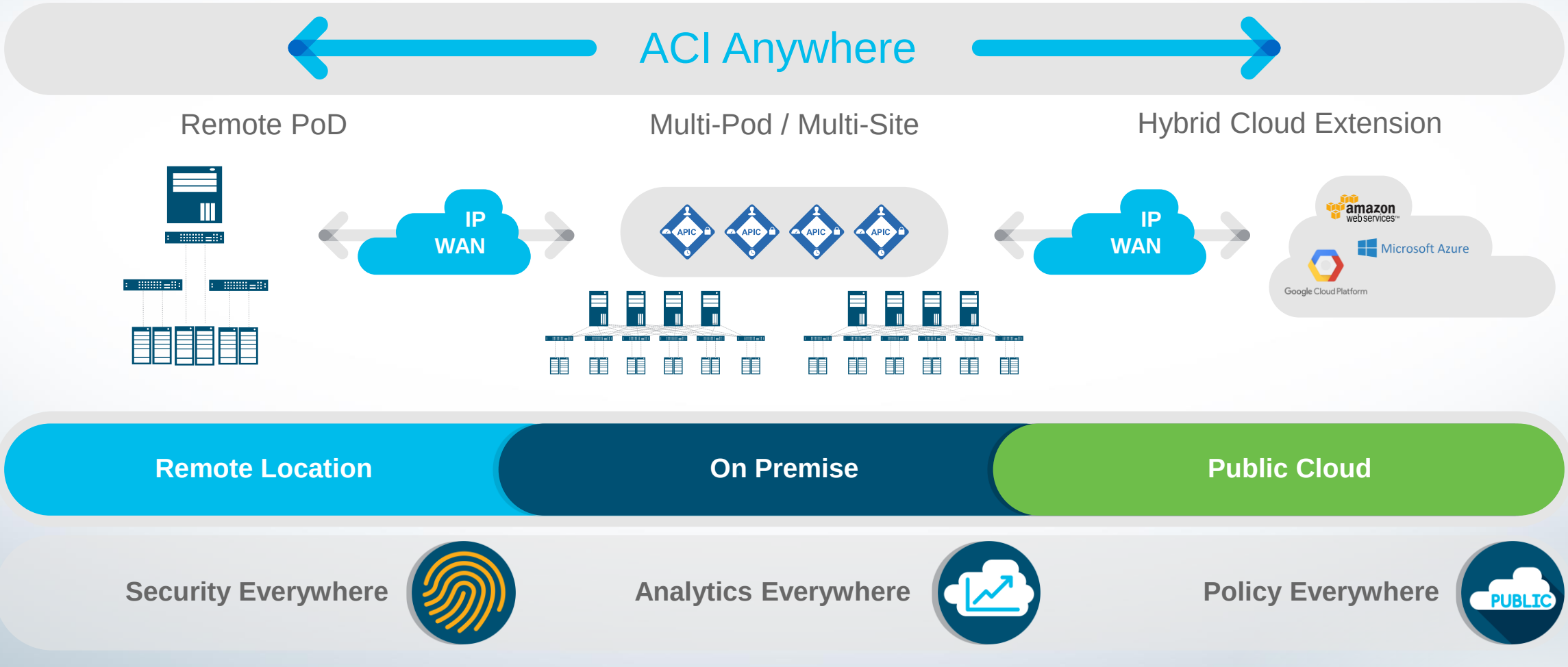
ACI 3.0 update

Brian Kvisgaard, System Engineer - Datacenter Switching
bkvisgaa@cisco.com

cisco

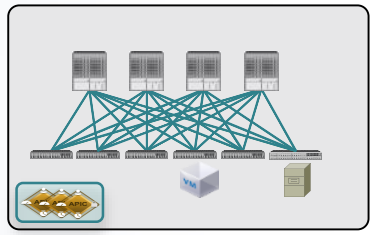
ACI Anywhere - Vision

Any Workload, Any Location, Any Cloud



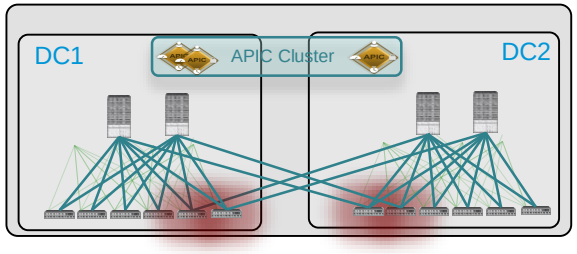
Cisco ACI
Fabric and Policy Domain Evolution

ACI Single Pod Fabric



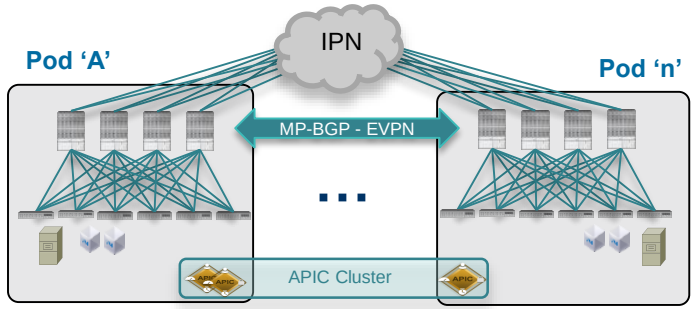
ACI 1.0 Leaf/Spine
Single Pod Fabric

ACI Stretched Fabric



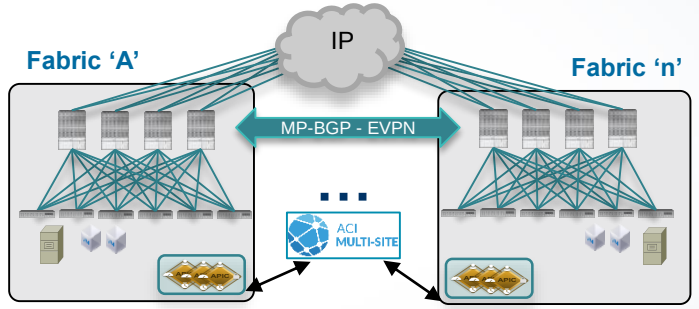
ACI 1.1
Geographically
Stretch a single fabric

ACI Multi-Pod Fabric



ACI 2.0 - Multiple
Networks (Pods) in a
single Availability Zone
(Fabric)

ACI Multi-Site

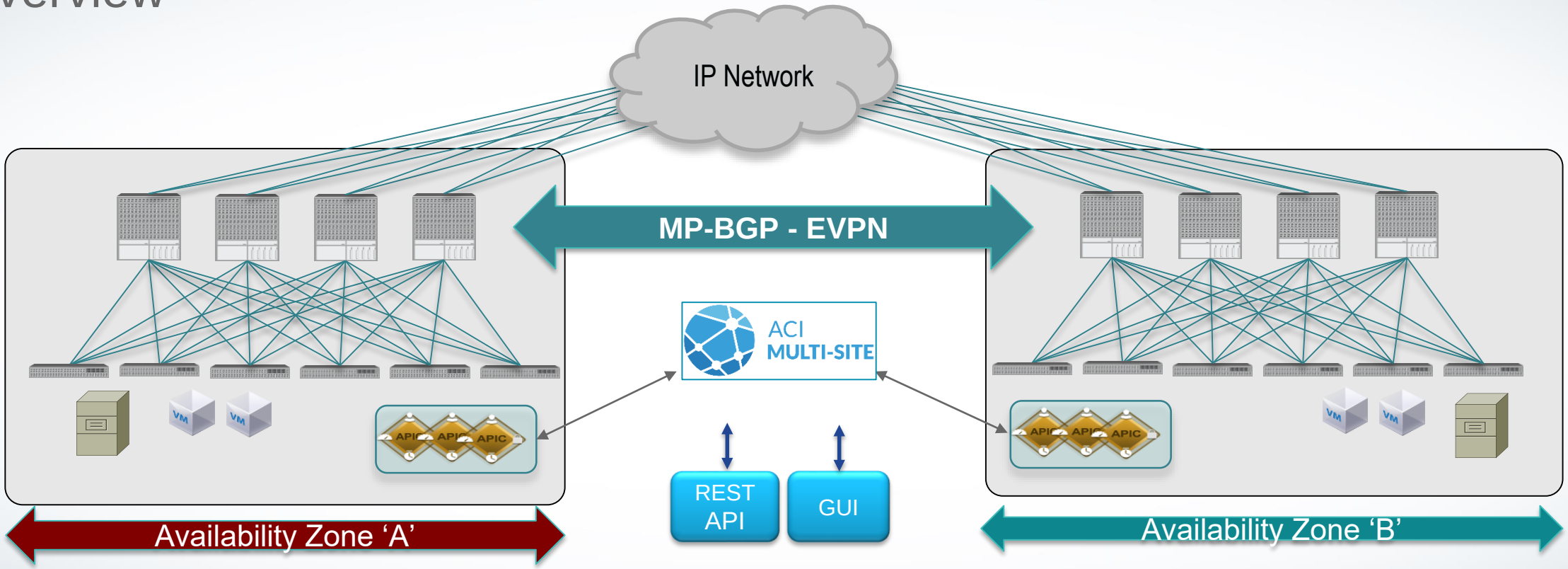


ACI 3.0 - Multiple
Availability Zones
(Fabrics) in a Single
Region 'and' Multi-
Region Policy
Management

…more to
come!

ACI Multi-Site
Overview

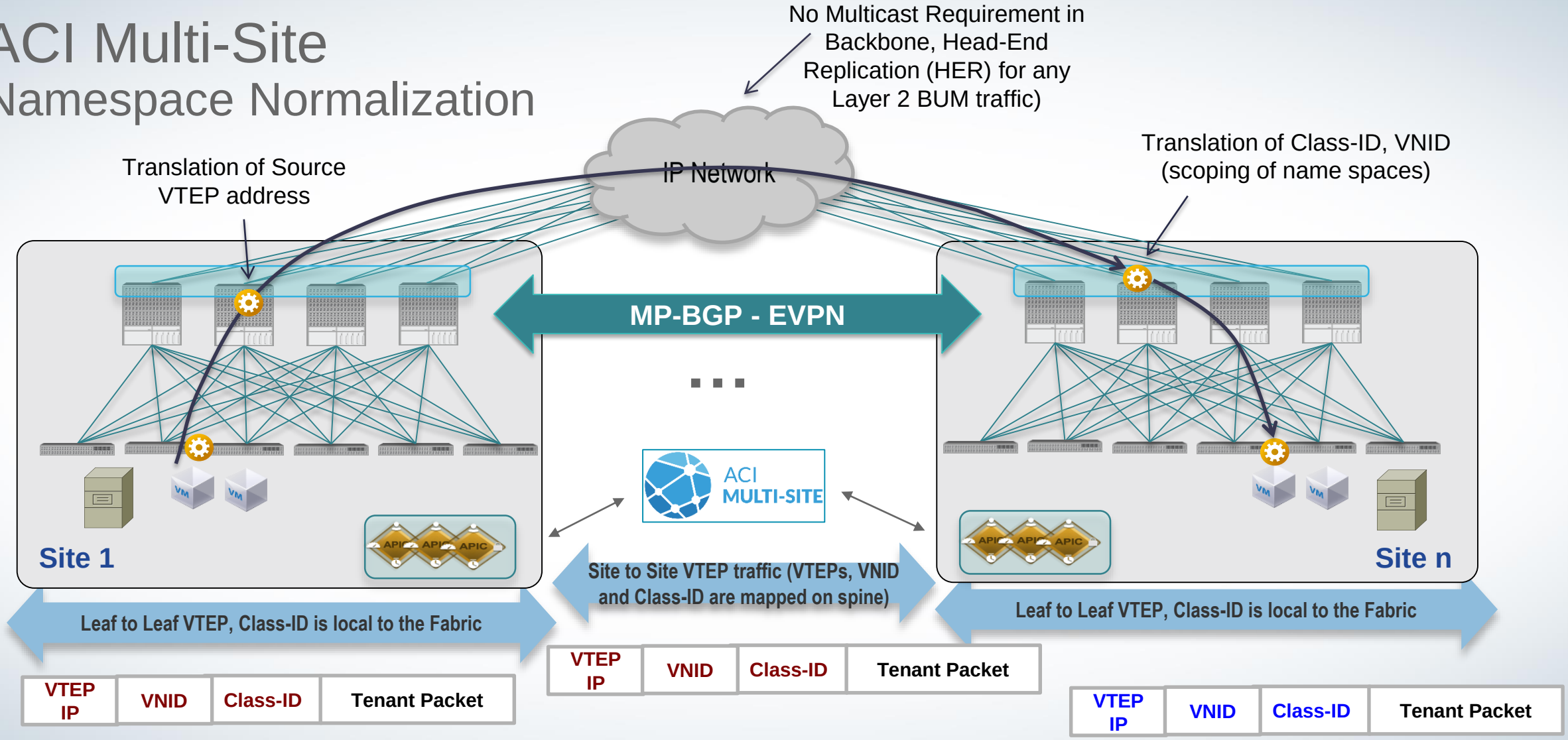
From ACI 3.0 Release



- Separate ACI Fabrics with independent APIC clusters
- ACI Multi-Site defines and pushes cross-fabric policies to multiple APIC clusters providing scoping of all configuration changes

- MP-BGP EVPN control plane between sites
- Data Plane VXLAN encapsulation across sites
- End-to-end policy definition and enforcement

ACI Multi-Site
Namespace Normalization

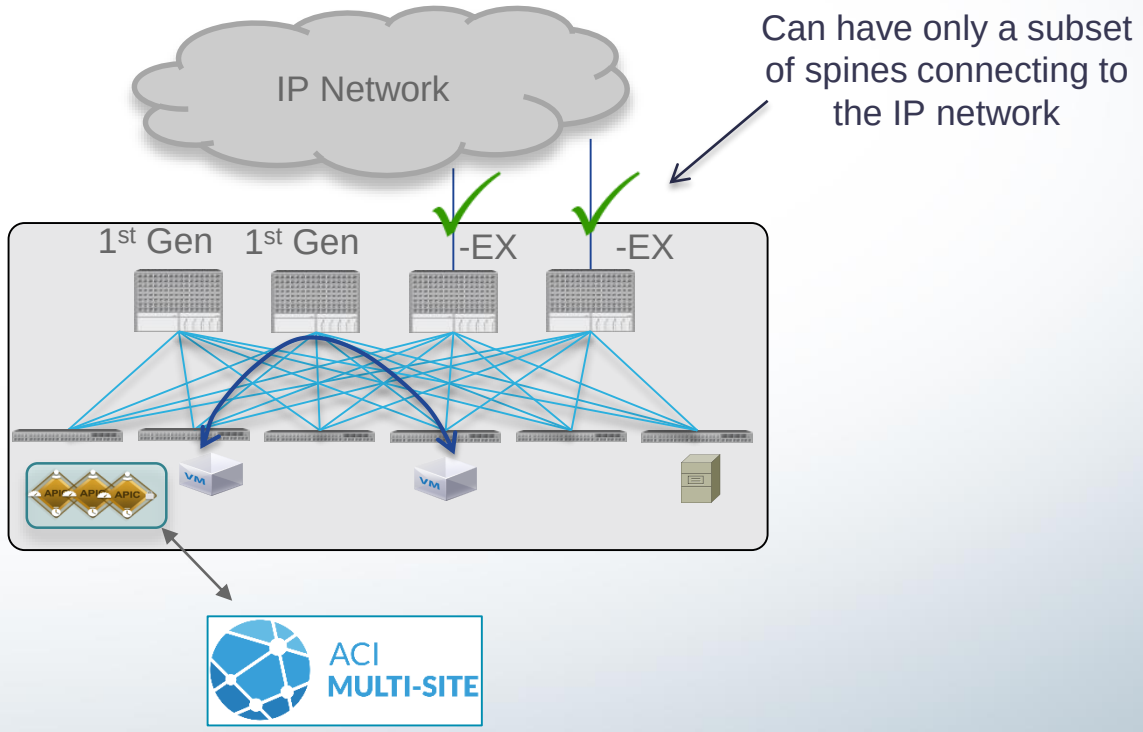


- Maintain separate name spaces with ID translation performed on the spine nodes
- Requires specific HW on the spine to support for this functionality

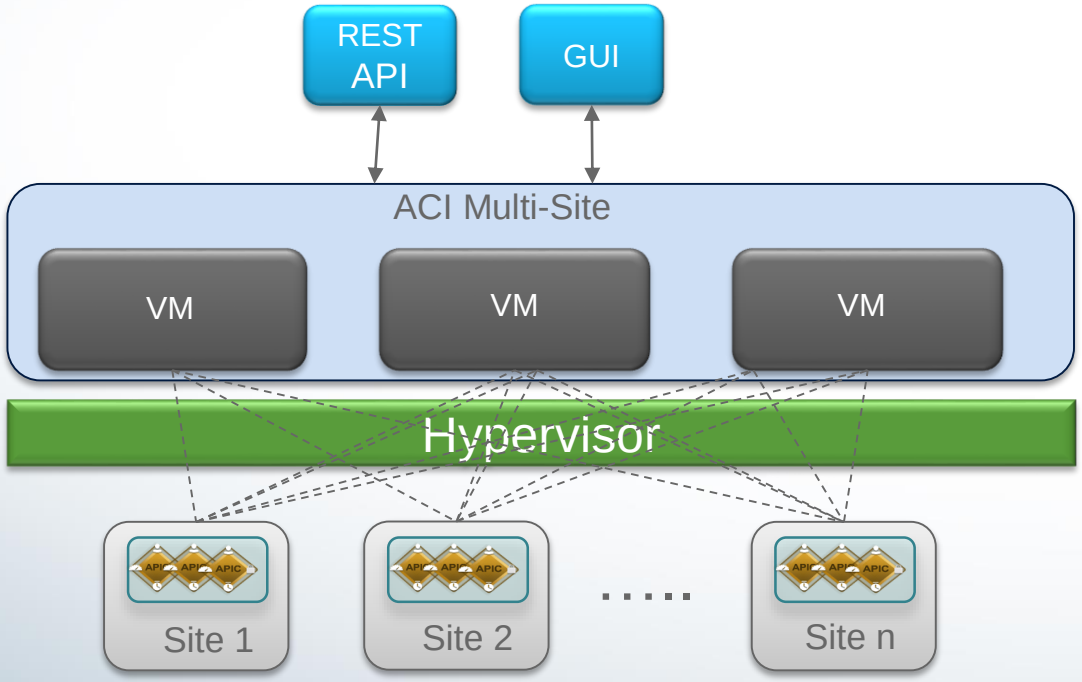
ACI Multi-Site
Hardware Requirements

- Support all ACI leaf switches (1st Generation, -EX and -FX)
- Only -EX spine nodes (or newer) to connect to the inter-site network
- New FX non modular spine (9364C, 64x40G/100G ports) will be supported for Multi-Site in Q1CY18 timeframe
- 1st generation spines (including 9336PQ) not supported

  Can still leverage those for intra-site leaf to leaf communication



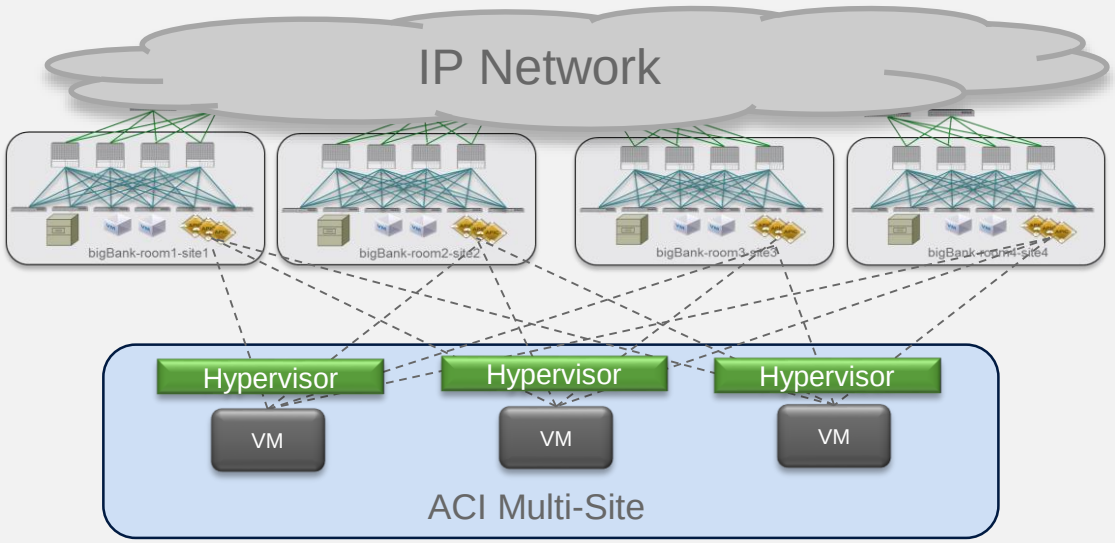
ACI Multi-Site
Multi-Site Policy Manager



- Micro-services architecture
  - Multiple VMs are created and run concurrently (active/active)
  - vSphere only support at FCS (KVM and physical appliance support scoped for future releases)
- OOB Mgmt connectivity to the APIC clusters deployed in separate sites
  - Support for 500 msec to 1 sec RTT
- Main functions offered by ACI Multi-Site:
  - Monitoring the health-state of the different ACI Sites
  - Provisioning of day-0 configuration to establish inter-site EVPN control plane
  - Defining and provisioning policies across sites (scope of changes)
  - Inter-site troubleshooting (post-3.0 release)
- Recommended to deploy ACI Multi-Site for a single ACI site to plan for a future Multi-Site deployment

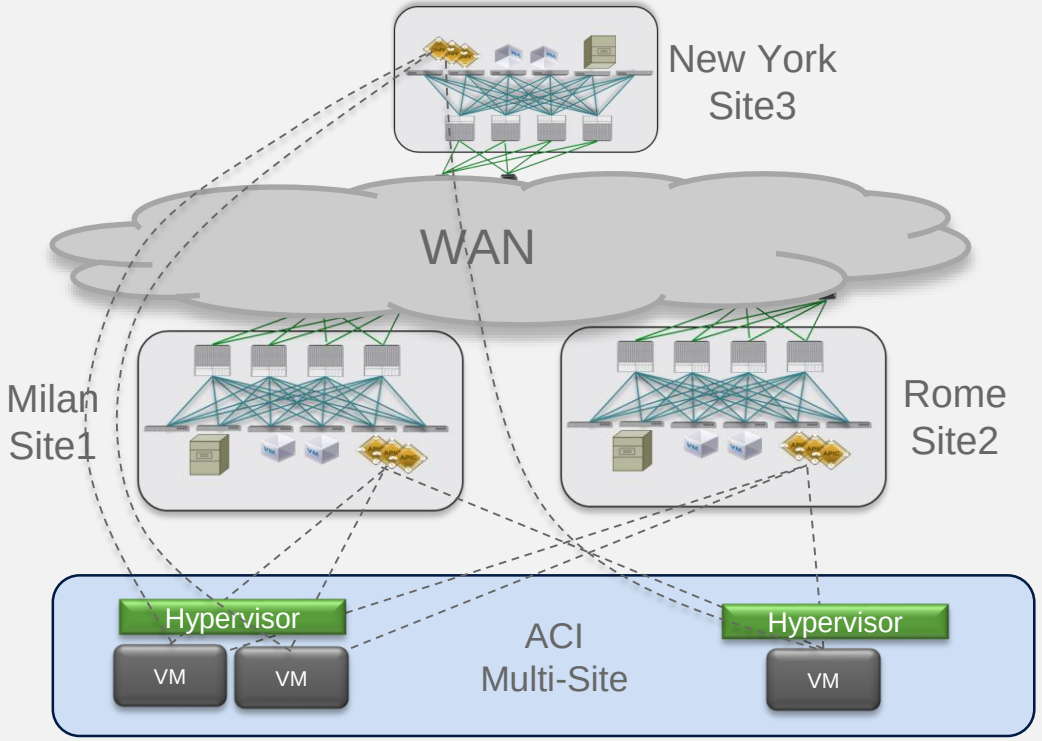
ACI Multi-Site Policy Manager
Deployment Considerations

Intra-DC Deployment



- Hypervisors can be connected directly to the DC OOB network
- Each ACI Multi-Site VM has a unique routable IP
- Async calls from ACI Multi-Site to APIC

Interconnecting DCs over WAN



- Moderate latency (up to 150 msec) between ACI Multi-Site nodes
- Higher latency (500 msec to 1 sec RTT) between ACI Multi-Site nodes and remote APIC clusters
- If possible deploy a node in each site for availability purposes (network partition scenarios)

APIC vs. ACI Multi-Site Policy Manager Functions

- Central point of management and configuration for the Fabric
- Responsible for all Fabric local functions

Fabric discovery and bring up
Fabric access policies
Service graphs
Domains creation (VMM, Physical, etc.)
…

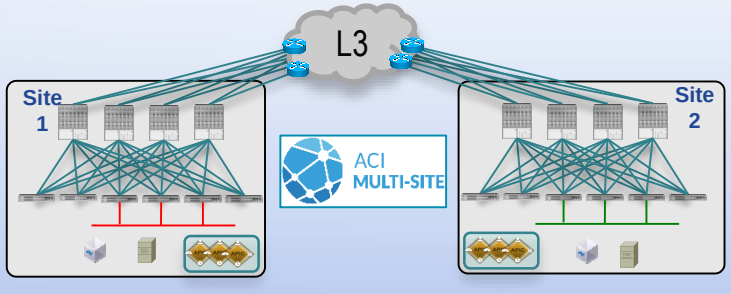
- Integration with third party services
- Maintains runtime data (VTEP address, VNID, Class_ID, GIPo, etc.)
- No participation in the fabric control and data planes

- Complementary to APIC
- Provisioning and managing of **“Inter-Site Tenant and Networking Policies”**
- Scope of changes

  Granularly propagate policies to multiple APIC clusters
- Can import and merge configuration from different APIC cluster domains
- End-to-end visibility and troubleshooting
- No run time data, configuration repository
- No participation in the fabric control and data planes

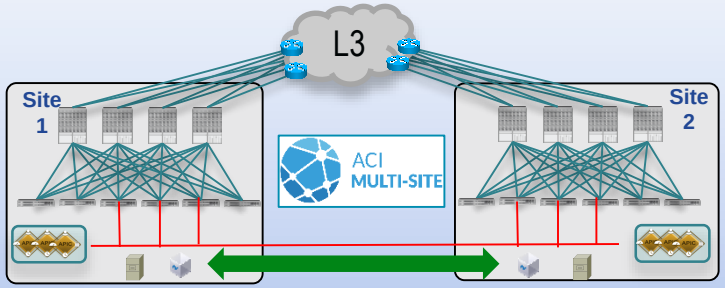
ACI Multi-Site Networking Options
Per Bridge Domain Behavior

Layer 3 only across sites



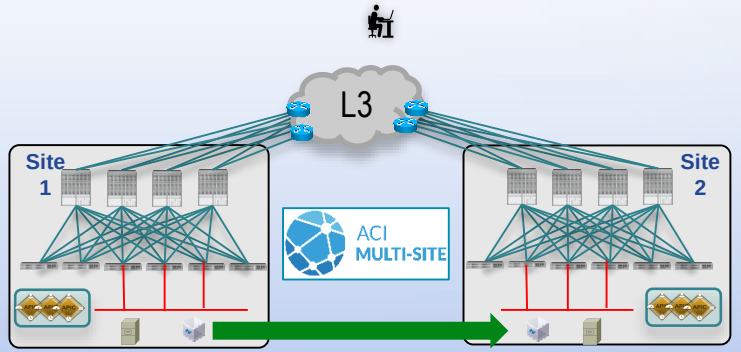
- Bridge Domains and subnets not extended across Sites
- Layer 3 Intra-VRF or Inter-VRF communication only

Full Layer 2 and Layer 3 Extension



- Interconnecting separate sites for fault containment and scalability reasons
- Layer 2 domains stretched across Sites (Support for 'hot' VM migration)
- **Layer 2 flooding across sites**

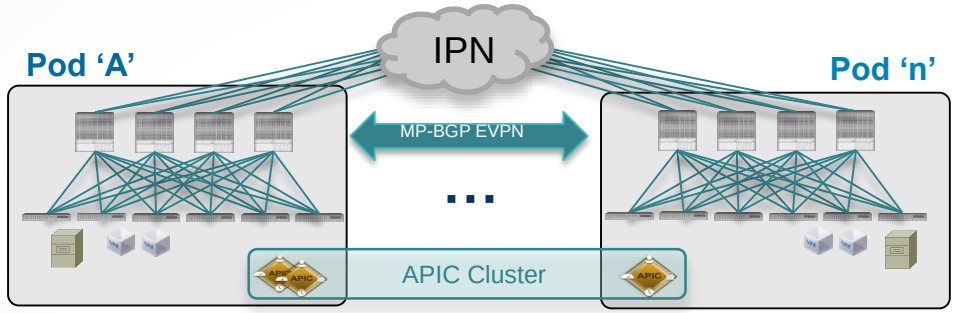
IP Mobility without L2 flooding



- Same IP subnet defined in separate Sites
- Support for IP Mobility ('cold' VM migration) and intra-subnet communication across sites
- **No Layer 2 flooding across sites**

ACI Multi-Pod and ACI Multi-Site
Summary of Main Differences

Multi-Pod



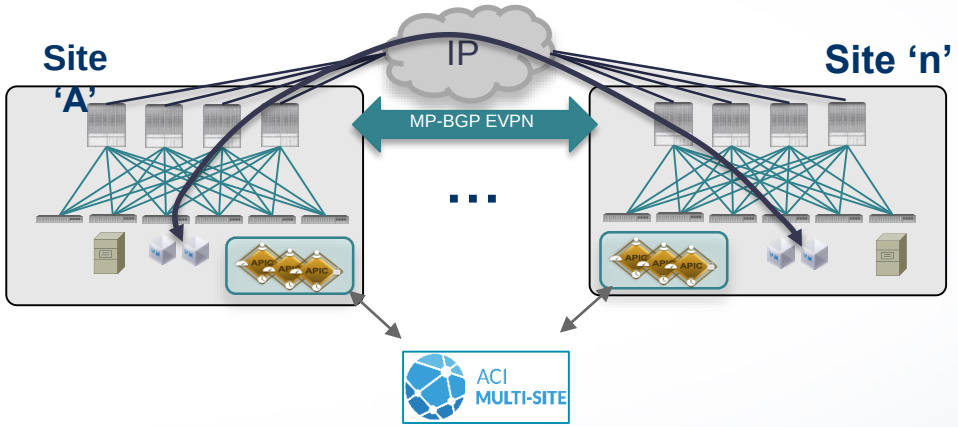
Operational Simplicity

Feature Richness across Pods

Lower Number of APIC Nodes

Single VMM across Pods

Multi-Site



Change Domain Isolation

Fabric Nodes Scale

High Latency across Sites

No Multicast required in the IP Network

New Hardware Support

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Nexus 9364C 64p 40/100G - ACI 40/100GE Fixed Spine

Ideal for space constrained fabrics

Support for mixed 1st & 2nd gen ACI leaf designs

Support for mixed 40/100G fabrics speed designs

100G line rate MACSEC and VTEP-VTEP overlay encryption on 16 ports*

40 MB buffer w/ smart buffer feature

Flexible TCAM templates
1M+ IPv4 routes

VXLAN Routing

QSFP28 Connector, Pin compatible with 40G QSFP+

Flexible Speed 64 ports with 1,10,25,40,50,100G

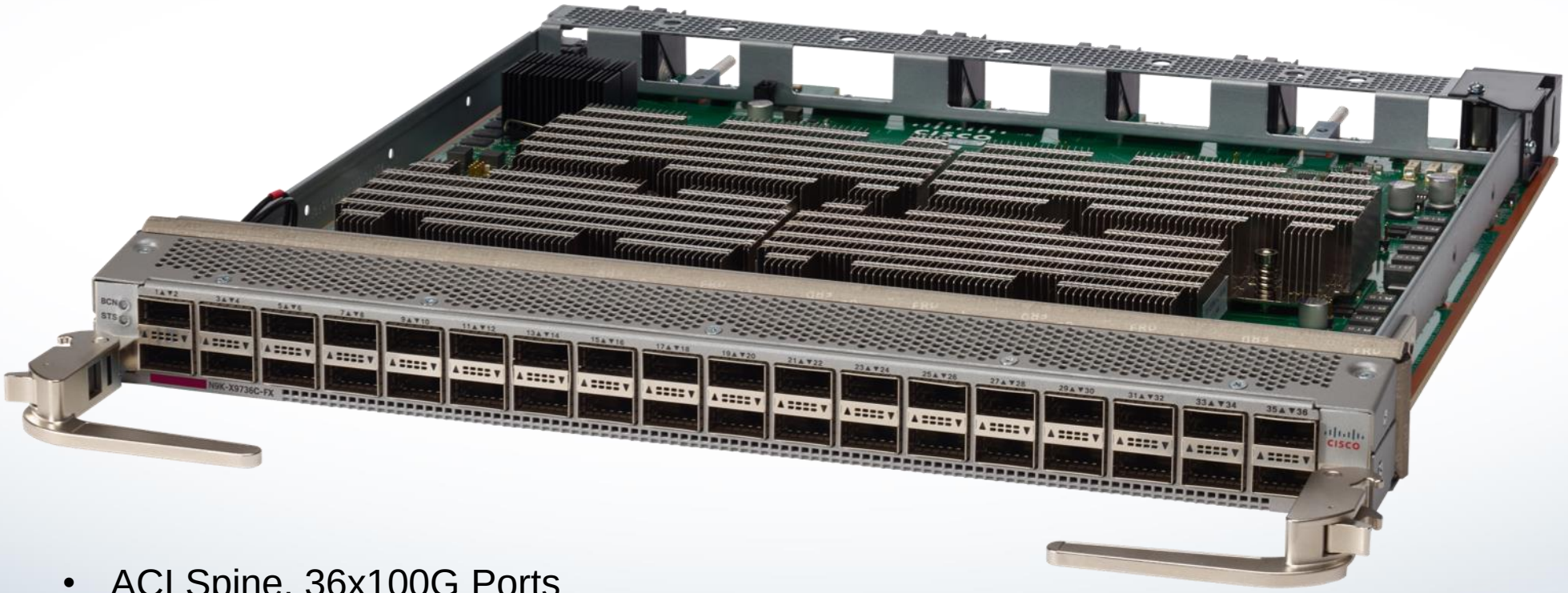
6.4 T full feature L2/3 ASIC

* future

Note: Roadmap: ACI 3.1 onwards 16 ports of MACSEC will be supported

Nexus 9500 Spine 40/100GE Linecard

N9K-X9736C-FX: 36p 40/100G



- ACI Spine, 36x100G Ports
- MACSEC and CloudSec (VTEP-VTEP encryption) capable

Nexus 9348GC-FXP

Data Center 100M/1G Access Optimized (NXOS/ ACI)

Nexus 9300 Fixed
1/10G Copper
(NX-OS / ACI)

Cisco Cloud Scale

N3048 Fixed Series
NX-OS

Merchant Silicon

N2248 Fabric Extender
NX-OS

Price Optimized
Feature Enhanced

Nexus 9348GC-FXP
48p 100M/1G Base-T +
4p 10/25G SPF+ +
2p 40/100G QSFP+
Dual 350W PSU

Nexus 9348GC-FXP

ACI Leaf: 48p 100M/1G, 4p 10/25G, 2p 40/100G

Dual 350W power supply for enhanced availability

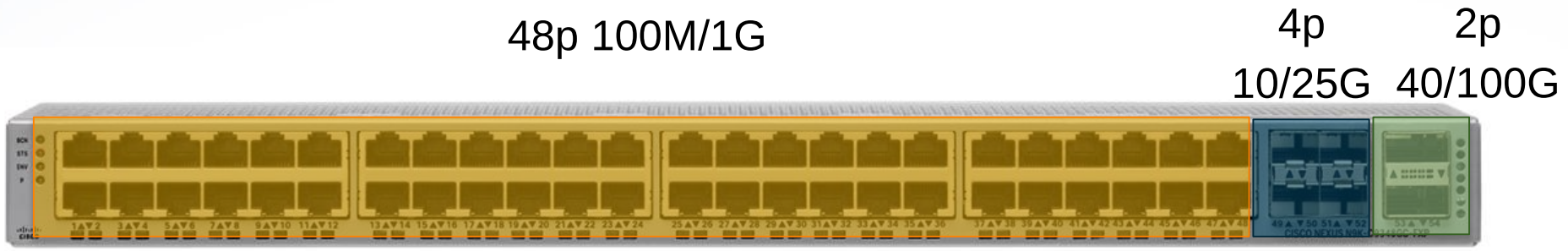
- Gigabit Ethernet application
- Up to 696 Gbps of bandwidth and 250+ mpps
- 2 and 4 post rack mount options

Flexible Speeds w/ 100M, 1GE and 10, 25, 40,100G uplink support

Full featured L2/L3 ASIC (Homewood)

40 MB buffer w/ smart buffer features

Nexus 9348GC-FXP Port Configuration



Port configuration supported:

- **Ports 1 – 48 support 100M/1G**, (*APICs cannot connect to these ports*)
- **Ports 49-52 support 10/25G**, (*APICs connectivity compatible only to these ports, can also connect to 10/25G host servers)*
- **Ports 53-54 are 40/100G uplinks** for spine connectivity

Additional Information on APIC Connectivity here

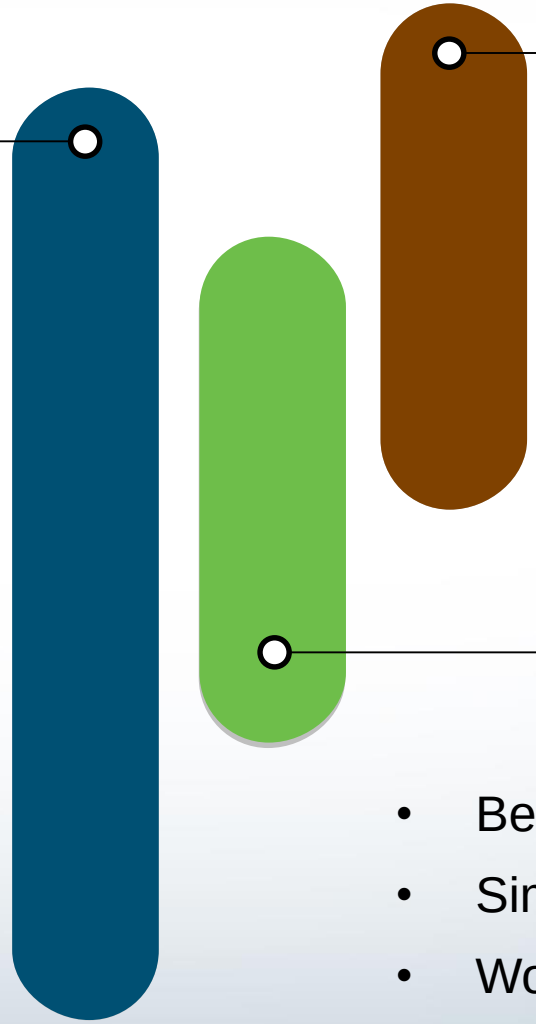
ACI 3.0 Optimizations

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

APIC GUI Enhancements – The Journey Begins

Usability

- New Look and Feel across Applications
- Consistent Layout across Tabs
- Collaborate by Sharing Objects
- Simplified Topology Views
- Release Bulletin
- Troubleshooting
- User Profiles
- Alerts



Operations

- Personalized User Profile
- Dashboard Widgets
- Improved Health Score and Fault Counts

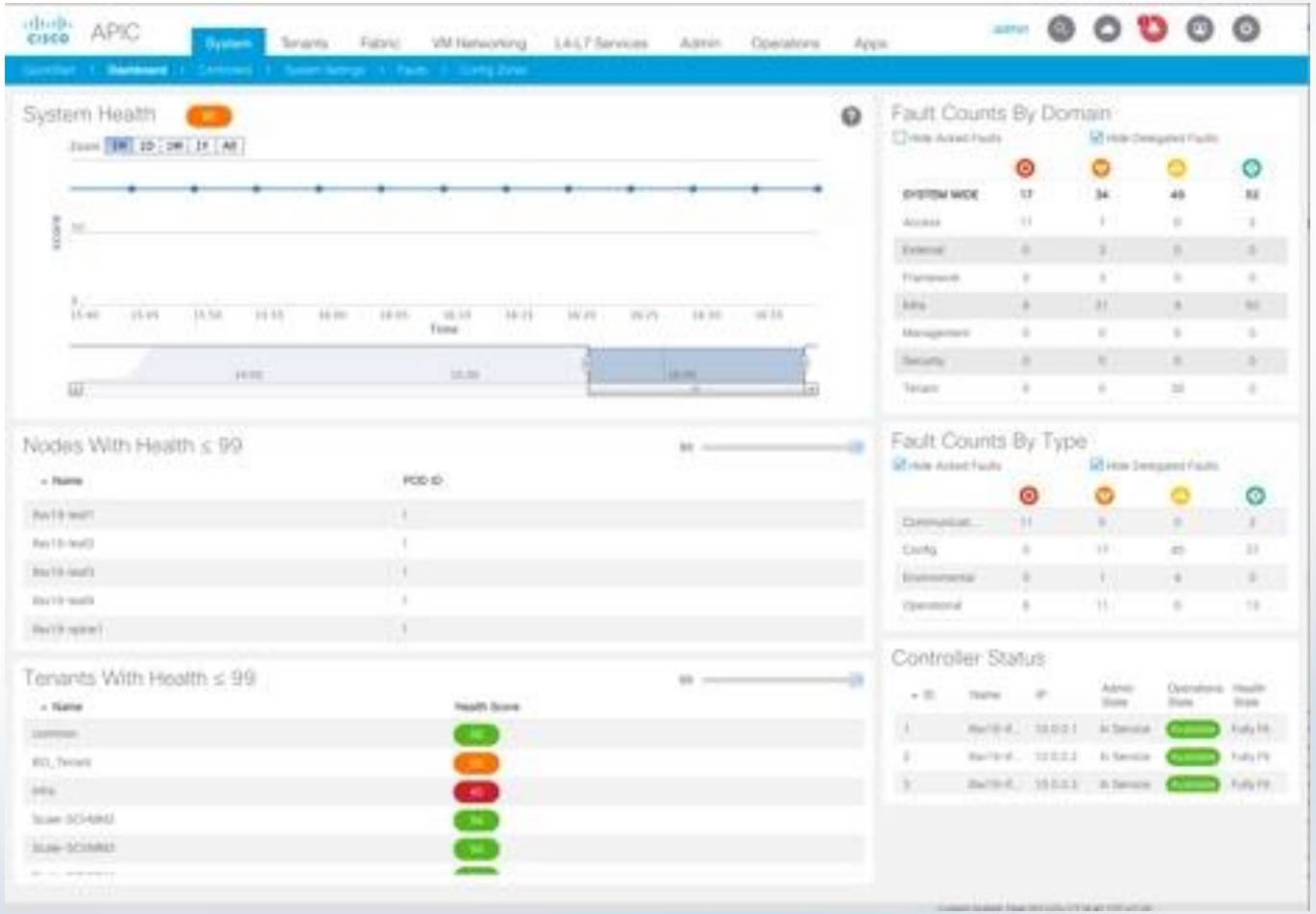
Configuration

- Best of both Basic and Advanced UI
- Simplified Port Selectors
- Workflows simplified
- New APIC Postman App

Update GUI Look and Feel – Cont.
Overview

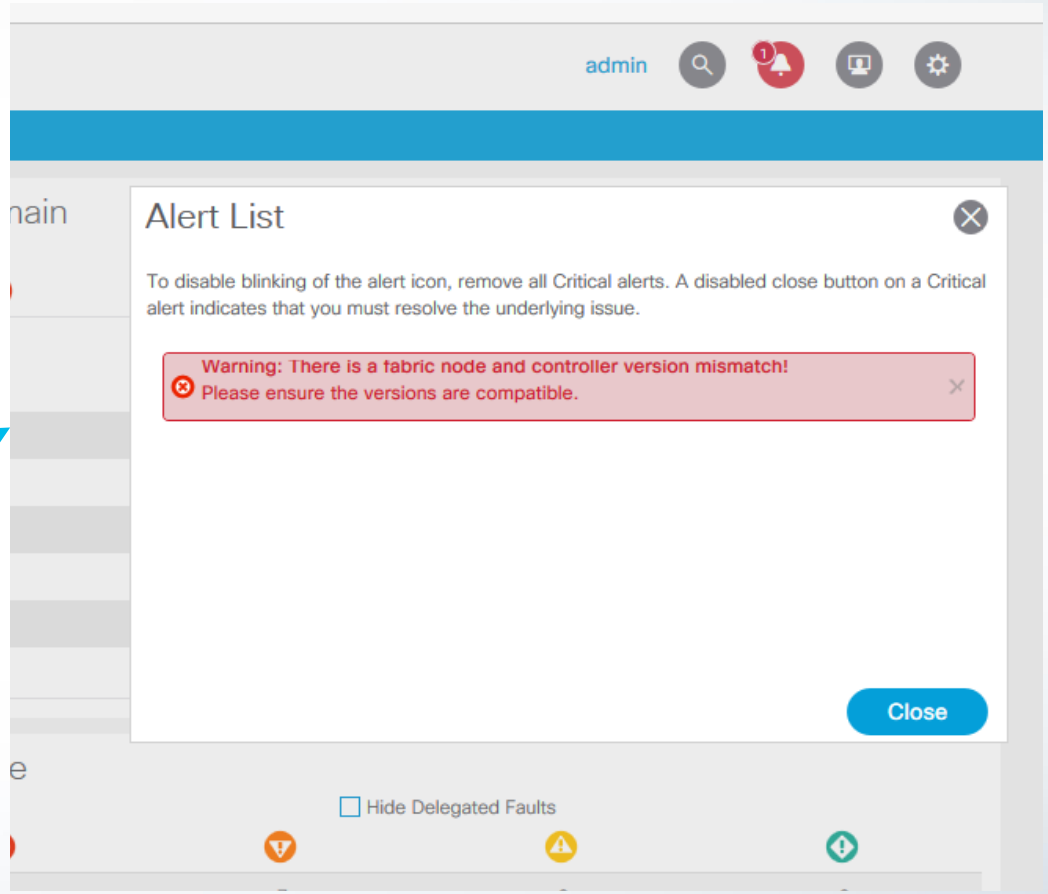
Earlier than APIC 3.0

APIC 3.0



Alert List

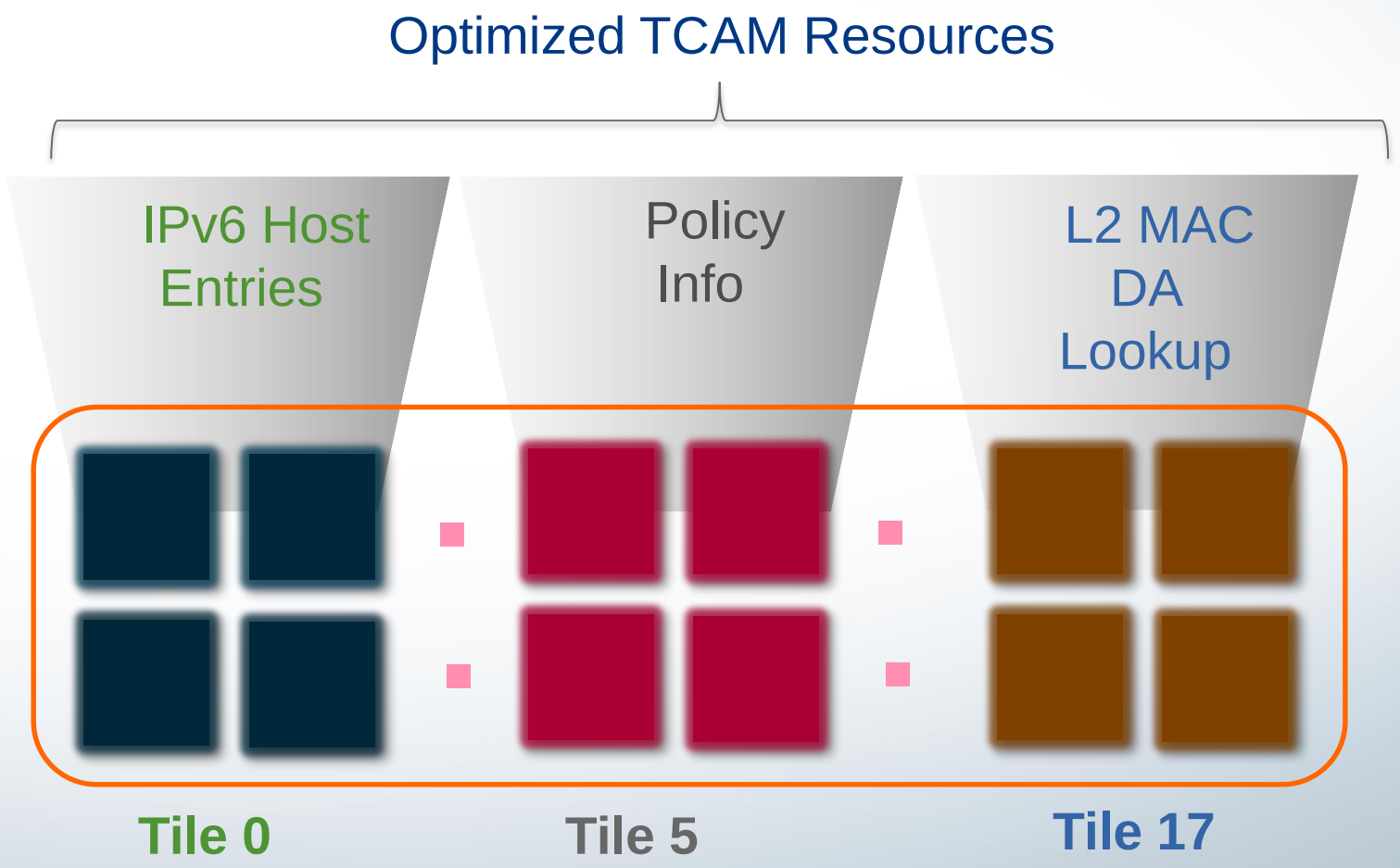
- Provides a list of any critical alerts that user should be aware
- Shows each alert when user logs in (can be turned off in user preferences mentioned earlier)
- Shows a flashing alert icon in the top menu
- All critical alerts shown when the user clicks the alert icon
- Critical alerts don’t go away until the issue that raised the alert has been resolved



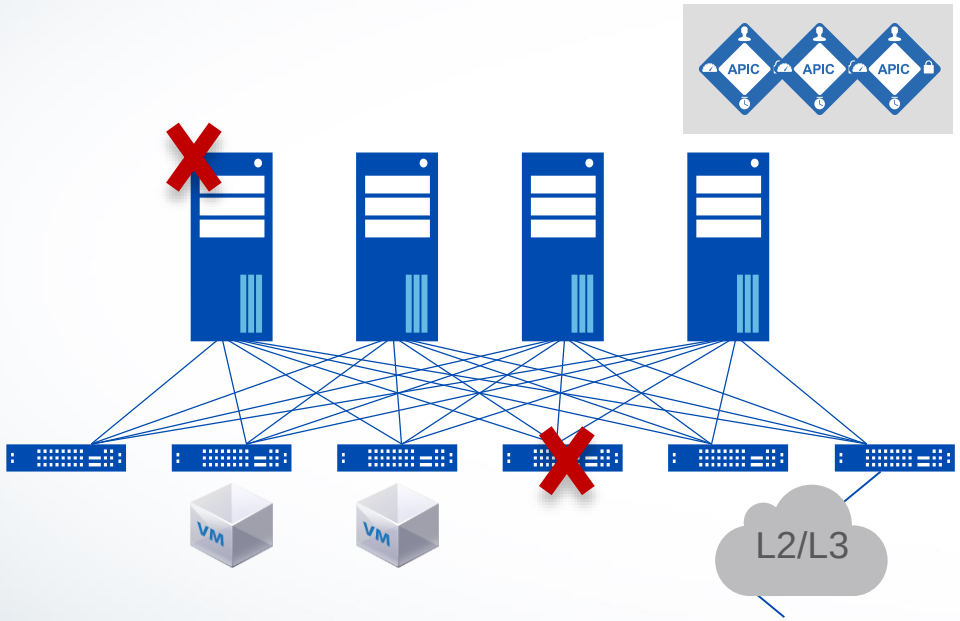
Optimizing Hardware Resources - Forwarding Scale Profile Policy

Flexibility To Choose TCAM Profile Based On Your Infrastructure Needs

- Profile 1: Dual Stack (Default)
- Profile 2: IPv4 Only
- Profile 3: Policy Optimized
- Profile 4: Multicast



Maintenance Mode to allow Graceful Insertion and Removal (GIR) of switches

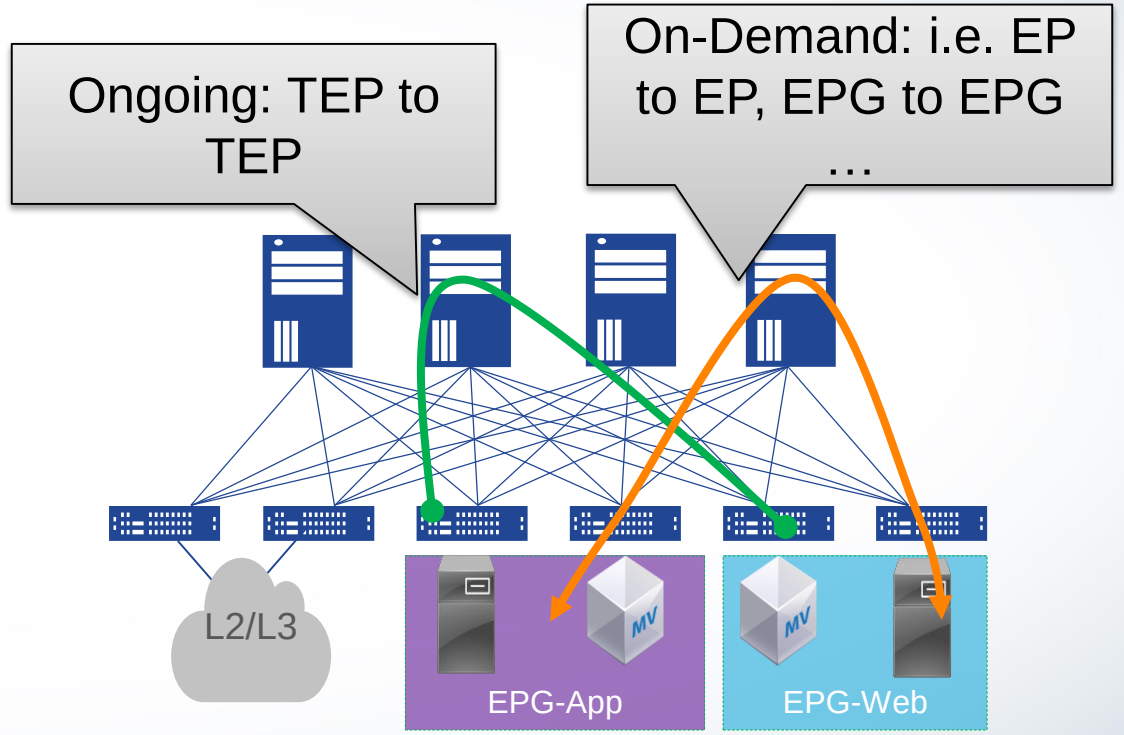


1. Gracefully isolate the node from fabric
2. Troubleshoot (if required)
3. Gracefully Re-commission the node

GIR diverts the data traffic to alternate paths and allows node troubleshooting, maintenance and upgrade.

Latency Monitoring (EX and FX switches in 3.0)

- New Troubleshooting tool to measure fabric latency
- **Differentiator**: No other Fabric or VXLAN SDN solution supports this
- Leverages **Precision Time Protocol** (PTP) to synchronize all the nodes in the fabric
- Two type of latency measurements:



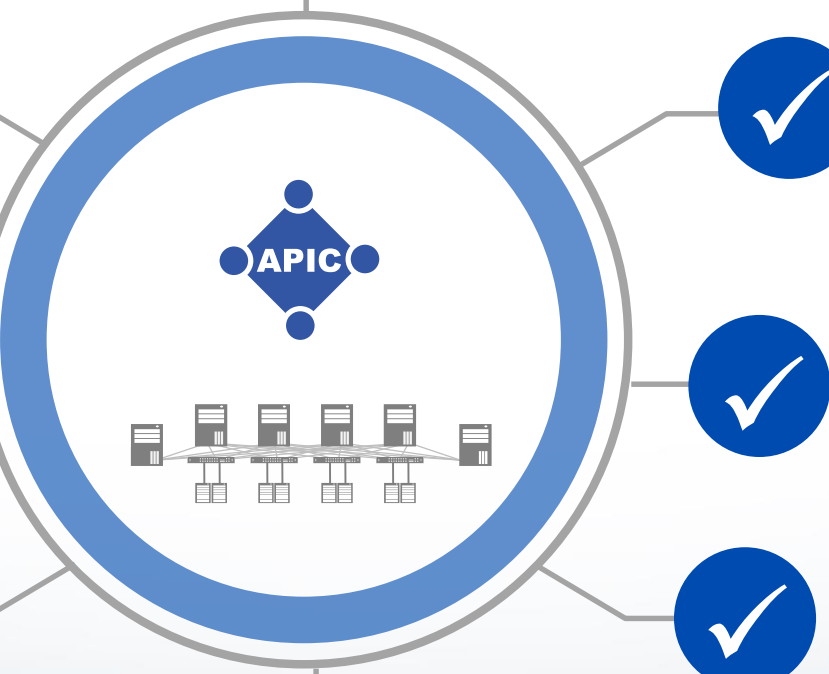
On-going: Measures latency across Tunnel End Points (TEPs)

On-Demand: Configured by the tenant to troubleshoot issues at the level of individual applications

On-Demand Latency

- Measurement is done for IP flows matching a Flow rule programmed in the Latency TCAM.
- Flow rules semantics are similar to the Atomic counter flow rules
- Both Atomic counters and Latency measurements can be independently enabled or disabled on the same IP flow rules.
- We can enable latency measurement for the below flow rules:
  - EP to EP
  - EP to EPG
  - EP to External IP
  - EPG to EP
  - EPG to IP
  - EPG to EPG
  - External IP to EP
  - Any to EP
  - IP to EPG
  - EP to Any

Verified Scale Improvements



Layer-3
50 VRFs Per Tenant , 1k IPs/MAC

Leafs
Up to 400 Per Fabric
8 Border Leafs per L3 Out

Tenants
Up to 3000

EPGs
Up to 15000
Up to 1k L3 EPGs/EX-Leaf
4k L3 EPGs for one tenant
& one context
250 Isolated EPGs

FEX
Up to 650 / Fabric
Up to 20 / Leaf

Bridge Domains
Up to 21,000 (L2), 15,000 (L3)
Up to 1750 Bridge Domains/VRF
3967 VLANs per leaf
3967 VLANs + BDs

Multicast Groups
Up to 8,000 (S,G) routes with
Convergence of 5 seconds

Other
Up to 200 vCenters
Up to 2,000 Contracts
Up to 61k TCAM Rules
500 Service Graphs Per Cluster
Up to 12 Pods in Multi-Pod

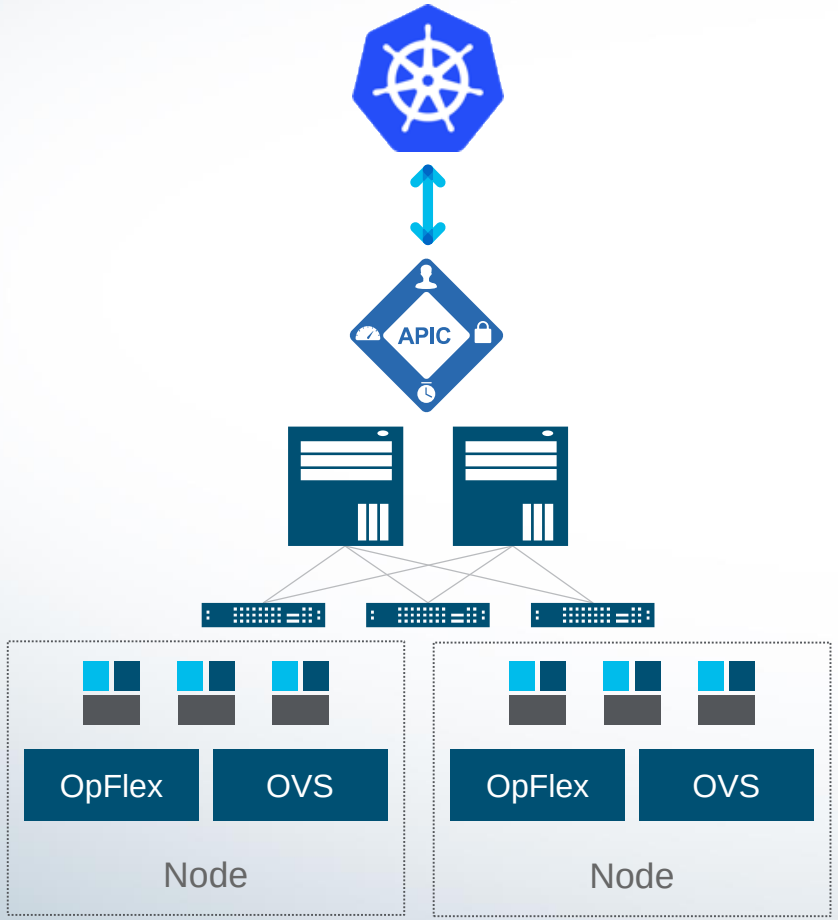
Integration with Kubernetes: Containers as First Class Citizens in IT Enterprise

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Container Cluster Managers

- Cluster multiple nodes (servers) that can run containers into logical units
- Schedule the placement of containers on cluster resources
- Provide service discovery tools for containers (who runs what where?
- Provide API services to other tiers and app developers

APIC Integration with Kubernetes



ACI and Containers

Unified networking: Containers, VMs, and bare-metal

Micro-services load balancing integrated in fabric for HA / performance

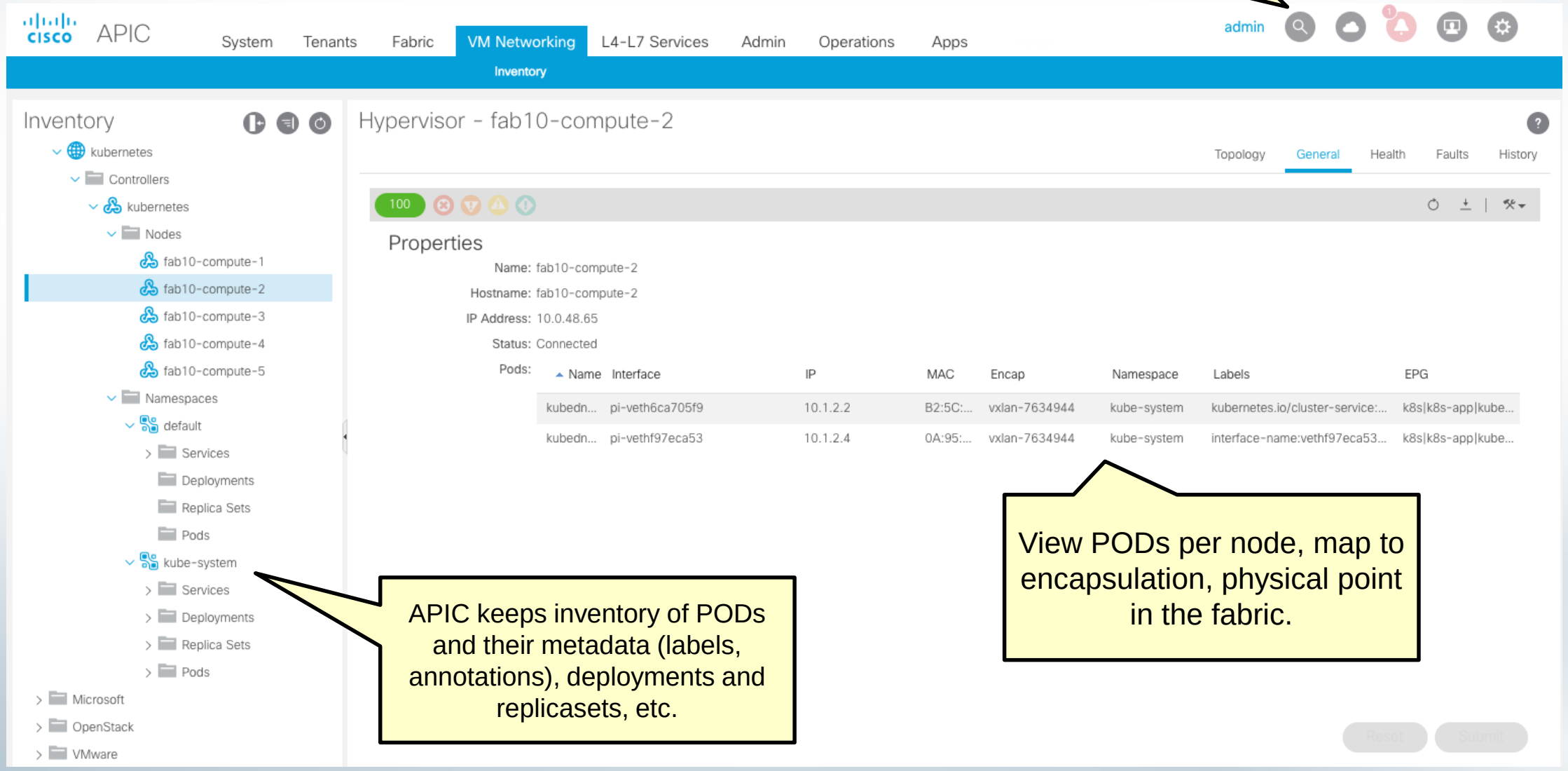
Secure multi-tenancy and seamless integration of Kubernetes network policies and ACI policies

Visibility: Live statistics in APIC per container and health metrics

Integration based on existing APIC, no need for a new controller!!!

Kubernetes VMM Domain

Fabric admin can search APIC for k8s nodes, masters, PODs, services …

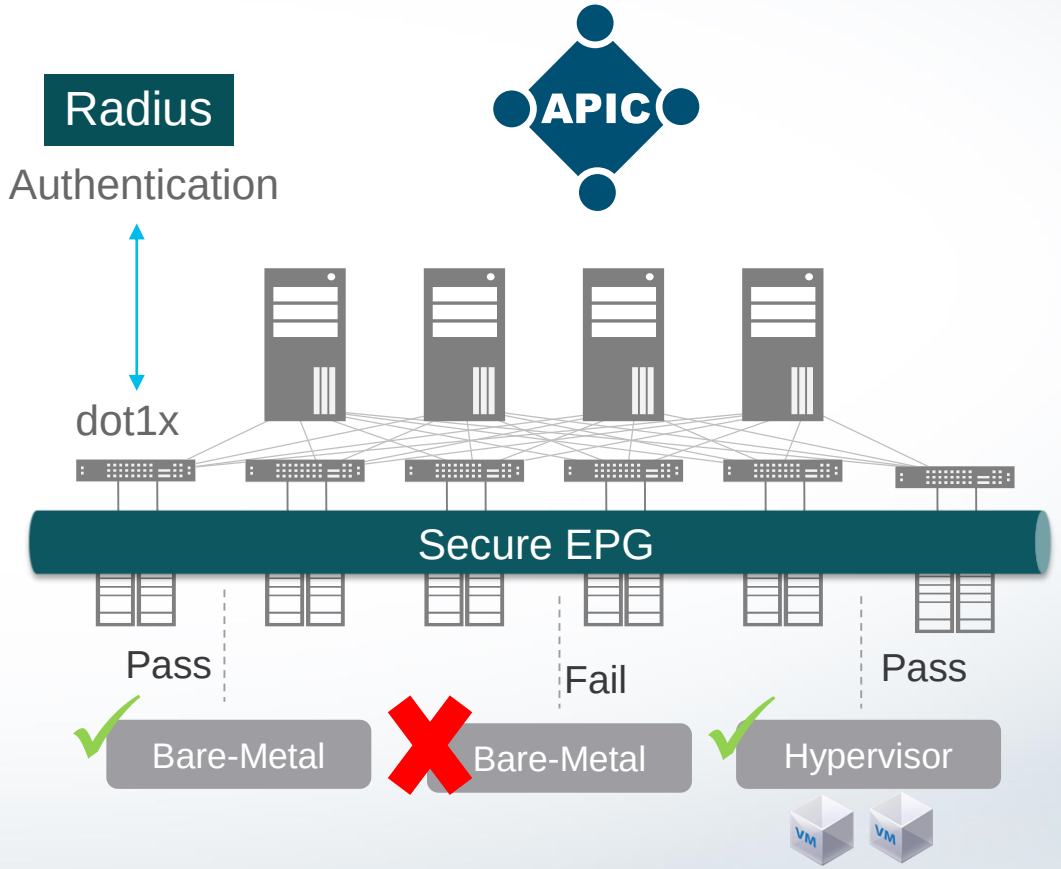


APIC keeps inventory of PODs and their metadata (labels, annotations), deployments and replicasets, etc.

View PODs per node, map to encapsulation, physical point in the fabric.

Zero Trust Security
Dot1X Authentication

- ✓ End Point Authentication For EPG Classification
- ACI 3.0 — Supported on Bare-metal Only
- ACI 3.0 — Supported on '-EX' & '-FX' Leaf
- Futures — Hypervisor & Container Workloads



Admit Only Authenticated Endpoints Into Secure EPG

ACI 1st Hop Security Features

- 5 features supported for IPv4/v6
  - BD Level
    1. DHCP Snooping/Inspection
    2. Dynamic ARP/ND Inspection
    3. IP Source Guard
    4. RA Guard v6
  - EPG Level
    1. Trusted EPG
- FHS policy is implemented on Leaf switch only
- ACI 3.0 supports PhyDoms only
- VMM domain support will be in future (vDS requires PVLAN)

Support for Intra-EPG Contracts

Before the Ebro release intra-EPG traffic is either:

- Allow all (by default); or
- Deny all (when intra-EPG isolation is enabled)

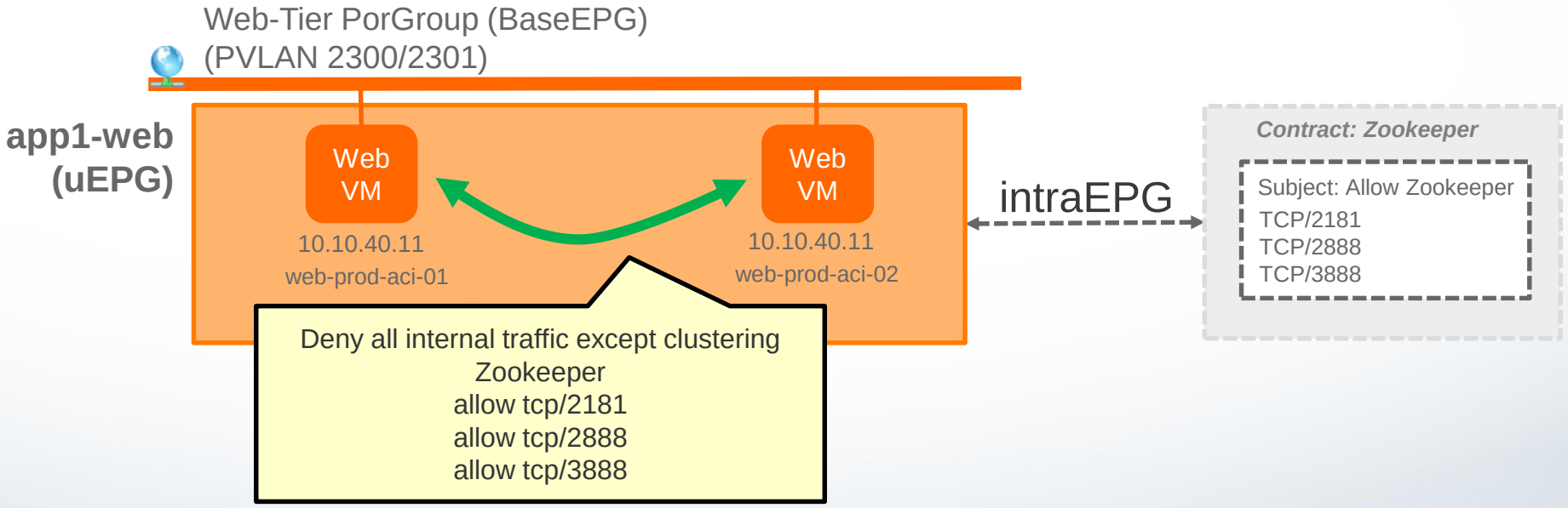
We needed a way to define the policy for intra-EPG traffic to enforce an “intra EPG contract.” Something like “permit ICMP and ssh” between all the end points in an EPG.

Why do we need intra-EPG security?

- Customers want a way to restrict the communication between any two endpoints in an EPG
- Although micro-segmentation EPGs can be used to satisfy this requirement, it involves complicated configurations and micro EPGs should be used to classify endpoints with variable attributes.

Support for Intra-EPG Contracts

Example: a cluster of web applications running Zookeeper. You want only the required protocols to be allowed between the VM inside the portgroup, or between the servers in a VLAN, etc.



Supported on Bare Metal and VMware VDS with ACI 3.0 (only EX/FX switches)

BGP Multi-Path
Overview

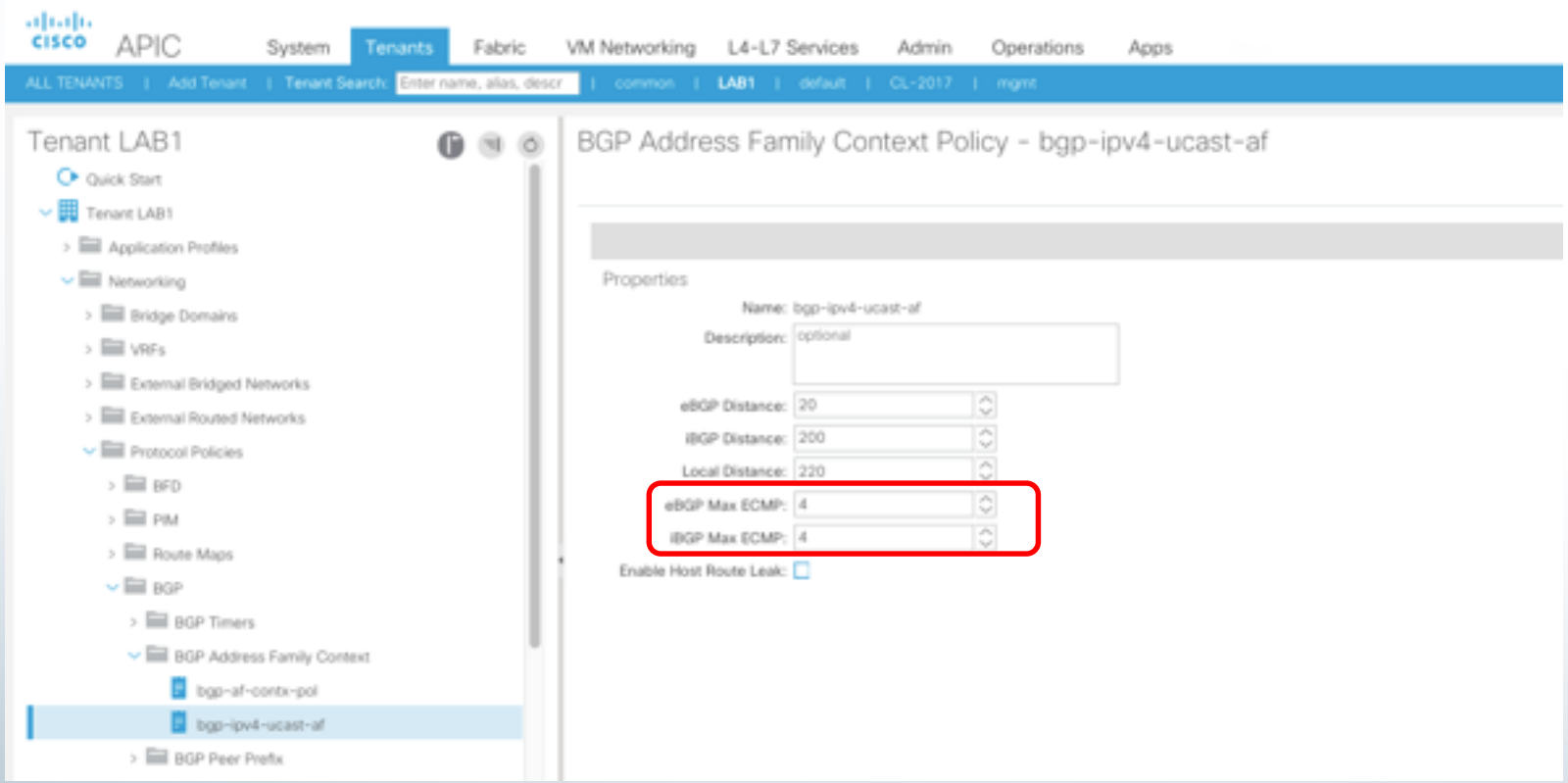
- BGP typically selects only one best path for each prefix to install in the forwarding table. When BGP multipath is enabled, the device selects multiple equal-cost BGP paths to reach a given destination prefix, and all these paths are installed in the forwarding table for ECMP (equal-cost multi-path) load sharing

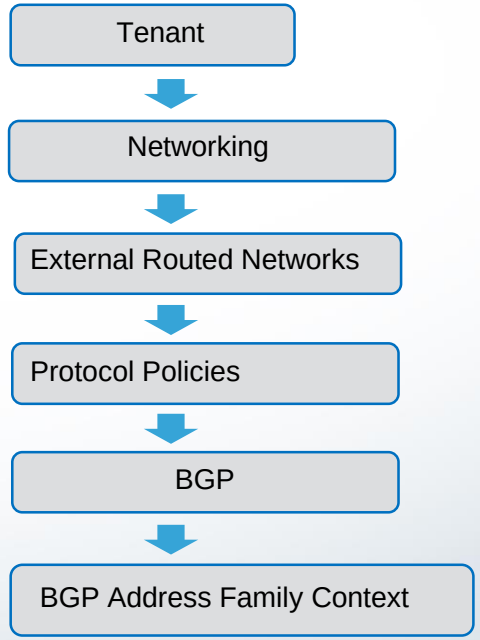
- BGP multipath allows to install multiple internal BGP paths and multiple external BGP paths to the forwarding table for load sharing. The maximum number of ECMP BGP routes is 16

- For this feature, two new properties, maxEcmp (for eBGP) and maxEcmpIbgp (for iBGP) are added in bgpCtxAfPol MO. They are to configure the maximum number of paths that BGP selects and installs into the forwarding table. They are configured for each address family. The default setting is 16

Configuration
APIC GUI

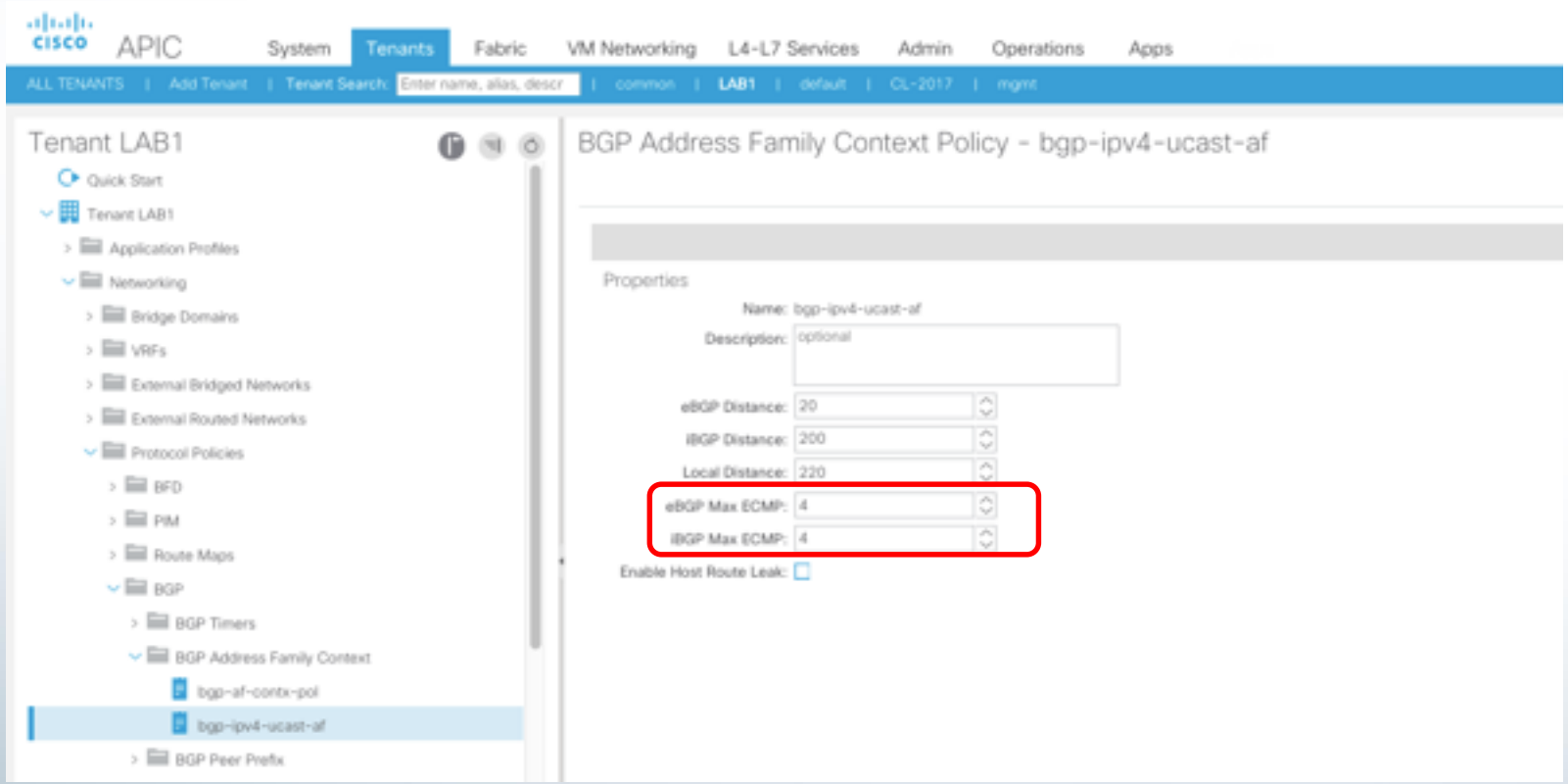
- Configure the maximum number of ECMP paths for eBGP and iBGP in BGP Address Family Context:

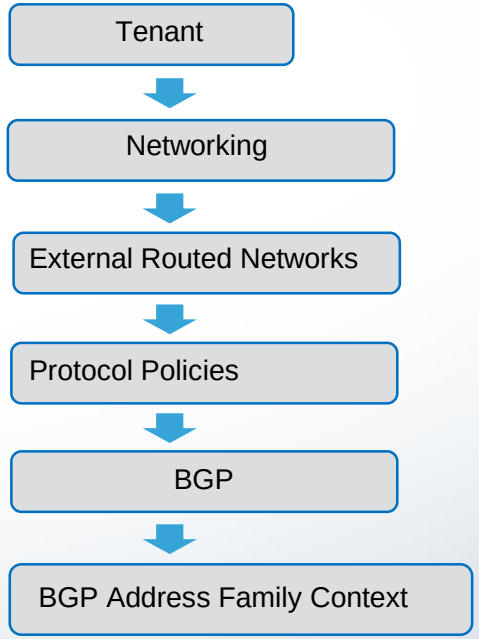




Configuration (Cont.)
APIC GUI

- Apply the BGP Address Family Context to VRF:





Q-in-Q to EPG Mapping
Overview

- Map two dot1q tags (combined 24 bits value of inner and outer VLAN ID) of incoming QinQ frame to an EPG
- Support of dualEncap port mode on ACI front panel interface
- Support of QinQ encap for static path binding of a dualencap port to an EPG, including:
  - Static Ports
  - Static Leafs
  - Static Endpoints
- Only QinQ traffic is allowed on dual Encap port. Single tagged and untagged traffic would be dropped
- Requires FX hardware

Cisco ACI 3.0 Release

Infrastructure

Hardware

- Fixed Spine N9K-9364C – No Multipod, Multisite & MACSEC support at FCS
- N9K-X9736C-FX with 9504/08/16 chassis support
- N9K-C9348GC-FXP (48p 10/100/1000, 4p 10/25 G & 2p 40/100G port)
- N9K-C9508-FM-E2 – 8 slot Fabric module

Features

- Multi Site v1.0
- Maintenance Mode (Graceful decommission)
- Forwarding Scale Profile
- Intra EPG contracts for DVS and Bare Metal
- First hop security (ARP, ND & DHCP Inspection, RA Guard, IPv4/v6 Source Guard): EX platform onwards
- SAML Integration with OKTA IDP & Microsoft ADFS – 2 factor authentication
- Q-in-Q Mapping to EPG
- 802.1 x support (Single host mode)
- Ingress QoS policing – per EPG per interface policing (Interface, EPG/VLAN)
- AS Path prepend
- BGP Multi-Path
- Enforced VRF Bridge Domain (EX and FX only)
- Latency & Precision Time Protocol (PTP)
- Password-less Authentication
- APIC GUI Overhaul

Tetration

- TEP ID and additional MO support

Virtualization and Operations

- Logical Operators (AND/OR) for attribute-based uSeg - HyperV
- Intra EPG Isolation for HyperV
- VMM Domain / OpFlex support for Kubernetes
- AVS: QoS Marking
- AVS: Distributed Netflow
- Delay Endpoint Detach

With Cisco ACI, you can build a better network…

anywhere.