10G, 4 port 40G, 1RU Switch

Cisco Nexus® 5000 Series

- Nexus 5020
  - 56-Port 2RU Switch

- Nexus 5548
  - 48-Port 1RU Switch

- Nexus 5596
  - 96-Port 2RU Switch

- Nexus 5596T
  - 10Gbase-T switch

- 4-port QSFP+ GEM

New products available:
- Nexus 5010
  - 28-Port 1RU Switch

- Nexus 5548
  - 48-Port 1RU Switch

- Nexus 5596
  - 96-Port 2RU Switch

- Nexus 5596T
  - 10Gbase-T switch

- 4-port QSFP+ GEM

New! Nexus 6000 Series
Introducing Nexus 6000

High Performance
- Line rate for L2 and L3 for all packet size
- Line rate SPAN
- 1 us port to port latency for all frame size
- Cut through switching at 10Gig and 40Gig
- FCoE at 40 Gig
- 25M buffer for 3 QSFP ports
- 31 active SPAN

High Scalability
- Nexus 6004: 96 x 40Gig or 384 x 10Gig
- Nexus 6001: 48 x 10Gig and 4 x 40Gig
- Up to 256K MAC
- Up to 128K ARP
- 32K LPM routes
- 16K Bridge domain

Feature Rich
- L2 and L3 features
- vPC
- Fabric Path / TRILL
- Segment-ID
- Adapter FEX / VM FEX
- NAT, Tunneling

Visibility / Analytic
- Sampled Netflow
- Buffer monitoring
- Latency monitoring
- SPAN on drop
- SPAN on high latency
- Microburst monitoring
Nexus 7000 and Nexus 6004: DC Considerations

Customer Requirements: Decision Points

- Virtualization
- Scalability
- DCI/Mobility
- Environmentals
- L4-7 Services
- High Availability
- Latency
- Investment Protection

Decision Criteria in the Aggregation

Lead Platform: Modular, High-End Solution

Recommended when:
- Scale and Flexibility (100M/1G/10G/40G/100G/UP*)
- Highest Availability (HA)
- Investment Protection
- Multi-Protocol / Services VDC / OTV / MPLS / VPLS / LISP

Nexus 7000 Series
Up to 36 Tbps

Fixed, Mid-Range Solution

Recommended when:
- High density compact 10G/40G/100G*/UP*
- Low footprint & low power
- Low latency & jitter
- Extensive FEX Features

Nexus 6000 Series
Up to 7.68 Tbps

* Roadmap
Nexus 6000 Deployment Scenarios

**Compact Aggregation**
- Nexus 7000
- Nexus 7000
- Nexus 6004
- Nexus 6004
- Nexus 6004/Nexus 6001
- Nexus 2000

**Large Scale Fabric (Layer 2 and Layer 3)**
- Nexus 7000
- Nexus 6004
- Nexus 6004/Nexus 6001

**High Performance Computing (HFT)**
- Nexus 6004
- Nexus 3548
- Nexus 6001/6004

**Multi hop FCoE**
- Nexus 7000
- Nexus 6004
- Nexus 6001

Brkarc-3453 © 2012 Cisco and/or its affiliates. All rights reserved.
Nexus 6004 Chassis
Rear View

- 48 Fixed QSFP Interfaces
- 12 QSFP ports
- Line card Expansion Module (LEM)
Nexus 6004 Chassis

Front View

- Minimum 3 PS and 3 FAN are required

Power Supply 3+3 grid redundancy or 3+1 redundancy

Fan Module 3+1 redundancy

Console Mgmt0 USB
Nexus 6001 Chassis

Rear View

- 1RU
- 30”

- 48 Fixed SFP+ Interfaces
- 4 Fixed QSFP interface
Nexus 6001 Chassis

Front View

- Power Supply 1+1 redundancy
- Fan Module 2+1 redundancy
- Console Mgmt0 USB
Nexus 6000 Airflow

- Front to back or back to front air flow
- Port side exhaust at FCS
- Port side intake (Reversed airflow)

Nexus 6004 with 6.0(2)N1(2) release-Q2CY13
Nexus 6001 with 6.0(2)N2(1) release-Q3CY13

- Different PS and Fan modules are required for different air flow directions
Nexus 6004 Physical Connections

- **QSFP-SR4**: 100m over OM3 MMF, 150m over OM4 MMF
- **QSFP-CSR4**: 300m over OM3 MMF, 400m over OM4 MMF
Nexus 6004 Supported Transceivers

<table>
<thead>
<tr>
<th>Type</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>QSFP-40G-LR4</td>
<td>10KM with SMF</td>
</tr>
<tr>
<td>FET-40G</td>
<td>100m with OM3</td>
</tr>
<tr>
<td>QSFP-40G-SR4</td>
<td>100m with OM3</td>
</tr>
<tr>
<td></td>
<td>150m with OM4</td>
</tr>
<tr>
<td>QSFP-40G-CSR4</td>
<td>300m with OM3</td>
</tr>
<tr>
<td></td>
<td>400m with OM4</td>
</tr>
<tr>
<td>QSFP-4x10G-AC7M</td>
<td>7m</td>
</tr>
<tr>
<td>QSFP-4x10G-AC10M</td>
<td>10m</td>
</tr>
<tr>
<td>QSFP-H40G-1M</td>
<td>1m</td>
</tr>
<tr>
<td>QSFP-H40G-3M</td>
<td>3m</td>
</tr>
<tr>
<td>QSFP-H40G-5M</td>
<td>5m</td>
</tr>
<tr>
<td>QSFP-H40G-AC7M</td>
<td>7m</td>
</tr>
<tr>
<td>QSFP-H40G-AC10M</td>
<td>10m</td>
</tr>
</tbody>
</table>
FET-40G (Q3CY13)

- Low cost QSFP optical transceiver connecting FEX to N6004
- Supported on N6004 and Nexus 2248PQ-10G
- Interoperable with FET-10G
- Support 100m distance with OM3
Unified Port LEM

- Unified ports expansion module
- 2/4/8/FC port
- 1/10GE and FCoE
- For Nexus 6004 only
- SFP+ port allows support for more variety of optical transceiver
100 Gig LEM

- 100G unified ports expansion module
- 100GE support including FCOE
- For Nexus 6004 only
- Not CC’d and EC’d. Form factor and features under consideration
Port Speed Configuration

• By default ports are in 40GE mode
• Port speed can be changed at group of 3 QSFP ports.
• The group of 12 QSFP ports need to be reset after port speed change.

```
N6004(config)# interface breakout slot 8 port 1-12 map 10g-4x ?
<CR>
```
Nexus 6000 Key Forwarding Tables

- **Host Table**: 256K entry hashing table. Actual capacity is slightly below 256K.
  - Host Table is shared between MAC, ARP/ND and /32 host route and (*,G)
  - Host Table default carving: 128K MAC, 128K IP host
- **LPM Table (32K)**
  - Summary routes
- **Mroute Table (64K)**
  - (S,G)

- **MAC Region**: (128K by default)
- **IP Host Region (ARP/ND/host route)**: (*,G)
  - (128K by default)
Nexus 6000 Unicast Table Scaling

- Each IPv6 ND consumes 2 entries in IP host region in host table
- Each IPv6 route consumes 4 entries in LPM table

**IPv4:** 10.1.1.1  
**IPv6:** 2001::0DB8:800:200C:417A

```
10.1.1.0/24  
2001::/64
```

**Host Table**
- MAC Region: 00:02:B3:01:02:03
- IP Host Region: 10.1.1.1
- ARP: 2001:0:0:0:
- 0DB8:800:200C:417A

**LPM Table**
- IPv4 route: 10.1.1.0
- 4 HW entries for IPv6 route
- 2 HW entries for IPv6 ND
## Nexus 6000 Host Table Scaling

<table>
<thead>
<tr>
<th>Deploy Scenario</th>
<th>Scalability</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 switch</td>
<td>256K MAC</td>
</tr>
<tr>
<td>L2L3 Gateway with IPv4 only*</td>
<td>128K hosts</td>
</tr>
<tr>
<td>L2L3 Gateway with IPv6 only*</td>
<td>85K hosts</td>
</tr>
<tr>
<td>L2L3 Gateway with dual stack*</td>
<td>50K hosts</td>
</tr>
</tbody>
</table>

- In L2 mode the IGMP snooping is stored at IP host region. So the actual MAC region will be less than 256K. At FCS software statically allocate 128K entries for MAC and 128K entries for IP host(ARP/ND/host route).
- Host table is hashing table. Actual capacity will be slightly below the number in the table.

* Assume one IPv4 or IPv6 per host. Hardware scaling number.
Nexus 6004 Control Plane

- CPU - 1.66 GHz Intel LV Xeon
- DRAM - 2 GB of DDR2 400 (PC2 3200) in two DIMM slots
- On-Board Fault Log - 64 MB of flash for failure analysis
- 1G Flash
- NVRAM - 2 MB of SRAM: Syslog and licensing information

- CPU - 1.7 GHz Intel Jasper Forest (Dual Core)
- DRAM - 8 GB of DDR3 in two DIMM slots
- On-Board Fault Log (OBFL) - 64 MB
- 2G Flash
- NVRAM - 6 MB of SRAM to store Syslog and licensing information

- Built-in single supervisor
- ISSU with L2 at FCS
- CPU – 4 Core Intel Gladden 2.0GHz
- DRAM - 4 DIMM slots 16GB by default.
- 8G Flash for NX-OS software and user files
- 6MB NVRAM 64MB OBLF(On-board Fault Logging)
## Nexus 6000 versus Nexus 5500

### HW Performance and Scalability

<table>
<thead>
<tr>
<th></th>
<th>Nexus 55xx</th>
<th>Nexus 6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>~1.8us</td>
<td>~1us</td>
</tr>
<tr>
<td>MAC table</td>
<td>32K</td>
<td>256K MAC/ARP (flexible)</td>
</tr>
<tr>
<td>LPM Routes</td>
<td>16K</td>
<td>32K</td>
</tr>
<tr>
<td>Hosts</td>
<td>16K</td>
<td>128K</td>
</tr>
<tr>
<td>Multicast route</td>
<td>8K</td>
<td>32K</td>
</tr>
<tr>
<td>Bridge Domains</td>
<td>4K</td>
<td>16K</td>
</tr>
<tr>
<td>ACLs</td>
<td>4K flexible</td>
<td>4K flexible</td>
</tr>
<tr>
<td>IGMP Snooping groups</td>
<td>4K</td>
<td>32K</td>
</tr>
<tr>
<td>ECMP</td>
<td>64 way</td>
<td>1K</td>
</tr>
<tr>
<td>VRFs</td>
<td>1K</td>
<td>4K</td>
</tr>
<tr>
<td>SPAN</td>
<td>4</td>
<td>31, 16 can be ERSPAN</td>
</tr>
<tr>
<td>Buffer</td>
<td>640K per port dedicated</td>
<td>25MB per 3 QSFP</td>
</tr>
<tr>
<td>VDC</td>
<td>N/A</td>
<td>8</td>
</tr>
</tbody>
</table>
Nexus 6004 Architecture
Nexus 6004 Architecture

Switch Fabric

UPC

UPC

UPC

UPC
Nexus 6004 Unified Port Controller (UPC)

- Multimode MAC
- Packet Parser/Editing
- LU Engine
- Buffering
- Queuing
Nexus 6004 Switch Fabric

10 Gig Fabric Mode

Switch Fabric
192 x 384

14 Gig

40 Gig Fabric Mode

Switch Fabric
48 x 96

4 x 14 Gig

Scheduler

14 Gig

Scheduler

4 x 14 Gig

14 Gig

4 x 14 Gig
Nexus 6004 Internal Architecture

- Fabric Mode 10 Gig
Nexus 6004 Internal Architecture

- Fabric Mode 40 Gig

**40 Gig Fabric Mode**

- Ingress UPC
  - 3 x 40GE
  - 4 x 56 Gbps = 224 Gbps

- Egress UPC
  - 8 x 56 Gbps = 448 Gbps

- Switch Fabric 48 x 96
  - 56 Gbps

- Switch Fabric 48 x 96
  - 56 Gbps

- Switch Fabric 48 x 96
  - 56 Gbps

- Switch Fabric 48 x 96
  - 56 Gbps
Nexus 6004 Internal Architecture

- Fabric Mode 10 Gig

Ingress UPC 1

4 x 14Gig

10G mode (192x384)

8 x 14Gig

Egress UPC 1

10G mode (192x384)

10G mode (192x384)

10G mode (192x384)

Ingress UPC 32

Egress UPC 32

Nexus 6004 Internal Architecture

• Fabric Mode 10 Gig
Nexus 6004 Internal Architecture

• Fabric Mode 40 Gig

Ingress UPC 1

56Gig

40G mode (48x96)

2 x 56Gig

Egress UPC 1

Ingress UPC 32

40G mode (48x96)

40G mode (48x96)

40G mode (48x96)

Egress UPC 32

56Gig

2 x 56Gig

Egress UPC 1
Nexus 6004 Unified Port Controller (UPC)

- Multimode MAC:
  - Packet protocol function: encoding/decoding for physical medium dependent
  - Flow control, PFC, 802.3x, B2B credit for Fibre Channel

- Packet Parser/Editing:
  - Parse the packet, extract relevant fields, form a lookup vector and pass it to LU
  - Create SPAN copy if needed
  - Put 1588 clock

- LU Engine and ACL:
  - Perform all the forwarding decision based on the vector received from FW
  - AC module performs ACL, QoS classification, FC zoning, NAT, PBR

- Buffering:
  - Packet buffering on ingress and egress
  - Pass queuing information to ingress QS
  - Transmit frame to Fabric Interface
  - Perform DWRR based on output queues

- Queuing:
  - Maintains ingress VoQs for unicast, multicast and SPAN
  - Maintains per port egress queues for unicast and multicast
  - Performs egress replications
Nexus 6004 UPC block (animation)
Cut Through and S&F

- Depending on the port speed combination and switch fabric mode, Nexus 6004 performs either cut through switching or store& forward switching.
- We can summarize the cut through switching as follow.
- In 10 Gig fabric mode, we do cut through switching when the egress is 10Gig.

<table>
<thead>
<tr>
<th>Ingress</th>
<th>Egress</th>
<th>10GE</th>
<th>40GE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GE</td>
<td>10GE</td>
<td>Cut-through</td>
<td>Store-N-Forwarding</td>
</tr>
<tr>
<td>40GE</td>
<td>40GE</td>
<td>Cut-through</td>
<td>Store-N-Forwarding</td>
</tr>
</tbody>
</table>

- In 40 Gig fabric mode, we do cut through switching when the ingress is 40 Gig.

<table>
<thead>
<tr>
<th>Ingress</th>
<th>Egress</th>
<th>10GE</th>
<th>40GE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GE</td>
<td>10GE</td>
<td>Cut-through OR Store--Forwarding</td>
<td>Store-N-Forwarding</td>
</tr>
<tr>
<td>40GE</td>
<td>40GE</td>
<td>Cut-through</td>
<td>Cut-through</td>
</tr>
</tbody>
</table>
Fabric Mode, 40Gig or 10Gig?

• If all ports are operating at 10Gig, use the 10Gig fabric mode.
• If all ports are operating at 40Gig, use 40 Gig fabric mode.
• A change to the fabric mode requires a reload.
• If there is a mix of 10Gig and 40Gig ports, use the cut through / S&F matrix in previous slides to see what traffic needs low latency, the following also needs to be considered
  – In 10Gig fabric mode, a 40GE interface can only carry 10Gbps flows
  – In 10Gig fabric mode, there could be a latency improvement of 200 ns for 10Gig ports
  – At FCS, ISSU is disabled when fabric mode is 10Gig mode. Post FCS, ISSU will be enabled in 10Gig fabric mode
Nexus 6000 SPAN
Nexus 6004 SPAN Enhancements

- Total 31 active SPAN sessions. 16 ERSPAN sessions.
- Support ERSPAN termination
- Wire speed SPAN throughput. Extra fabric link bandwidth for SPAN traffic
- Best effort for SPAN traffic. Drop SPAN traffic in case of fabric link congestion
- Hardware support multiple SPAN destination ports per SPAN session
- Support PortChannel as SPAN destination port. Source port based hashing.
Data Protection with SPAN

- SPAN packets are dropped first when there is congestion
- For rx SPAN data packets are replicated at ingress UPC. For tx SPAN data packets are replicated at egress UPC
- UPC gives priority to data traffic and schedule data traffic first with remaining fabric link bandwidth allocate for SPAN.
- Separate buffer pool for SPAN traffic to avoid buffer hog due to SPAN destination port congestion
Line Rate SPAN

- Extra bandwidth capacity is built in for SPAN traffic
- Any fabric bandwidth not taken by data traffic can be used for SPAN traffic
- Can achieve line rate SPAN throughput with extra fabric link bandwidth
- First product in the market support 16 line rate SPAN sessions

*Note: The internal header is not taken into consideration.*
SPAN with Multiple Destination Ports

- Hardware support multiple SPAN destination ports per session
- Each destination port is counted as one SPAN session
- Packets will be replicated multiple times, one per each SPAN destination port
# Nexus 6004 vs Nexus 5500 SPAN

<table>
<thead>
<tr>
<th>SPAN Features</th>
<th>Nexus 6000</th>
<th>Nexus 5500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total SPAN sessions</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>Local SPAN sessions</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>ERSPAN sessions</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Prioritize data over SPAN</td>
<td>Yes (through scheduling)</td>
<td>SPAN policing</td>
</tr>
<tr>
<td>Line rate SPAN throughput</td>
<td>Yes</td>
<td>Yes for limited scenarios</td>
</tr>
<tr>
<td>ERSPAN destination session</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Truncated SPAN/ERSPAN</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ACL based SPAN/ERSPAN</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SPAN on drop</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SPAN on high latency</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SPAN with multiple destination ports</td>
<td>Yes (each destination port burns one SPAN session)</td>
<td>Yes (each destination port burns one SPAN session)</td>
</tr>
</tbody>
</table>
Nexus 6000 Buffering & QoS
Nexus 6004 Ingress Buffer Management

Ingress Buffer

- Shared
- Dedicated

Ingress UPC

- Default class
  - Control
  - SPAN
  - Default
Nexus 6004 Ingress Buffer Allocation

### 40 Gig Port

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dedicated per 40Gig Port</td>
</tr>
<tr>
<td>Control Plane</td>
<td>67 KB</td>
</tr>
<tr>
<td>SPAN</td>
<td>152 KB</td>
</tr>
<tr>
<td>Default Class</td>
<td>100 KB</td>
</tr>
</tbody>
</table>

### 10 Gig Port

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dedicated per 10Gig Port</td>
</tr>
<tr>
<td>Control Plane</td>
<td>64 KB</td>
</tr>
<tr>
<td>SPAN</td>
<td>38 KB</td>
</tr>
<tr>
<td>Default Class</td>
<td>100 KB</td>
</tr>
</tbody>
</table>
Nexus 6004 FCoE Buffer Allocation

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300 m</td>
</tr>
<tr>
<td>FCoE</td>
<td>187 KB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300 m</td>
</tr>
<tr>
<td>FCoE</td>
<td>165 KB</td>
</tr>
</tbody>
</table>

- By default no buffer is allocated to class FCoE
- The buffer allocation for FCoE is a function of port speed and distance
- There is enough buffer to support one port per UPC for FCoE over 100KM
- FCoE over 100KM requires 9.6 MB dedicated on a single port
Nexus 6004 Ingress Buffer Management

- By default all the shared buffer is allocated to default class
- A CLI is provided to change the amount of shared buffer, the remaining buffer is dedicated and divided between ports
  - hardware shared-buffer-size <0-14.2 MB>
- A Drop class can take half of the shared buffer.
- The queue-limit change the fixed ingress buffer allocation for a class.
Nexus 6004 Egress Buffer Management

Egress Buffer

Unicast

Multicast

Egress UPC

Pool 1
Pool 2
Pool 3
Pool 4

Pool 1
Pool 2
# Nexus 6004 Egress Buffer Management

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast</td>
<td>212 KB</td>
</tr>
<tr>
<td>Multicast</td>
<td>None</td>
</tr>
</tbody>
</table>

### Fabric Mode 10G

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast</td>
<td>212 KB</td>
</tr>
<tr>
<td>Multicast</td>
<td>None</td>
</tr>
</tbody>
</table>

### Fabric Mode 40G

<table>
<thead>
<tr>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast</td>
<td>212 KB</td>
</tr>
<tr>
<td>Multicast</td>
<td>None</td>
</tr>
</tbody>
</table>
## Nexus 6004 Egress Buffer Management

<table>
<thead>
<tr>
<th>Fabric Mode</th>
<th>Traffic Type</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>10G</td>
<td>Unicast</td>
<td>672 KB</td>
</tr>
<tr>
<td></td>
<td>Multicast</td>
<td>None</td>
</tr>
<tr>
<td>40G</td>
<td>Unicast</td>
<td>795 KB</td>
</tr>
<tr>
<td></td>
<td>Multicast</td>
<td>None</td>
</tr>
</tbody>
</table>

- **Traffic Type**
  - Dedicated / 40 Gig Port
  - Shared

- **Buffer**
  - None
  - 6.1 MB
Nexus 6004 Ingress Queuing

- **4K Unicast VoQ**
  - Per port, Per Class of Service

- **8K Multicast VoQ**
  - Per UPC Fan out

- **32 SPAN VoQ**
  - Per SPAN Session

**UPC**

**Switch Fabric**

224 Gig
Nexus 6004 Egress Queuing

Switch Fabric

Egress UPC

12 x 8 = 96

Unicast Queues

Multicast Queues

8 queues

P1

8 queues

P12

12 x 8 = 96

448 Gig
Nexus 6004 QoS flow

**Ingress UPC**
- Trust CoS/DSCP L2/L3/L4 info with ACL
- Traffic Classification
- Ingress CoS/DSCP Marking
- Ingress Policing
- MTU checking
- Buffer usage Monitoring
- If buffer usage crosses threshold:
  - Tail drop for drop class
  - Assert pause signal to MAC for no-drop system class
- Truncate or drop packets if MTU is violated

**Egress UPC**
- PAUSE ON/OFF signal
- ECN Marking
- Egress Policing
- Egress scheduling
- Egress UPC
- Strict priority + DWRR scheduling
- Proxy Queues
- Unicast queues
- Multicast queues
- Egress Queues
- SPAN queues
- VoQs for unicast (8 per egress port)
- Multicast queues

Switch Fabric
Nexus 6004 QoS Features

• 8 Classes of service. 2 reserved for control traffic. 6 for data traffic
• Traffic classification
  – DSCP/CoS/ ACL
• Strict Priority queuing and DWRR
  – DCBX 802.1Qaz
• Packet marking
  – DSCP/CoS/ECN
• Ingress policing/egress policing
  – 4K policer per ASIC
• No drop system class
• Flexible buffer management
Nexus 6004 ACL Table Scaling

• 4K TCAM entries per ASIC.
• TCAM entries are shared if same policy are applied for multiple interfaces or VLANs
• There are 24 L4ops for ingress and 24 L4ops for egress. 16 L4ops are used for tcp/udp port range (each protocol can use up to 12 L4ops) and 8 are usef for TCP flag
• Software can choose to expand ACL entry with UDP/TCP port to multiple ACE in TCAM automatically
Nexus 6004 ACL Entries Sharing

- When same ACL policy (security ACL such as PACL, VACL, RACL and QoS ACL) is applied to multiple interfaces or VLAN there is only one copy stored in TCAM. The same ACL rules are shared by multiple interfaces and VLANs.

- Each ACL policy has label. By assign same label to multiple interfaces and VLANs same TCAM rule can be applies to multiple interfaces or VLANs

```
interface Ethernet1/10
  ip port access-group ip-list-1 in
interface Ethernet1/11
  ip port access-group ip-list-1 in
interface Ethernet1/12
  ip port access-group ip-list-1 in
```

<table>
<thead>
<tr>
<th>Label</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>xyz</td>
<td>eth1/10</td>
</tr>
<tr>
<td>xyz</td>
<td>eth1/11</td>
</tr>
<tr>
<td>xyz</td>
<td>eth1/12</td>
</tr>
</tbody>
</table>

IPV4 ACL ip-list-1

10 permit ip 100.1.1.0/24 200.1.1.0/24
20 permit ip 100.1.2.0/24 200.1.2.0/24
30 permit tcp 100.1.3.0/24 200.1.3.0/24 range 100 2000
Nexus 6004 TCAM Partition

Default TCAM Partition

- VACL(1024)
- IFACL(1152)
- QoS(448)
- RBACL(1024)
- SPAN(64)
- Control Traffic(256)

User configured TCAM Partition

- VACL
- IFACL
- QoS
- RBACL
- SPAN
- Sup

4096 Entries

Default TCAM Partition

User configured TCAM Partition
Nexus 6000 Multicast
Nexus 6004 Multicast Highlights

High Performance
- Line rate L2 and L3 multicast throughput with all frame sizes
- Low latency at scale

High Scalability
- 64k mroute table.

Larger buffer for burst absorption

Optimized multicast replication
- Fabric replication and egress replication

Enhanced features
- IP based forwarding for IGMP snooping
- PIM-BiDir support
- Flow based hashing for multicast over PortChannel
- Better traffic visibility
IP Based Forwarding for IGMP Snooping

- Source IP and group address based forwarding for IGMPv3 snooping even when N6k is L2 switch
- Can filter traffic based on source IP for IGMPv3
- No concern of overlapping multicast MAC addresses

Multicast MAC based forwarding

Vlan10 0100.5E01.0101 eth1/1

IP based forwarding

Vlan10 10.0.0.1 224.1.1.1 eth1/1
Nexus 6004 Multicast Packet Replication

- Fabric replication: One copy is sent to each egress UPC that has at least one receiver
- Egress replication: UPC replicates packets locally to each port and multiple copies to same port if needed
- Egress buffering for microburst and oversubscription
- Drop multicast packet at egress on the per port per queue basis for congestion
Multicast Replication and VoQ

- Optimized replication: fabric and egress replication
- 8K Multicast VOQ at ingress to avoid HOLB
- Track the fanout of the egress UPC. Packets with different egress UPC fanout are assigned to different VoQ
Multicast Hashing over Port Channel

- Real flow based hashing for multi-destination traffic
- Traffic is replicated to all egress UPC where PortChannel member resides
- Egress UPC run hash calculation and choose an egress port for a particular flow. Rest of UPCs drops the packet.
- Extra bandwidth from fabric to egress UPC to handle the extra multi-destination packets
Nexus 6000

Summary Key Enhancements

**Performance**
- L2 and L3 line rate at 10Gig and 40Gig, Latency at 1us, 1K way ECMP, Line rate SPAN

**Scalability**
- 256K MAC/IP host routes, 32K LPM routes, 16K Bridge Domains, 4K VRF, 64K RPF, 64K Mroute table

**Buffering**
- 25M shared buffer per 3 QSFP ports

**Multicast**
- Optimized multicast replication, flow based hashing, IP (S,G) / (*,G) lookup even for L2 multicast, egress midx translation

**SPAN**
- Line rate SPAN, 31 SPAN sessions, SPAN on drops, SPAN on high latency

**Analytics**
- Sampled Netflow, buffer monitoring, microburst monitoring, latency monitoring, SPAN on Drop, SPAN on High Latency
Your feedback is important to us.

Complete the session survey at: www.ciscolivelondon.com/onsite or via the Cisco Live Mobile App