White paper Cisco public IIIIII CISCO The bridge to possible

# Flexible Forwarding Table on Nexus 9000

# Contents

Introduction	3
Flexibility with Cisco Cloud Scale ASIC	3
Cloud Scale ASIC forwarding block	4
ASIC	4
Slices	5
Forwarding	6
Routing templates	8
Nexus 9000 routing template for Cloud Scale	9
Verified scale on different ASICs for fixed platforms	10
Verified scale on different ASICs for modular platforms	14
Configuration and verification of routing template	16
Configuration	16
Verification	17
Conclusion	20
Further references	20

# Introduction

The Cisco Nexus<sup>®</sup> 9000 Series Switches are industry-leading data center switches designed to cater to the needs for modern-day data centers. The Nexus 9000 family offers a wide range of switches built using both merchant and custom Cisco<sup>®</sup> ASICs.

Cisco Cloud Scale is a family of ASICs, developed in-house by Cisco, that power the latest generation of Nexus 9000 switches.

The Cloud Scale Nexus 9000 product family consists of both Nexus 9300 series fixed form-factor and Nexus 9500 series modular switches.

Some of the key highlights of Cloud Scale ASICs are as listed below:

- Ultra-high port densities
- Multi-speed from 100 Mbps to 400 Gbps
- Rich feature-set
- Flexible forwarding scale
- Intelligent buffering
- Built-in analytics and telemetry

# Flexibility with Cisco Cloud Scale ASIC

Cisco Nexus 9000 switches power data centers across a variety of verticals – be it enterprise, financials, universities, media and entertainment, or service providers. Depending on the use case, the requirements when it comes to scale could change. Network deployment within an enterprise data center typically requires a large host route scale, a service provider using the Nexus 9000 as an internet peering edge requires a large prefix scale or for customers in financials and media and entertainment, who are multicast heavy, require a large multicast route scale. The Cloud Scale ASIC forwarding table allows the capability to resize the forwarding table based on the requirements driven by the use cases.

# Cloud Scale ASIC forwarding block

## ASIC

The most important component of the switch is its ASIC. The ASIC is a custom-built chip designed to perform a specific function. In the case of a switch, this ASIC is designed to perform packet forwarding operations, which are switching and routing.

To perform forwarding lookup operations, ASICs have the following building blocks:

- MAC layer responsible for transmitting and receiving the packets to and from the wire
- Parser determines if a particular frame/packet should be L2 or L3 processed
- Forwarding engine determines, based on the frame/packet header, what egress port this frame/packet should exit through
- Forwarding table stores various forwarding tables such as MAC, ARP, IPv4/V6 route tables, etc., which are consumed by the forwarding engine to make forwarding decisions
- Classification TCAM mainly used to store ACL and QoS policies
- Buffer reserved memory space to store the packet during congestion

The following table shows the Cisco Nexus 9300 Cloud Scale family switches and their ASICs:

#### Table 1. Nexus 9000 Cloud Scale models and ASICs



## Slices

Cisco Nexus 9000 Cloud Scale ASICs are further divided into slices. A slice is a self-contained forwarding block controlling a subset of ports on the ASIC. Additionally, each slice has its own dedicated resources (discussed above) to perform packet forwarding.

The idea behind implementing a slice is to build multiple parallel forwarding pipelines to achieve greater throughput. Hence a slice is a switch in itself and therefore is called an SoC (a "switch on chip").

Depending on the ASIC form factor, the number of slices in an ASIC varies.



Figure 1. Cloud Scale ASICs

Forwarding resources such as TCAM, buffer, etc., are assigned to the slices. Each slice has its own dedicated TCAM and buffer resources that are not shared with any other slices on the same ASIC.

## Forwarding



#### Figure 2.

Cloud Scale ASIC forwarding block

In Cloud Scale ASICs, the forwarding information is stored in two different resources:

- Forwarding TCAM
- Flexible tiles

**Note:** Forwarding TCAM is NOT used for classification ACLs. That is a separate resource ACL/QoS TCAM.

#### Forwarding TCAM

TCAM memory, front-ending flexible forwarding lookups

- Handles overflow/hash collisions
- Divided into smaller portions called banks (usually 1K in size)

#### Flexible forwarding TCAM (flex tiles)

Provide fungible pool of table entries for lookups

Variety of functions, including:

- IPv4/IPv6 unicast longest-prefix match (LPM)
- IPv4/IPv6 unicast host-route table (HRT)
- IPv4/IPv6 multicast (\*,G) and (S,G)
- MAC address table / adjacency table
- ECMP tables
- ACI policy



#### Figure 3.

Forwarding TCAM and Flex tiles

Majority of forwarding lookup tables are programmed in flex tiles.

The size of the forwarding TCAM varies based on the ASIC and hence the number of banks and tiles. The following table shows the forwarding TCAM and flex tiles size per slice:

|--|

ASIC	No. of slices	Forwarding TCAM	Flex tiles
LS1800 EX	2	8 x 2K	68 x 8K
LS1800 FX	1	16 x 1K	136 x 8K
LS3600 FX2	2	24 x 1K	68 x 8K
LS1800 FX3	1	24 x 1K	136 x 8K

ASIC	No. of slices	Forwarding TCAM	Flex tiles
LS6400 GX	4	24 x 1K	136 x 8K
S 6400	4	16 x 1K	44 x 8K

Depending on the routing mode (we call it the routing template) that is configured, flex tiles are carved out and allocated to various forwarding tables. This provides flexibility to increase and decrease the various table sizes per use case.

# **Routing templates**

Cisco NX-OS offers predefined routing templates that can be used depending on requirements.

Once these templates are applied, it applies to all the slices on the device.

As of now, NX-OS does not offer user-defined templates. Also, only one template can be applied to the device at any time.

Table 3. Routing templates

Routing template	Description
Dual-stack host scale	Maximizes ARP/ND scale Increases the ARP/ND scale to double the default mode value. No IPv4/IPv6 LPM routes No Mcast
Internet peering	Maximizes IPv4/IPv6 LPM table To support IPv4/IPv6 Internet scale routing table Decreases ARP, ND, host route, Mac table No Mcast
L2 heavy	Maximizes Mac-address table scale 200K Decreases IPV4/IPv6 LPM routes No Mcast
LPM heavy	Maximizes IPv4/IPv6 LPM scale Decreases Mac, host route, ARP and ND tables
MPLS heavy scale	Maximizes MPLS table scale and ECMP MPLS table and ECPM groups are increased Decreases ARP, ND, host route, Mac table No Mcast
Multicast extended heavy scale	Increases IPv4/IPv6 Mcast table Reduction in IPv4/IPv6 LPM, ARP,ND and Mac tables
Multicast heavy scale	To further increase IPv4 Mcast table

# Nexus 9000 routing template for Cloud Scale

Not all these templates are supported on all platforms and NX-OS software versions. The following table shows the routing templates supported on various Nexus 9000 Cloud Scale ASICs running Cisco NX-OS Release 9.3(5):

			тс	DR			EC	DR
ASIC	EX	FX	FX2	FX3	GX	S 6400	EX	FX
Routing template								
Default	$\checkmark$							
LPM heavy	$\checkmark$							
L2 heavy	$\checkmark$	x	X	x	$\checkmark$	X	x	x
L3 heavy	x	x	x	x	x	x	x	x
L3 Scale	X	x	X	x	x	X	x	x
Internet peering	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	X	$\checkmark$	$\checkmark$
Dual-stack host scale	X	x	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Dual-stack Mcast	x	x	x	x	x	x	x	х
MPLS scale	x	$\checkmark$						
Mcast heavy	$\checkmark$							
Mcast ext heavy	x	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	x	X	x
Service provider	x	$\checkmark$	$\checkmark$	$\checkmark$	х	x	$\checkmark$	$\checkmark$

 Table 4.
 Routing template support matrix

# Verified scale on different ASICs for fixed platforms

Table 5.Forwarding scale on LS1800 EX

LS1800 EX											
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)	
Default	98304	458752	206438	65536	49152	32768	8192	2048	1000	500	
LPM heavy	32768	786432	353894	32768	32768	16384	8192	2048	1000	500	
L2 heavy	212992	196608	88473	65536	49152	32768	0	0	1000	500	
L3 heavy											
L3 scale											
Internet peering	40960	1000448	500224	32768	32768	16384	0	0	1000	500	
Dual-stack host scale	114688			262144	98304	98304	0	0	1000	500	
MPLS scale	90112	471859	265420	32768	32768	16384	0	0	4000	2000	
Mcast heavy	16384	589824	265420	49152	32768	24576	32768	8192	2000	1000	
Mcast ext heavy											
Service provider											

**Table 6.**Forwarding scale on LS1800 FX

LS1800 FX											
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)	
Default	114688	1153433	628224	196608	98304	98304	32768	8192	4000	2000	
LPM heavy	32768	786432	442368	32768	32768	16384	8192	2048	1000	500	
L2 heavy											
L3 heavy											
L3 scale											
Internet peering	16384	1256448	628224	98304	32768	32768	0	0	1000	500	

LS1800 FX												
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)		
Dual-stack host scale	106496			262144	98304	98304	0	0	1000	500		
MPLS scale	90112	471859	265420	32768	32768	16384	0	0	4000	2000		
Mcast heavy	16384	471859	265420	49152	32768	24576	32768	8192	1000	500		
Mcast ext heavy	32768	104857	58982	163840	32768	32768	131072	8192	1000	500		
Service provider	65536	629145	353894	65536	49152	32768	8192	2048	1000	500		

**Table 7.**Forwarding scale on LS3600 FX2

LS3600 FX2											
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)	
Default	98304	524288	294912	65536	49152	32768	8192	2048	1000	500	
LPM heavy	32768	786432	442368	32768	32768	16384	8192	2048	1000	500	
L2 heavy											
L3 heavy											
L3 scale											
Internet peering	40960	1000448	500224	32768	32768	16384	0	0	1000	500	
Dual-stack host scale	106496			327680	98304	98304	0	0	1000	500	
MPLS scale	49152	576716	324403	32768	32768	16384	0	0	4000	2000	
Mcast heavy	32768	471859	265420	49152	32768	24576	32768	8192	1000	500	
Mcast ext heavy	32768	104857	58982	163840	32768	32768	131072	8192	1000	500	
Service provider	65536	629145	353894	65536	49152	32768	8192	2048	1000	500	

LS1800 FX3												
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)		
Default	114688	1048576	589824	196608	98304	98304	8192	2048	1000	500		
LPM heavy	32768	786432	442368	32768	32768	16384	8192	2048	1000	500		
L2 heavy												
L3 heavy												
L3 scale												
Internet peering	16384	1256448	628224	98304	32768	32768	0	0	1000	500		
Dual-stack host scale	106496			327680	98304	98304	0	0	1000	500		
MPLS scale	49152	576716	324403	32768	32768	16384	0	0	4000	2000		
Mcast heavy	32768	471859	265420	49152	32768	24576	32768	8192	1000	500		
Mcast ext heavy	32768	104857	58982	163840	32768	32768	131072	8192	1000	500		
Service provider	65536	629145	353894	65536	49152	32768	8192	2048	1000	500		

#### Table 8.Forwarding scale on LS1800 FX3

Table 9.Forwarding scale on LS6400 GX

LS6400 GX												
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)		
Default	114688	1153433	628224	196608	98304	98304	32768	8192	4000	2000		
LPM heavy	32768	786432	442368	32768	32768	32768	8192	2048	4000	2000		
L2 heavy	212992	314572	176947	65536	49152	32768	0	0	4000	2000		
L3 heavy												
L3 scale												
Internet peering	16384	1256448	628224	98304	32768	32768	0	0	4000	2000		
Dual-stack host scale	106496			327680	98304	98304	0	0	4000	2000		

LS6400 GX												
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)		
MPLS scale	106496	681574	383385	196608	196608	98304	0	0	4000	2000		
Mcast heavy	32768	419430	235929	65536	32768	32768	32768	8192	4000	2000		
Mcast ext heavy	32768	104857	58982	163840	32768	32768	131072	8192	4000	2000		
Service provider												

### **Table 10.**Forwarding scale on S6400

\$6400												
	MAC	IPv4 LPM	IPv6 LPM	Host routes (v4/v6)	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)		
Default	32768	131072	65536	65536	32768	32768	16384	8192	1000	500		
LPM heavy	32768	262144	131072	32768	32768	16384	8192	2048	1000	500		
L2 heavy												
L3 heavy												
L3 scale												
Internet peering												
Dual-stack host scale	49152			163840	65536	65536	0	0	1000	500		
MPLS scale	32768			98304	90112	49152	0	0	4000	2000		
Mcast heavy	49152			98304	32768	32768	32768	8192	1000	500		
Mcast ext heavy												
Service provider												

**Note:** All these scales are as per Cisco NX-OS Release 9.3(5).

# Verified scale on different ASICs for modular platforms

Cisco Nexus 9500 Series Switches use a distributed forwarding architecture, where forwarding tables are distributed between line cards and fabric modules (mostly IPv6 tables are programmed on FMs).

EX													
		MAC	IPv4 Trie	IPv6 Trie	IPV4 Host routes	IPV6 Host routes	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)	ECMP Groups
Default	LC	98304	589824		65536		49152		8192	2048	1000	500	24574
	FM			176947		32768		32768		2048			
LPM heavy	LC	49152	786432		32768		32768		8192	2048	1000	500	24574
	FM			235929		32768				2048			
L2 heavy													
L3 heavy													
L3 scale													
Internet peering	LC	40960	1000448	500224	32768 (v4/V6)	32768	32768	16384	NA	NA	1000	500	24574
	FM			176947									
Dual-stack host scale													
Dual-stack Multicast	LC	65536	262144	147456	114688 (v4/v6)		65536		8192	8192	1000	500	24574
	FM							32768	8192				
MPLS scale	LC	90112	471859	265420	32768 (v4/v6)		32768	16384			4000	2000	7166
	FM												
Mcast heavy	LC	16384	471859	265420	49152 (v4/v6)		32768		32768	8192	1000	500	24574
	FM							24576		8192			
Mcast ext heavy													
Service	LC	65536	629145	353894	(v4/v6)		49152	32768	8192		1000	500	24574
provider	FM									8192			

Table 11. Forwarding scale on EX linecard

### Table 12. Forwarding scale on FX linecard

FX													
		MAC	IPv4 Trie	IPv6 Trie	IPV4 Host routes	IPV6 Host routes	ARP	ND	IPv4 Mcast	IPv6 Mcast	MPLS label (without ECMP)	MPLS label (with ECMP)	ECMP Groups
Default	LC	98304	589824		65536		49152		8192	2048	1000	500	24574
	FM			176947		32768		32768		2048			
LPM heavy	LC	49152	786432		32768		32768		8192	2048	1000	500	24574
	FM			235929		32768				2048			
L2 heavy													
L3 heavy													
L3 scale													
Internet peering	LC	40960	1256448	628224	32768 (v4/V6)	32768	32768	16384	NA	NA	1000	500	24574
	FM			176947									
Dual-stack host scale													
Dual-stack Multicast	LC	65536	262144	147456	114688 (v4/v6)		65536		8192	8192	1000	500	24574
	FM							32768	8192				
MPLS scale	LC	90112	471859	265420	32768 (v4/v6)		32768	16384			4000	2000	7166
	FM												
Mcast heavy	LC	16384	471859	265420	49152 (v4/v6)		32768		32768	8192	1000	500	24574
	FM							24576		8192			
Mcast ext heavy													
Service	LC	65536	629145	353894	(v4/v6)		49152	32768	8192		1000	500	24574
PIONICE	FM									8192			

**Note:** All these scales are as per Cisco NX-OS Release 9.3(5).

# Configuration and verification of routing template

Changing the routing template is a two-step process:

- 1. Select the desired routing template.
- 2. Save the configuration and reload the device.

# Configuration

```
N9K# conf
Enter configuration commands, one per line. End with CNTL/Z.
N9K(config)#
N9K(config)#
N9K(config) # system routing?
  template-dual-stack-host-scale Dual Stack Host Scale
  template-dual-stack-mcast
                                  Dual Stack Multicast
  template-internet-peering
                                  Internet Peering
  template-12-heavy
                                  L2 Heavy 200k MAC scale profile
  template-lpm-heavy
                                  LPM Heavy
  template-mpls-heavy
                                  MPLS Heavy Scale
  template-multicast-ext-heavy
                                  Multicast Extended Heavy Scale
  template-multicast-heavy
                                  Multicast Heavy Scale
                                  Service Provider
  template-service-provider
```

#### N9K(config) # system routing template-lpm-heavy

Warning: The command will take effect after next reload.Set the LPM scale using below CLI if multicast is needed hardware profile multicast max-limit lpm-entries <2048/4096> Note: This requires copy running-config to startup-config before switch reload. N9K(config)#

# Verification

The **Show hardware forwarding** command is used to check the configured routing template, and to check the scale and utilization of the various forwarding tables.

#### Configured System Routing Mode: LPM Heavy

```
Dynamic V6 Trie : False
```

\_\_\_\_\_ Max IPv4 Trie route entries: 786432 Max IPv6 Trie route entries: 353894 Max TCAM table entries : 16384 Max V4 Ucast DA TCAM table entries : 6144 Max V6 Ucast DA TCAM table entries : 2048 Max native host route entries (shared v4/v6) : 32768 Max v6 /128 learnt host route entries : 24576 Max ARP entries (Entries might overflow into tcam as Host-As-Route): 32768 Max ND entries (Entries might overflow into tcam as Host-As-Route): 16384 Total number of IPv4 host trie routes used : 0 Total number of IPv4 host tcam routes used : 1 Total number of IPv4 LPM trie routes used : 0 Total number of IPv4 LPM tcam routes used : 3 Total number of IPv6 host trie routes used : 0 Total number of IPv6 host tcam routes used: 0 Total number of IPv6 LPM trie routes used : 0 Total number of IPv6 LPM tcam routes used : 8 Total number of IPv4 host native routes used in native tiles : 1 Total number of IPv6 host native routes used in native tiles : 0 Total number of IPv6 ND/local routes used in native tiles : 0 Total number of IPv6 host /128 learnt routes used in native tiles : 0 IPv4 Host-as-Route count : 1 IPv6 Host-as-Route count : 0 Nexthop count : 3

Percentage utilization of IPv4 native host routes : 0.00 Percentage utilization of IPv6 native host routes : 0.00 Percentage utilization of IPv6 ND/local routes : 0.00 Percentage utilization of IPv6 host /128 learnt routes : 0.00 Percentage utilization of IPv4 trie routes : 0.00 Percentage utilization of IPv6 trie routes : 0.00 Percentage utilization of IPv4 TCAM routes : 0.08 Percentage utilization of IPv6 TCAM routes : 0.39 Percentage utilization of nexthop entries : 0.00

#### IPv4/Ipv6 Mcast host entry summary

-----

#### Max Mcast Route Entries Limit = 8192

Max Mcast V4 SA TCAM entries (S/m) = 2048 Max Overflow Mcast v4 SA TCAM entries = 0 Max Overflow Mcast V4 DA TCAM entries = 0 Total number of IPv4 Multicast SA LPM routes used = 0 Percentage utilization of IPv4 Multicast SA LPM routes = 0.00 Used Mcast Entries = 0 Used MCIDX Count = 1 Used \*, G Entries in HRT = 0 Used \*, G Entries in LPM = 0 Used (S,G) Entries = 0 Used Mcast S/32 Entries in HRT = 0 Max Reserved Mcast S/32 and G/32 in LPM = 0  $\,$ Used Mcast S/32 Entries in SA LPM = 0 Percentage utilization of S/32 in SA LPM = 0.00 Used Mcast G/32 Entries in DA LPM = 0 Percentage utilization of G/32 in DA LPM = 0.00

#### Max IPv6 Mcast Route Entries Limit = 2048

Used IPv6 Mcast Entries = 0 Used IPv6 \*,G Entries in HRT = 0 Used IPv6 \*,G Entries in LPM = 0 Used IPv6 (S,G) Entries = 0 Used IPv6 Mcast S/128 Entries in HRT = 0

```
Mcast Mac entry summary
```

Max Mcast Mac Route Entries Limit = 0 Used Mcast Mac Entries = 0

-----MPLS Hardware Resources-----

```
Max Label Entries ( without ECMP ) = 1000
Max Label Entries ( with ECMP ) = 500
Used Label Entries = 0
No. of MPLS VPN Labels Used = 0
```

-----ECMP GROUP/TILE INFO-----

ECMP group supported : Yes Max ECMP groups : 1022 ECMP groups used : 0 Num of ECMP group member tiles : 1 Num of ECMP group tiles : 1

QoS Resource Utilization

\_\_\_\_\_

Resource	Module	Total	Used	Free
Aggregate policers:	1	4094	50	4044
Distributed policers:	1	4094	0	4094
Policer Profiles:	1	4094	50	4044

N9K#

# Conclusion

The Cisco Nexus 9000 Cloud Scale ASIC delivers the needed performance and scale to meet and exceed the requirements for next generation data center networks. The ability to use the same hardware in a variety of deployments and use cases provides investment protection for our customers. Combined with rich features, granular hardware telemetry ensures data center design and operations can be optimized.

## Further references

Cisco Nexus 9000 Series NX-OS Unicast Routing Configuration Guide, Release 9.3(x):

https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/93x/unicast/configuration/guide/ b-cisco-nexus-9000-series-nx-os-unicast-routing-configuration-guide-93x/b-cisco-nexus-9000-series-nxos-unicast-routing-configuration-guide-93x\_chapter\_011001.html

Cisco Nexus 9000 Series NX-OS Verified Scalability Guide, Release 9.3(5): https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/93x/scalability/guide-935/cisconexus-9000-series-nx-os-verified-scalability-guide-935.html

Americas Headquarters Cisco Systems, Inc.

San Jose, CA

Asia Pacific Headquarters Cisco Systems (USA) Pte. Ltd. Singapore Europe Headquarters Cisco Systems International BV Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at https://www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: https://www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

Printed in USA

C11-2755883-00 01/22