Cisco Nexus One Architecture White Paper

Contents

Instructions for the reader 3

Introduction 5

Nexus One Architecture Overview 5

Managing Modern, Heterogeneous Data Center Fabrics 5

Operational Diversity in the Modern Data Center 6

Aligning Security and Policy Models 6

The Burden of Inter-Fabric Connectivity 6

Scalability, Failure Domains, and BUM Traffic 7

Automation and API Fragmentation 7

Use Cases 7

Disaster Recovery and Active-Active Data Centers 7

Unified Micro-Segmentation and Zero-Trust 8

L2/L3 Extension and Workload Mobility 8

Inter-Fabric Service Insertion and Chaining 8

Policy-Consistent Brownfield and Greenfield Expansion 9

Nexus One Innovations: Overview 9

The Cisco ACI Border Gateway: Opening Up ACI 9

The Genesis of the ACI Border Gateway 9

Key Innovations of the ACI BGW 9

NX-OS VXLAN EVPN Security Groups (Group Policy Option): Policy for All 10

The Evolution of Security in NX-OS 10

Key Innovations of Security Groups (GPO) 11

The Unified Nexus Dashboard: A Single Point of Control 12

The Need for Unified Management 12

Key Innovations of the Unified Nexus Dashboard 12

Nexus One Innovations: The Architectural Pillars Deep Dive 13

The Cisco ACI Border Gateway (BGW) 13

Role, Placement, and Hardware Requirements 13

Control Plane and Data Plane Overview 13

Data Plane Mechanics: Tunnel Stitching and Forwarding 18

Namespace Normalization: Symmetric and Asymmetric 19

Policy Enforcement 22

Distributed Anycast Gateway Across Domains 51

NX-OS VXLAN EVPN with Security Groups 52

The Emergence of VXLAN Group Policy Option 52

Core Use Cases of VXLAN GPO 53

Fundamental Constructs of VXLAN GPO 53
VRF Enforcement Modes 54
Operational Principles: Control and Data Plane 54
The Unified Nexus Dashboard 56
End-to-End Connectivity and Policy Enforcement 57
Control Plane Reachability Advertisement – Step by Step 65
The control plane journey for endpoints and external networks learned in the NX-OS fabric 67
The control plane journey for endpoints and external networks learned in the ACI fabric 97
The data plane journey for endpoints between ACI and NX-OS – Step by Step 117
Service Redirection 126
ACI Multi-Site With BGWs 128
The Evolution of ACI Multi-Fabric Architectures 128
A New Approach: Multi-Site via the ACI Border Gateway 129
Extending VRFs, Networks, and Security Groups Across Fabrics 130
Requirements 133
Caveats and Limitations 133
Deployment Considerations and Best Practices 134
Hardware, Software, and Licensing Prerequisites 134
Topology Design and BGW Placement 134
Control Plane and Connectivity Best Practices 135
Security Policy Design 135
Operational Considerations and Constraints 135
Conclusion 136
Acronyms 136
References 141

Instructions for the reader

This white paper is an architectural and technical deep dive designed to explain the concepts, mechanics, and design principles of the Nexus One Fabric solution, focusing on the interconnectivity between ACI and NX-OS VXLAN EVPN fabrics. It does not serve as a step-by-step configuration guide. For detailed configuration examples and command-line instructions, please refer to the separate Nexus One Implementation and Configuration Guide.

This document is comprehensive in scope. Due to the depth and breadth of the topics covered, ranging from high-level architectural vision to bit-level packet analysis, we recommend that readers approach this document based on their specific role and technical objectives.

For IT Leaders and Network Architects (High-Level Overview): If your primary goal is to understand the strategic value, architectural components, and use cases of Nexus One without getting bogged down in packet-level mechanics, focus on the following sections:

- Introduction: For the vision and the "Why" behind Nexus One.
- Nexus One Innovations: To understand the three pillars (ACI Border Gateway, NX-OS Security Groups, and Nexus Dashboard) and how they solve heterogeneity.
- Use Cases: To see practical applications for the solution.

For Network Engineers, Security Architects, and Implementation Specialists (Technical Deep Dive): If you are tasked with designing, deploying, troubleshooting the solution, or defining the security posture and Zero Trust model, the entire document is relevant. Attention should be paid to the technical mechanics and policy enforcement mechanisms:

- Nexus One Innovations: Read this for a detailed breakdown of the ACI Border Gateway hardware, roles, and control plane.
- End-to-End Connectivity and Policy Enforcement: This is the core technical chapter. It details the control plane and data plane interactions, including specific packet walks and troubleshooting outputs.
- Deployment Considerations and Best Practices: Critical for Day-0 planning, covering software baselines, topology constraints, and operational pitfalls.

Introduction

Nexus One Architecture Overview

Modern data centers are evolving quickly due to the growth of cloud-native applications and hybrid, multi-cloud architectures. This shift has moved infrastructure beyond monolithic designs toward dynamic, distributed ecosystems. In these complex scenarios, organizations often operate a diverse mix of data center fabrics. These may encompass highly programmable Cisco Application Centric Infrastructure (ACI) deployments, robust standard VXLAN EVPN fabrics running on Cisco NX-OS, and increasingly, specialized cloud-managed fabrics, such as Cisco Nexus Hyperfabric, or integrated campus networks. The main challenge is unifying connectivity, security, and management across these different domains to ensure interoperability and consistent policy enforcement.

Cisco Nexus One architecture emerges as Cisco's strategic architectural response to this challenge. Instead of just being a single product or a set of features, it is a comprehensive architectural framework designed to converge network connectivity, security policy, and operational management across all major Cisco data center fabric types. All this with a clear roadmap for the future inclusion of campus and public cloud-managed domains, as well as extensions to seamlessly integrate with third-party technologies, and networks such as hyperconverged solutions and CNIs.

At its core, Nexus One uses open networking, standards-based protocols, and a unified policy model to help organizations transform fragmented infrastructure into a cohesive, automated system.

The foundational vision of Nexus One is to deliver an Open NEtworking (ONE) experience that is inherently consistent, scalable, and secure. We achieve this through three key architectural pillars:

- Border Gateway (BGW): Border Gateways (BGWs) are VXLAN EVPN VTEPs designed to interconnect independent VXLAN EVPN data center fabrics by managing the exchange of reachability information and facilitating the forwarding of encapsulated frames between them. While this functionality has been a cornerstone of NX-OS environments since 2017, a recent ACI update now brings this standards-based BGW capability directly into ACI fabrics. This integration allows ACI to seamlessly extend Layer 2 and Layer 3 connectivity, alongside policy, to other BGW-enabled VXLAN EVPN fabrics, effectively enabling ACI to interoperate seamlessly across diverse, multi-technology environments.
- Endpoint Security Groups: Security Groups provide a robust framework for micro-segmentation, utilizing contract-based enforcement to control traffic flow between endpoints. This approach enables sophisticated service chaining, ensuring that security policies are applied consistently regardless of the underlying topology. Recent NX-OS versions have introduced the Security Groups functionality based on the Group Policy Option (GPO) standard, which leverages these capabilities to extend the policy enforcement for both intra-site and inter-site traffic, fostering a unified policy language across multi-technology environments.
- Unified Nexus Dashboard: Positioned as the central point of control and operations for all interconnected fabrics, the Unified Nexus Dashboard is designed to abstract away the burden of managing heterogeneous environments. It offers fabric-agnostic provisioning, comprehensive end-to-end visibility, automated troubleshooting, and a unified API gateway, thereby streamlining operational workflows and enhancing overall efficiency.

Managing Modern, Heterogeneous Data Center Fabrics

The evolution of data center architectures reflects the diverse requirements of modern enterprises. Organizations have strategically adopted various fabric types, such as ACI for policy-driven automation and

NX-OS for high-performance, programmable infrastructure, to meet specific application and business needs. As these environments scale, the priority shifts from managing individual fabrics to achieving operational harmony and connectivity across the entire ecosystem. The Nexus One architecture is designed to provide this unified layer, building upon the strengths of these established technologies to deliver consistent management, connectivity, security, and automation.

Operational Diversity in the Modern Data Center

Modern data centers often use specialized tools to optimize different environments, such as ACI, NX-OS, and campus networks. While this approach is highly efficient for individual fabrics, it creates operational silos that are difficult to manage. Each environment brings its own unique operational paradigm. Nexus One bridges these disparate environments. By providing a unified management layer, it reduces the overhead of managing multiple toolsets. This allows IT teams to shift their focus from platform-specific procedures to strategic business outcomes.

Aligning Security and Policy Models

Each fabric type offers distinct security constructs designed to segment traffic and enforce policy. The Nexus One architecture provides a unified, standard policy framework that translates these distinct security constructs into a cohesive posture. By abstracting the underlying differences, it ensures that security policies remain consistent as workloads move across the infrastructure, simplifying compliance and auditing without requiring the manual translation of policies between environments.

The Burden of Inter-Fabric Connectivity

Connecting different fabrics has traditionally been difficult and fragile. Traditional inter-fabric connectivity methods often introduce operational and architectural risks:

- Layer 2 VLAN Handoffs: Extending VLANs between distinct fabrics, while seemingly straightforward, creates the risk of large, unmanageable broadcast domains. This necessitates complex Spanning Tree Protocol (STP) topologies, which are notoriously difficult to design, implement, and troubleshoot. Critically, a broadcast storm originating in one fabric can easily cascade across the entire interconnected network, leading to widespread outages and severe business disruption.
- Layer 3 VRF-Lite: While offering improved isolation compared to Layer 2 extensions, VRF-Lite requires the establishment of a separate routing peer for every Virtual Routing and Forwarding (VRF) instance that needs to be extended across fabrics. This approach is inherently unscalable, particularly in multi-tenant environments with numerous VRFs. It leads to massive configuration complexity on border devices, making the network exceedingly difficult to manage, troubleshoot, and maintain over time.
- Traditional Data Center Interconnect (DCI) Technologies: Solutions such as Overlay Transport Virtualization (OTV), while effective in specific contexts, introduce an additional layer of technology to manage. Furthermore, they are not natively integrated with the advanced policy and automation models characteristic of modern fabrics, leading to disjointed operational workflows.
- Distributed Default Gateway Challenges: Traditional inter-fabric designs often struggle to support a distributed default gateway architecture, which is crucial for optimal east-west traffic flow and seamless host mobility. Inconsistent gateway placement and lack of native first-hop redundancy across fabrics can result in suboptimal routing, increased latency, and complicated failover scenarios, further increasing operational risk and complexity.

These traditional methods often fail to meet the demands of modern data centers. They lack the scalability, simplicity, and policy awareness needed for secure multi-fabric environments.

Scalability, Failure Domains, and BUM Traffic

Modern, distributed applications and microservices generate high volumes of east-west traffic between servers, VMs, and containers. This traffic is frequently latency-sensitive and must traverse multiple fabrics, often spanning geographically dispersed data centers or even continents. Traditional network designs often struggle with this reality:

- Some designs indiscriminately stretch Layer 2 domains, which, while simplifying workload mobility, dramatically increases the size of the failure domain. A localized issue, such as a misconfigured application or a software bug, can then impact the entire stretched network.
- Conversely, other designs route everything at Layer 3. While providing better isolation, this approach may not support seamless workload mobility or certain security requirements that necessitate Layer 2 adjacency.

A major challenge in stretched Layer 2 domains is managing BUM traffic. Without strict control, BUM storms can consume network resources and cause outages. In an environment with multiple fabrics, the risk is amplified, as without a proper control point at the fabric edge, a BUM storm originating in one fabric can easily flood across the inter-fabric network, consuming valuable bandwidth and potentially causing network instability in remote fabrics. Troubleshooting such events across disparate, interconnected fabrics is notoriously difficult, as the source of the storm may be hidden within a remote domain.

Automation and API Fragmentation

Infrastructure-as-Code (IaC) and DevOps rely on consistent, well-documented APIs. However, different API standards across fabric types create significant integration hurdles:

- Inconsistent APIs: Each fabric often exposes unique APIs, such as ACI's and ND's REST API compared to NX-OS's NETCONF/RESTCONF. This forces teams to maintain separate automation scripts for each fabric, increasing operational complexity.
- Lack of Unified Abstraction: Without a common abstraction layer, engineers must master the specific data models and APIs of every fabric. This steepens the learning curve, reduces productivity, and limits the scope of automation.
- Integration Challenges: Fragmented APIs hinder integration with orchestration tools, security platforms (like SIEM/SOAR), and public cloud management systems. This fragmentation prevents true end-to-end automation of application deployment and security policy enforcement.

Use Cases

The Nexus One architecture provides a unified operational and policy framework to seamlessly extend reachability information and connectivity between networks running on different technologies. It addresses key data center challenges by transforming disparate network domains into a single, cohesive infrastructure.

This powerful integration enables several key operational models. The following sections highlight the most common and validated use cases where this architecture delivers significant value.

Disaster Recovery and Active-Active Data Centers

Nexus One provides the architectural foundation for robust business continuity, seamlessly bridging disparate fabrics. This makes it possible to design recovery strategies that are independent of the underlying network technology at each site.

- Active-Active Deployments: Applications can run concurrently across an ACI fabric in one data center and an NX-OS fabric in another. Traffic is load-balanced between these active instances, ensuring continuous availability and optimal performance.
- Active-Passive (DR) Deployments: For disaster recovery, the L2/L3 extension capabilities allow a primary site to fail over to a secondary one, even if they run on different technologies. Applications are recovered with their existing network and security configurations intact, minimizing downtime and operational complexity.
- Workload Orchestration Integration: The architecture loosely integrates with higher-level orchestration tools to automate these failover and failback processes, ensuring a smooth and reliable recovery across the unified multi-fabric domain.

Unified Micro-Segmentation and Zero-Trust

Nexus One's most powerful capability is its ability to deliver a unified micro-segmentation policy that decouples security from network topology. Security is now based on logical, identity-based Endpoint Security Groups (ESGs).

An endpoint is classified into a group, and its identity is embedded and propagated within the network control plane and data packets. Policy is then defined through contracts that explicitly permit traffic between these groups. This "whitelist" approach, where all traffic is denied by default, is the cornerstone of a Zero-Trust security model, ensuring that only authorized communication can occur and significantly reducing the attack surface.

Crucially, policies are enforced at the earliest possible point in the data path. Ideally, this occurs at the ingress leaf switch. By learning the SGTs of both source and destination endpoints, the leaf can drop unauthorized traffic immediately, preventing unwanted packets from consuming fabric bandwidth and ensuring security is applied with maximum efficiency.

L2/L3 Extension and Workload Mobility

The current Nexus One architecture allows Layer 2 bridge domains and Layer 3 VRFs to be stretched across fabrics running different technologies. The initial support will be offered on NX-OS-based VXLAN EVPN fabrics and ACI. This capability is foundational to its multi-tenant design, as each stretched VRF provides complete routing and policy isolation for a specific tenant, even supporting overlapping IP address spaces.

This capability is the foundation for true workload mobility. A key feature is the unification of the ACI Pervasive Gateway with the standard VXLAN EVPN Distributed Anycast Gateway (DAG), allowing them to interoperate as a single, logical gateway within the same broadcast domain. As a result, an endpoint's default gateway is always active and local within its fabric, which optimizes traffic paths and allows virtual machines to be live migrated between sites without any IP address changes.

To prevent traffic disruption during endpoint migration, it is critical to maintain MAC address parity by ideally standardizing the VXLAN NX-OS DAG on the ACI default gateway MAC, or alternatively, by modifying the ACI Bridge Domain's custom MAC to match the NX-OS environment, thereby eliminating the need for endpoints to re-learn their gateway and ensuring a truly seamless mobility experience.

Inter-Fabric Service Insertion and Chaining

The architecture provides a framework where the distinct service insertion models of each fabric, like ACI Policy Based Redirection (PBR) and NX-OS Enhanced Policy-Based Redirection (ePBR), can coexist and operate effectively within the unified domain.

This allows each fabric to maintain its own independent, fabric-centralized service chain. For example, traffic within ACI can be steered through a local firewall using a Service Graph, while traffic in the NX-OS fabric is directed to a different set of services using its native PBR policy.

The key benefit of Nexus One is that the trigger for this redirection remains consistent across the entire solution. This is possible because all endpoints, regardless of their location, are classified into a common set of Security Groups-based on consistently and centrally defined rules.

As a result, if an endpoint moves from the ACI fabric to the NX-OS fabric, it retains its Security Group identity. This ensures that all associated policies, including complex service redirection rules, are seamlessly maintained, guaranteeing that the correct security posture is enforced automatically and without any manual intervention.

Policy-Consistent Brownfield and Greenfield Expansion

Nexus One excels at integrating existing networks with new deployments. The key to this is the sophisticated namespace normalization function performed by the ACI Border Gateways.

The BGWs automatically translate not only network identifiers (VNIDs) but also security identifiers (SGTs) between fabrics in both control and data planes. For example, an SGT from an NX-OS fabric can be mapped to a corresponding ESG tag in ACI, allowing policy to remain consistent even if the underlying numbering schemes are different. This enables organizations to extend a unified policy model to existing infrastructure without disruptive re-engineering.

Nexus One Innovations: Overview

The Nexus One architecture is a significant step forward in how we deliver network, security, and operational services across diverse Cisco (and in the future, non-Cisco) fabric types. This architectural paradigm shift is underpinned by three core technological innovations that, in concert, forge a unified, scalable, and secure multi-fabric environment. These pillars are fundamental components designed to enable interoperability and consistent policy enforcement.

The Cisco ACI Border Gateway: Opening Up ACI

Historically, Cisco ACI, while revolutionary in its policy-driven automation, has operated as a relatively "closed" ecosystem. Its native distributed solutions, ACI Multi-Pod and NDO Multi-Site (orchestrated by Nexus Dashboard Orchestrator), provided robust and scalable connectivity and policy extension, but exclusively within the ACI domain. This limitation often required complex, unscalable hand-offs to other non-ACI VXLAN domains, which reduced ACI's agility in mixed environments.

The Genesis of the ACI Border Gateway

The ACI Border Gateway is based on the standard VXLAN EVPN Multi-Site architecture. This architecture uses a 'Border Gateway' to provide a clean interface for inter-fabric connectivity and flood control. By adding this standards-based gateway to ACI, we enable ACI to integrate into a broader multi-fabric ecosystem. This allows ACI to function as part of a larger, open, and unified data center network.

Key Innovations of the ACI BGW

The Cisco ACI Border Gateway (BGW), introduced in ACI software release 6.1(1), fundamentally transforms ACI's interoperability. It is a new, specialized device role for Nexus 9000 switches within an ACI fabric that

functions as a fully standards-compliant VXLAN EVPN gateway, implementing the IETF EVPN Multi-Site draft (draft-sharma-bess-multi-site-evpn-05). This capability effectively provides the following benefits:

- Standardized L2/L3 Extension and Interoperability: The ACI BGW speaks standard MP-BGP EVPN, allowing it to peer natively with NX-OS BGWs and, in the future, other technologies implementing BGWs or dedicated border functions. This eliminates the need for complex, proprietary integrations or traditional L2/L3 hand-offs, providing a truly open and scalable mechanism for extending VRFs and bridge domains.
- Anycast BGW Model for Resiliency and Scale: The ACI BGW leverages an anycast IP address model. At least two VIPs are deployed: all BGWs within a given ACI pod share the same external Virtual IP (VIP), which serves as the next-hop for most incoming traffic from remote fabrics. All BGWs within the same fabric share a common internal Virtual IP which serves as the next-hop for most outgoing traffic to remote sites. This design ensures seamless failover and native load balancing.
- Intelligent Tunnel Stitching: The BGW acts as a sophisticated tunnel termination and re-origination point. It intelligently stitches the ACI-internal iVXLAN tunnels to standard external VXLAN tunnels. This process involves decapsulating incoming packets, performing necessary lookups and translations, and then re-encapsulating them in the appropriate format for the destination fabric. This hardware-accelerated stitching ensures line-rate performance without introducing latency.
- Advanced Namespace Normalization: A critical capability of the ACI BGW is its ability to perform namespace normalization. This means it can translate ACI-specific VNIDs and pcTags to the corresponding values used in the external VXLAN EVPN domain, and vice-versa. This is particularly vital in brownfield scenarios where segment numbering or policy tags are not symmetrically defined across different fabrics. This dynamic translation ensures seamless connectivity without requiring a re-numbering of existing networks.
- Hierarchical Overlay for Scalability: By terminating tunnels at the BGW, Nexus One creates a hierarchical overlay network. ACI leaf nodes only need to establish tunnels to their local BGWs (or proxy TEPs on the local spines), not to every VTEP in every remote fabric. This dramatically reduces the number of tunnels and routing entries required on individual leaf nodes, significantly improving scalability and containing failure domains.
- Crucial Flood Control: The BGW serves as a vital control point for BUM (Broadcast, Unknown Unicast, Multicast) traffic. It can inspect incoming BUM frames, apply intelligent policies (e.g., rate limiting, filtering), and then forward them to remote sites only if permitted. This capability is paramount in preventing broadcast storms from cascading across the entire interconnected network, thereby enhancing overall network stability and resiliency. The ACI BGW storm control feature was released with ACI 6.2(2)F.

NX-OS VXLAN EVPN Security Groups (Group Policy Option): Policy for All

While ACI is known for its micro-segmentation capabilities, standard VXLAN EVPN fabrics were limited to legacy IP ACLs. This created a policy gap when unifying different fabrics under one security model.

The Evolution of Security in NX-OS

The journey of security in NX-OS has progressed from basic ACLs applied at different levels (VLANs, SVIs, L3 ports) to more advanced features like enhanced policy-based redirection (ePBR). However, these mechanisms remained tightly coupled to network constructs. The need for a more agile, identity-based security model, like ACI's ESG, became evident as applications became more distributed and dynamic. The

introduction into NX-OS of VXLAN EVPN Security Groups marks a pivotal moment in this evolution, bringing a robust, flexible, and scalable macro- and micro-segmentation capability to standard fabrics built on top of NX-OS.

Key Innovations of Security Groups (GPO)

NX-OS VXLAN EVPN Security Group feature, based on the IETF Group Policy Option (GPO) draft, fundamentally elevates the security posture of NX-OS fabrics. This innovation establishes a similar policy language that can be consistently interpreted, understood, and enforced across both ACI and NX-OS domains, thereby enabling a unified security framework for Nexus One.

With this feature security is no longer intrinsically tied to IP subnets, VLANs, or physical interfaces. Instead, endpoints are classified into logical Security Groups (SGs). This abstraction allows for policies to be defined based on business intent (e.g., "Web Tier," "Database Tier," "PCI Workloads") rather than network addresses, providing unprecedented flexibility and agility. Classification into SGs is based on flexible criteria, including IP addresses, VLAN IDs, ingress ports, MAC addresses or VM attributes with the help of automation.
Security rules, or contracts (referred to as SGACLs - Security Group Access Control Lists - in NX-OS), are defined between these SGs. These contracts explicitly specify the applied actions for traffic flowing between a source SG and a destination SG. Actions include permit (allow traffic), deny (block traffic), and redirect (steer traffic to a service device). This enables a true zero-trust security model, where only explicitly allowed communication can occur.
GPO standard integrates both the control plane and data plane to enable highly efficient and optimized policy enforcement.

- Data Plane Integration: Each VXLAN packet can carry a 16-bit Security Group Tag (SGT) within its VXLAN GPO header. This SGT uniquely identifies the security group associated with the source of the inner, real packet. Additionally, a "policy applied" bit ("A bit") can be set in the header, signaling to downstream VTEPs that the policy has already been enforced at the ingress point, thus preventing redundant processing.
- Control Plane Integration: The SGT information is not just carried in the data plane; it is first included as a BGP EVPN extended community within endpoint reachability advertisements (Type-2 for MAC/IP, Type-5 for prefixes). This crucial innovation allows every VTEP (VXLAN Tunnel Endpoint) in the fabric to learn the SGT of any destination endpoint via the control plane. This knowledge enables optimal ingress policy enforcement, as the first-hop VTEP has all the necessary information (source SGT, destination SGT, and applicable contract) to make a policy decision at the earliest possible point.
- Flexible and Powerful Service Chaining: GPO provides native and highly flexible support for Enhanced Policy-Based Redirection (ePBR). Traffic can be easily steered to firewalls, load balancers, or other inspection/security services-based on SG-to-SG contracts. This eliminates the need for complex, IP-based PBR configurations and enables dynamic service insertion, even across fabrics part of the same campus or in geographically dispersed data centers.
- Scalability and Resilience: By decoupling security from network topology and leveraging the EVPN control plane, GPO offers a highly scalable solution. Policies are enforced consistently regardless of endpoint mobility or network changes, and the distributed nature of enforcement enhances overall resilience.

The Unified Nexus Dashboard: A Single Point of Control

Managing multiple fabrics with different tools, APIs, and models is inefficient and unsustainable. The Unified Nexus Dashboard centralizes the provisioning, monitoring, and automation of all data center fabrics. It currently supports the Nexus One architecture for ACI and NX-OS VXLAN EVPN, with future support planned for Hyperfabric and Campus.

The Need for Unified Management

Managing diverse network domains independently creates an operational burden. It leads to:

- Operational Silos: Separate teams, tools, and processes for each fabric, hindering collaboration and end-to-end visibility.
- Increased TCO: Higher training costs, more licenses for disparate management platforms, and greater manual effort.
- Slower Time to Deploy: Complex, multi-step provisioning processes delay application deployment and service activation.
- Higher Risk of Error: Manual translation of policies and configurations across different platforms is inherently error-prone.

The Unified Nexus Dashboard directly addresses these pain points by consolidating management functions into a cohesive, centralized platform.

Key Innovations of the Unified Nexus Dashboard

The Unified Nexus Dashboard is more than just a management console; it is a sophisticated orchestration and automation platform that will soon provide a consistent operational experience across the entire Nexus One domain.

The Nexus Dashboard abstracts away the underlying complexities of individual fabric types. Operators will define logical constructs, such as VRFs, Networks, Security Groups, Contracts, and related objects, only once, at a high level of intent. Nexus Dashboard will then intelligently translate and deploy this intent into the specific configurations required for each underlying fabric (ACI or NX-OS), ensuring consistency without requiring deep, fabric-specific knowledge from the operator.
Unified Health Monitoring: A single, consolidated view of the operational data, health status, faults, KPIs, and events across all interconnected sites and fabrics.
End-to-End Visibility: The ability to trace a traffic flow from an endpoint in an ACI fabric, through the BGWs, to an endpoint in an NX-OS fabric, and visualize the forwarding decisions applied at each hop.
Correlated Troubleshooting: Advanced analytics and correlation engines to quickly identify the root cause of issues that span multiple fabrics, reducing mean time to innocence (MTTI) and mean time to resolution (MTTR).
A single, consistent REST API is exposed by the Nexus Dashboard for all provisioning and operational tasks. This better enables Infrastructure-as-Code (IaC) adoption for the multi-fabric data center, allowing seamless integration with external automation tools, DevOps pipelines, cloud orchestration platforms, and security tools (e.g., SIEM/SOAR).

Note: Support for unified ACI and NX-OS fabric configurations is available starting with the Nexus Dashboard 4.2 release.

Nexus One Innovations: The Architectural Pillars Deep Dive

This section provides a detailed technical examination of the core components of the Nexus One architecture, focusing on the ACI Border Gateway, NX-OS Security Groups, and the role of the Nexus Dashboard.

The Cisco ACI Border Gateway (BGW)

The VXLAN EVPN Border Gateway is the center of the Nexus One forwarding architecture, serving as the bridge between the ACI fabric's proprietary control and data planes (COOP and iVXLAN) and the standards-based world of VXLAN EVPN.

Role, Placement, and Hardware Requirements

BGWs must be deployed as dedicated leaf nodes connected to the ACI spines. As of ACI release 6.1, the BGW function cannot be co-located on a spine device. At the time of writing of this document, BGWs must be dedicated to the inter-site connectivity only; they cannot be used for end-host access nor for external connectivity via tenant L3Out.

Note: Future releases will allow the co-existence of tenant L3Out in the BGW leaves.
For redundancy, a minimum of two BGWs should be deployed per site (or per pod in a Multi-Pod fabric).

Note: In an ACI Multi-Pod deployment, BGWs must be deployed in each pod that requires connectivity to the external VXLAN EVPN domain.

Platform

Cisco Nexus 9300-FX, -FX3, -GX, -GX2, and -HX series switches can operate in the BGW role, if they are equipped with a minimum of 32 GB of RAM. FX2 devices cannot function as BGWs because they do not meet this memory requirement.
ACI Software: All nodes in the ACI fabric (APICs, spines, and leaves) must be running ACI software release 6.1(1) or later. For full functionality, including asymmetric namespace normalization, release 6.1(5)M or later is recommended.

Licensing

The ACI Border Gateway, like any other device in the ACI fabric must be licensed with the Advantage tier.

Control Plane and Data Plane Overview

Control Plane Architecture: Overlay and Underlay

The Cisco ACI BGW is essential for integrating ACI and VXLAN EVPN domains, operating through a multi-protocol control plane. To facilitate this connection, the APIC configures each BGW with a combination of unique and shared loopback IP addresses. These addresses include:

- Two unique loopbacks per BGW: **ptep** and **vxlan-evpn-rtep.** They are respectively used to establish intra-site and inter-site control plane overlay adjacencies; they might also be used as source or destination for the VXLAN outer headers depending on the traffic type and direction.
- Two shared anycast loopbacks: the external VIP (**vxlan-evpn-anycast-v4-external-tep**) and the internal VIP (**vxlan-evpn-anycast-v4-internal-tep**). They are exclusively used as source or destination for the VXLAN outer headers.

The specific roles of each of these loopback addresses in the underlay and overlay will be explained in the following sections.

```
aci-bgw-3102# show ip interface
lo0, Interface status: protocol-up/link-up/admin-up, iod: 3, mode: ptep
  IP address: 172.21.10.133, IP subnet: 172.21.10.133/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 0, tag: 0
lo1, Interface status: protocol-up/link-up/admin-up, iod: 73, mode: vxlan-evpn-anycast-v4-external-tep
  IP address: 172.20.0.3, IP subnet: 172.20.0.3/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 0, tag: 0
lo2, Interface status: protocol-up/link-up/admin-up, iod: 74, mode: vxlan-evpn-anycast-v4-internal-tep
  IP address: 172.21.11.164, IP subnet: 172.21.11.164/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 0, tag: 0
lo3, Interface status: protocol-up/link-up/admin-up, iod: 75, mode: vxlan-evpn-rtep
  IP address: 172.20.0.2, IP subnet: 172.20.0.2/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 0, tag: 0
```

Note: The loopback IDs are randomly allocated and configured by the APIC

Note: Readers familiar with NX-OS BGWs may notice differences between what is described above and their existing knowledge of standard NX-OS deployments. Specifically, NX-OS BGWs utilize three distinct loopbacks: the RID (Router ID) for establishing internal and external control-plane adjacencies, the PIP (Primary IP) as the source and sometimes destination for VXLAN packets, and the Multi-Site VIP, an anycast IP shared by all BGWs in the same site that serves as the next-hop for re-originated prefixes and the destination for VXLAN traffic.

Underlay Control Plane

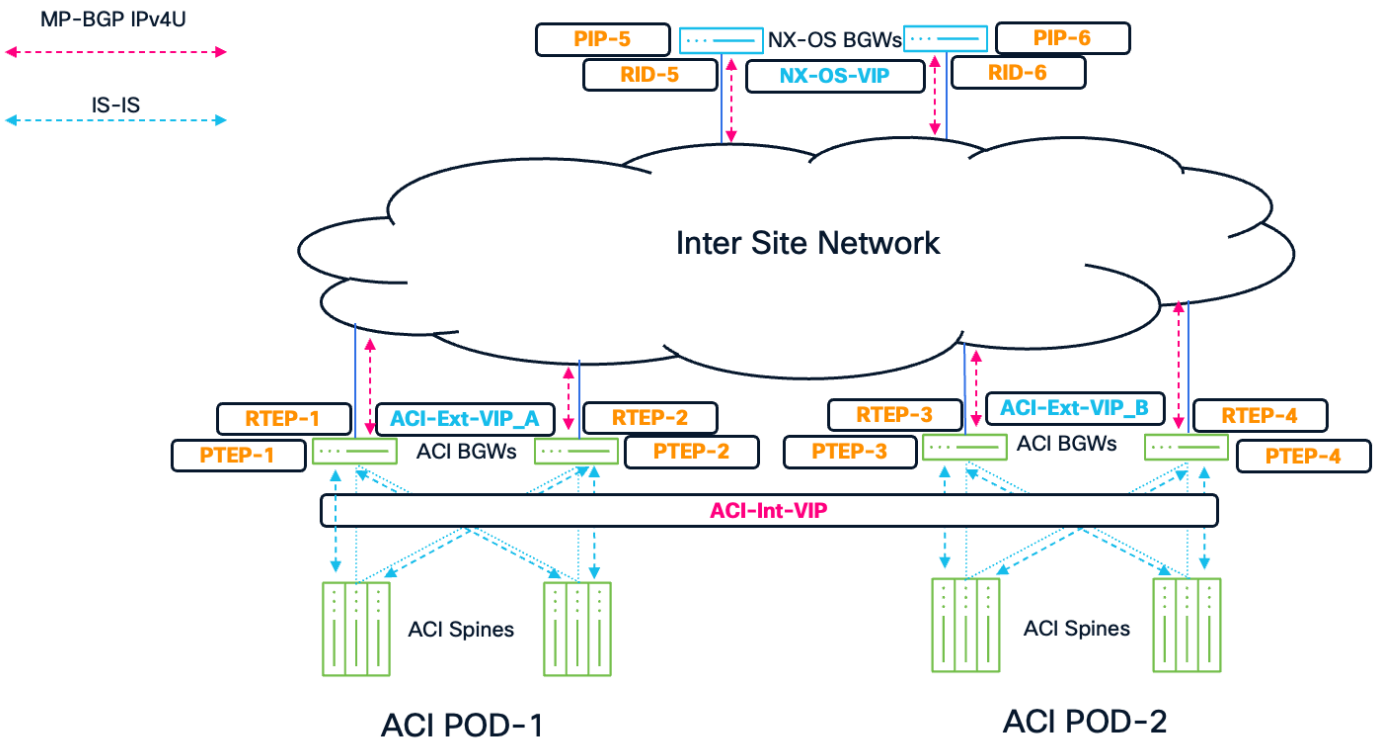

Figure 1 Underlay Control Plane
PTEP = ptep
RTEP = vxlan-evpn-rtep
ACI-Int-VIP = vxlan-evpn-anycast-v4-internal-tep
ACI-Ext-VIP = vxlan-evpn-anycast-v4-external-tep

To ensure end-to-end connectivity, a BGW must advertise its various loopback addresses across different network domains.

Within the local ACI fabric, BGWs adhere to the standard underlay model by forming IS-IS adjacencies with spine switches over IP unnumbered point-to-point links. This provides the internal reachability needed for the BGW to advertise its locally significant loopbacks, **ptep** and **vxlan-evpn-anycast-v4-internal-tep**, throughout the fabric.

For connectivity between sites, BGWs establish IPv4 unicast eBGP sessions with the first hop routers in the Inter-Site Network (ISN) to exchange the **vxlan-evpn-rtep** and the **vxlan-evpn-anycast-v4-external-tep** loopbacks.

This comprehensive advertisement strategy ensures that the correct MP-BGP sessions can be established and that overlay iVXLAN/VXLAN packets can be correctly routed in tunnels, both within the fabric and across the ISN.

Overlay Control Plane

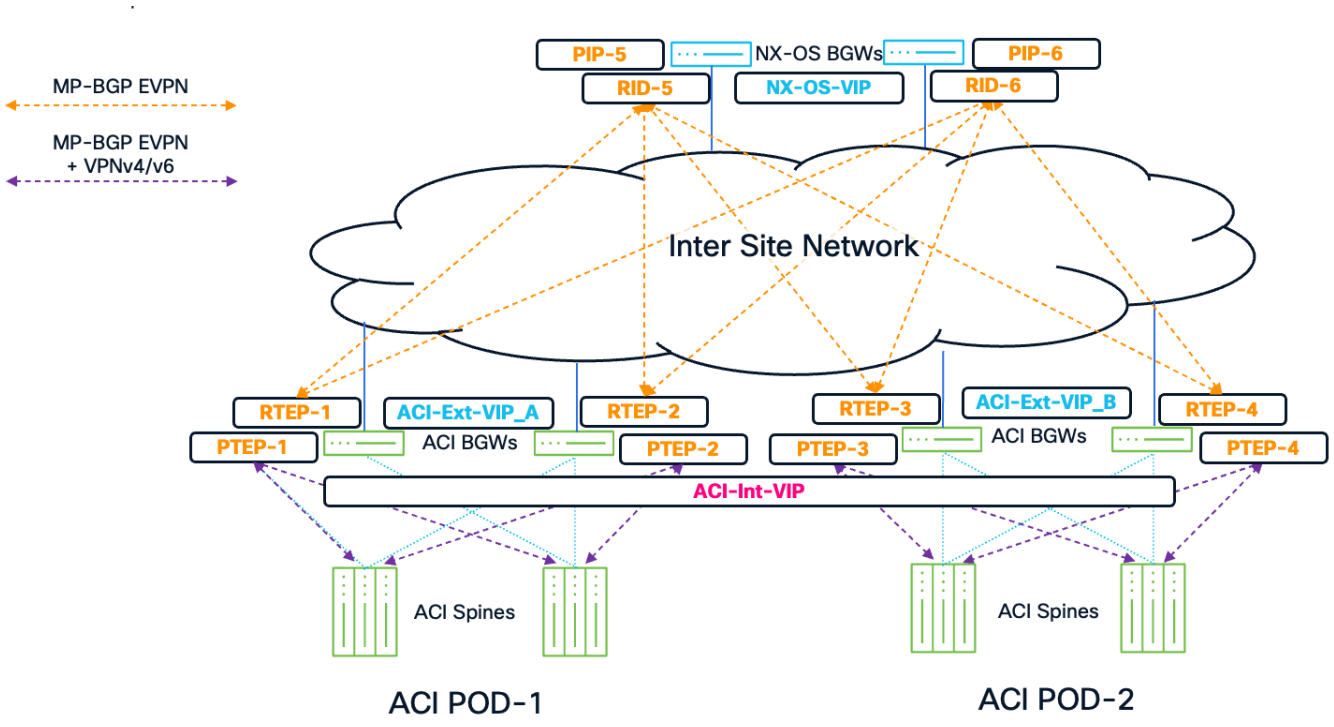

Figure 2 Overlay Control Plane

PTEP = ptep
RTEP = vxlan-evpn-rtep
ACI-Int-VIP = vxlan-evpn-anycast-v4-internal-tep
ACI-Ext-VIP = vxlan-evpn-anycast-v4-external-tep

ACI BGW provides network administrators with granular control to selectively advertise overlay reachability information, including endpoints and prefixes, to remote sites. This is managed by configuring which L2VNIs (Bridge Domains) or L3VNIs (VRFs) need to be stretched across the sites.

To facilitate this route exchange, a significant enhancement is made to the ACI control plane by establishing new MP-BGP EVPN sessions between each BGW and the spines within the same pod. In parallel, for inter-site communication, each ACI BGW establishes EVPN sessions with every remote NX-OS BGW to exchange Type-2, Type-3, and Type-5 routes. Route-Server support, to avoid a full mesh of eBGP sessions in the Inter Site Network was introduced in ACI 6.2(2)F. This document will be updated accordingly when ACI train 6.1 will also be validated for the Route-Server use case.

The flow for an endpoint advertisement originating from ACI begins when a leaf learns about its existence and publishes it to the spines via the Council of Oracle Protocol (COOP). The spines store this in the COOP database, install a hardware forwarding entry, and then, if the ACI Bridge Domain is configured to be stretched, redistribute the endpoint as an EVPN Type-2 route to the BGWs over the new internal EVPN session. The next-hop for these routes is the ACI spine-proxy TEP, an anycast address shared by all spines in the pod. This design optimizes endpoint mobility, as endpoints that move between leaves within the fabric do not require updates to the BGWs. Finally, the BGW re-originates the route toward the remote NX-OS BGWs, normalizing the VNI and SGT values in the process. When re-originating the advertisements, the BGWs set the EVPN next-hop as their **vxlan-evpn-anycast-v4-external-tep.**

When ACI BGWs receive remote Type-2 routes from NX-OS fabrics, they re-originate, normalize, and send them to the local spines, which then redistribute these into the COOP database. This allows the implementation within the ACI fabric to remain consistent for all types of endpoints (local and remote). It achieves a highly scalable fabric because leaf nodes learn the endpoint information on other nodes only when needed, allowing the optimized use of resources on each node.

Note: The section “End-to-End Connectivity and Policy Enforcement” contains a detailed explanation on how endpoints and external networks are exchanged across the devices in the different fabrics

A similar process occurs for L3 prefixes (EVPN Type-5 routes): external prefixes learned on ACI border leaves via L3Outs are advertised to spines using MP-BGP VPNv4/v6, reflected to BGWs, and then redistributed as EVPN Type-5 routes to remote sites. Conversely, incoming Type-5 routes from remote sites, whether representing external networks learned via L3Outs in those fabrics or the locally configured L2VNI subnets, are redistributed by the BGW into the fabric's VPNv4/v6 sessions with the spines which will finally reflect to all ACI leaves.

The next-hop selection is critical for forwarding traffic correctly. Depending on the type of EVPN route advertised, and the direction, different next-hops might be used:

- When ACI BGWs advertise Type-2 routes internally toward the spines, the next-hop is set to the BGW's internal shared anycast VTEP loopback (**vxlan-evpn-anycast-v4-internal-tep**). When these Type-2 and Type-5 routes are advertised externally toward the BGW nodes in remote sites, the next-hop is the external shared anycast VTEP loopback (**vxlan-evpn-anycast-v4-external-tep**).
- With the advertisement of EVPN Type-3 routes, which are used to signal a BGW's willingness to receive BUM traffic for a stretched bridge domain the BGW always originates them from the **vxlan-evpn-rtep** as the next-hop. Remote BGWs collect these next-hops from the received Type-3 updates to build their list of replication peers, ensuring that a copy of the BUM traffic is sent to each remote BGW. A Designated Forwarder is elected per bridge domain to prevent loops.
- When Type-5 advertisements are received from NX-OS fabrics, each ACI BGW will redistribute them in VPNv4/VPNv6 and use its own **ptep** as the next-hop.

Finally, a crucial operational consideration concerning the control plane, but also data plane, is the MTU. Control-plane sessions are highly sensitive to MTU settings, so it is vital to ensure a proper and consistent MTU configuration on all ISN links and devices.

While a BGP session may initially establish (as Open/Keepalive messages are small), problems can arise later when a BGP update larger than the path MTU needs to be transmitted, causing intermittent failures or complete downtime. Therefore, the APIC setting that controls control plane’s default MTU value may need to be adjusted based on the maximum MTU supported across the entire Inter-Site Network (ISN) path to prevent such issues.

By default, ACI nodes generate control plane updates with an MTU of 9000 bytes. NX-OS devices, instead, by default use IP MTU of the physical interface used to reach the BGP neighbor.

This default behavior can be globally changed on ACI by setting a lower value in the Control Plane MTU Policy. Although the change will affect both intra-site and inter-site adjacencies no performance degradation will be caused by this.

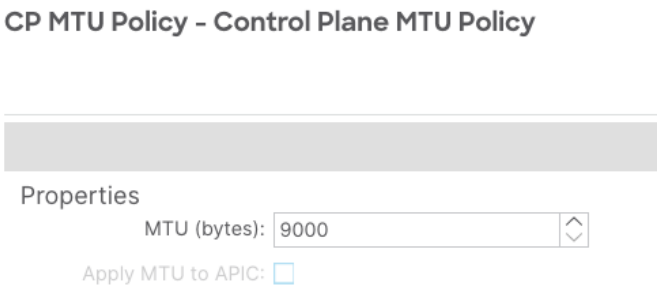

Figure 3 APIC Control Plane MTU Policy

Note: A similar behavior can be achieved on the NX-OS devices by using the “ip tcp mss”. Reducing the MSS will consequently reduce the packet maximum size.

Note: The TCP MSS will be negotiated by the two conversation endpoints. The lower value will always be honored and that will dictate the overall size of the IP packet.

Data Plane Mechanics: Tunnel Stitching and Forwarding

In the data plane, the BGW functions as a high-performance forwarding and translation engine. Its primary role is to "stitch" together two different VXLAN tunnels: the internal fabric's iVXLAN and the standard VXLAN used for inter-site traffic.

The core data plane mechanism it uses is tunnel stitching, a process of decapsulation and re-encapsulation.

To illustrate this process at a high level, let’s pick a real scenario, consider a packet being **routed** from an endpoint in the local ACI fabric to an endpoint in a remote NX-OS fabric:

- An ACI leaf encapsulates the original packet in an iVXLAN packet and forwards either:
  - Toward the ACI BGW`s internal VIP. This can happen when the subnet is stretched and the remote NX-OS endpoint has been learned by the ACI leaf via an ARP packet.
  - Toward the spine-proxy TEP, if the destination subnet is stretched and the remote endpoint has not been locally learned. In that case the spine performs a lookup on the inner destination address and based on its COOP database entry, forwards the iVXLAN packet to the ACI BGW's internal VIP.
  - Or directly to the ACI BGW's internal ptep, if the destination subnet is not stretched.
- The BGW receives and decapsulates the iVXLAN packet, exposing the original inner frame.
- It then performs a lookup on the inner frame's destination and re-encapsulates it into a new, standard VXLAN packet.
- During re-encapsulation, the BGW applies namespace normalization, translating the VNI to the globally significant value required for the remote site, as configured by the administrator.
- Finally, the BGW forwards the new VXLAN packet across the ISN to the remote fabric's BGWs.

This entire tunnel stitching process, including decapsulation, lookup, normalization, policy enforcement and re-encapsulation—is performed in hardware at line rate. This ensures there is no performance penalty and

makes it the key mechanism enabling both inter-site scalability and VNI translation between asymmetric network domains.

Namespace Normalization: Symmetric and Asymmetric

A critical function of the BGW is to handle differences in how network and security segments are numbered across fabrics.

Symmetric Namespace: In a greenfield deployment, it is possible to configure the same VNID and SGT values for a given segment in both the ACI and NX-OS fabrics thanks to the flexibility that NX-OS gives to the administrators when defining L2VNIs and L3VNIs and Security Group Tags. In this case, no translation is needed. However, this is often impractical, especially in brownfield scenarios where an existing NX-OS VXLAN EVPN VNI needs to be extended into ACI, or when an existing ACI VNI needs to be extended to a different ACI fabric via the BGW and not via the classic ACI Multi-Site architecture.

For additional information on ACI Multi-Site forwarding via the BGW consult the ACI Multi-Site With BGWs chapter.

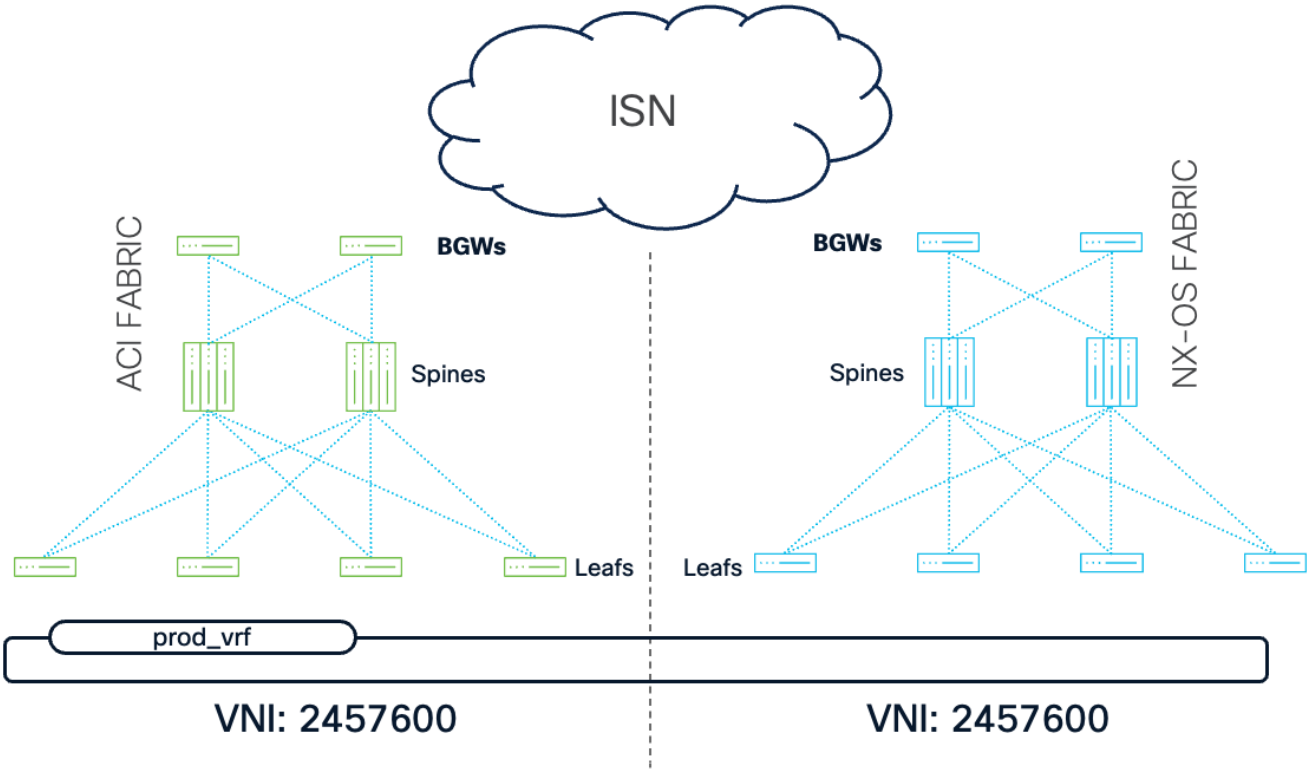

Figure 4 Symmetric VNI Namespaces

Asymmetric Namespace: This is the more common scenario where the ACI BGW will play a pivotal role. For the same extended resource, like a VRF, a BD or an ESG, the ACI fabric may use any VNID while the NX-OS fabric uses a different one.

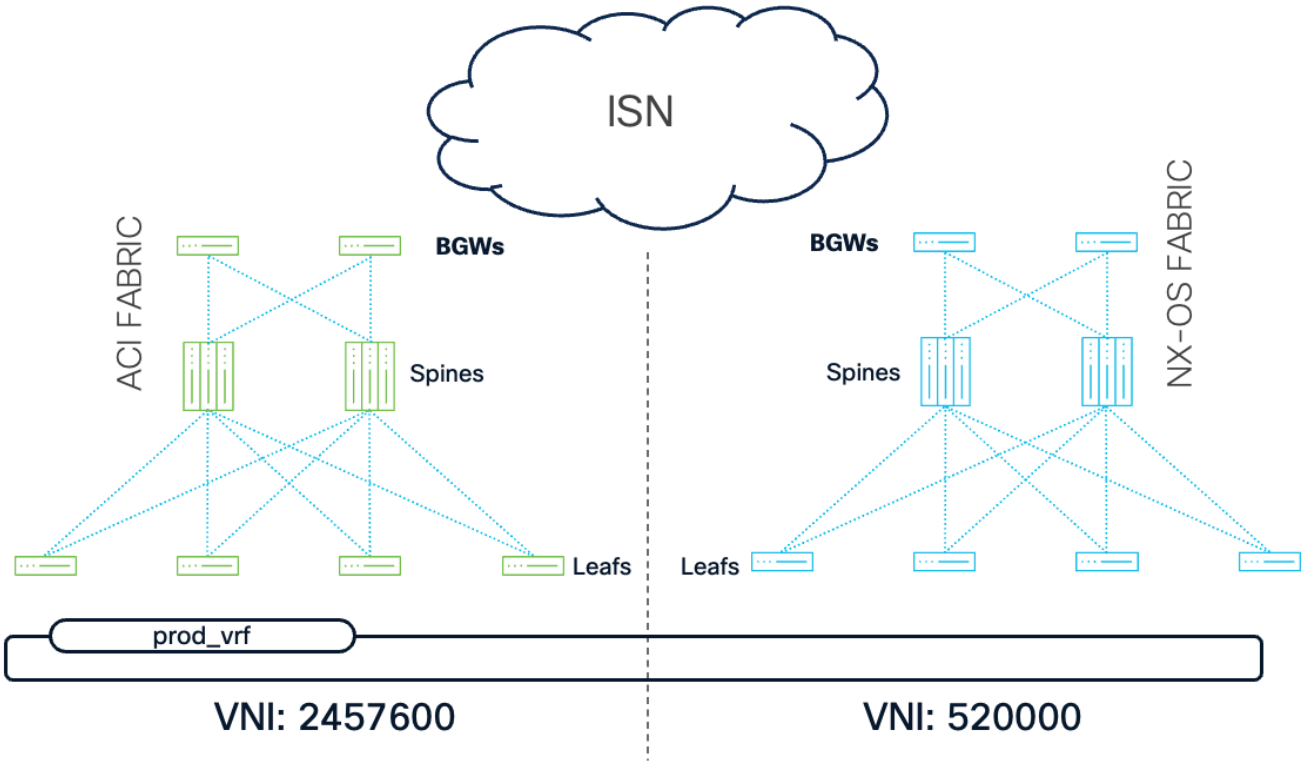

Figure 5 Asymmetric VNI Namespaces

Administrators can then configure via the APIC the proper translation entries that will preserve the end-to-end multi-tenancy. This two-way translation is performed on the ACI BGW and is essential for integrating existing fabrics. ACI BGWs support it starting with 6.1(4).

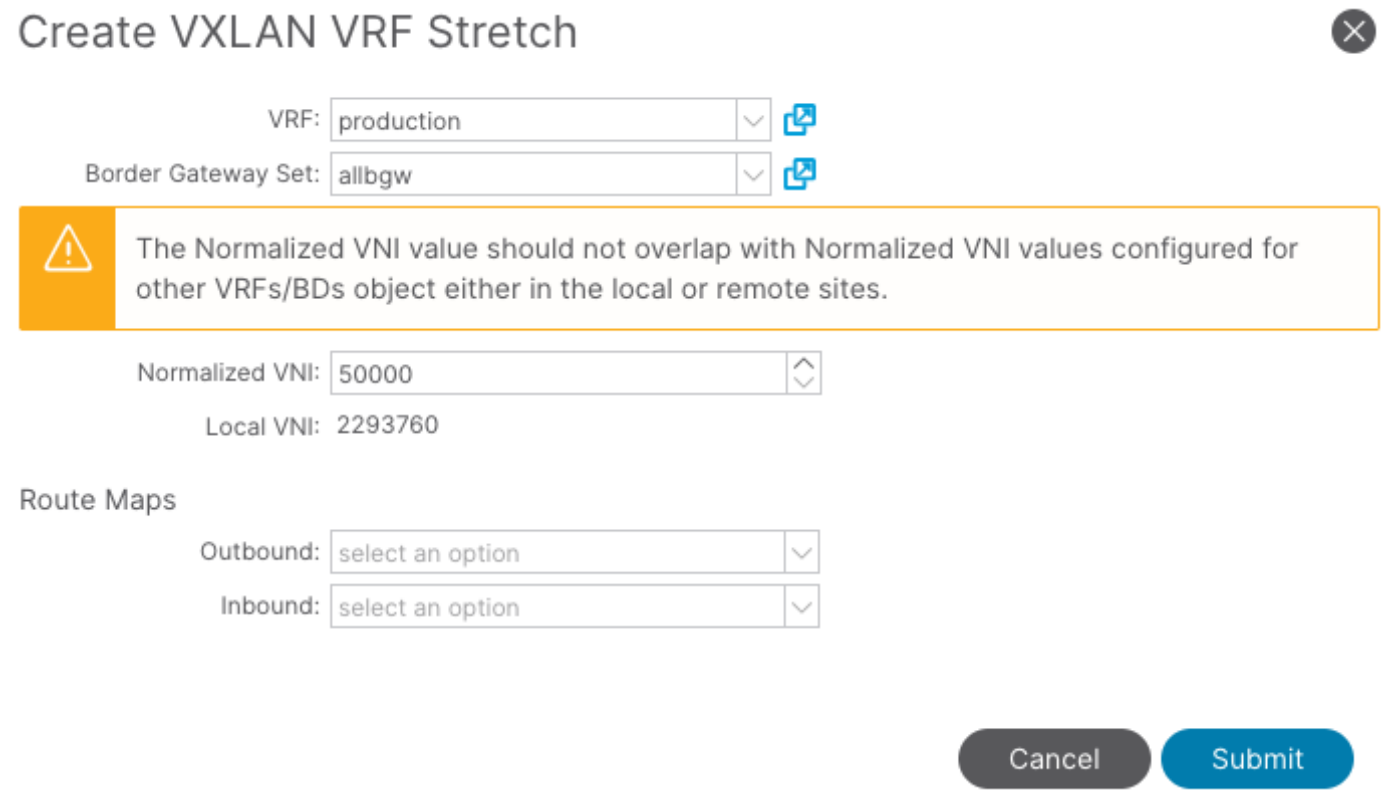

Figure 6 VRF Stretch and Normalization

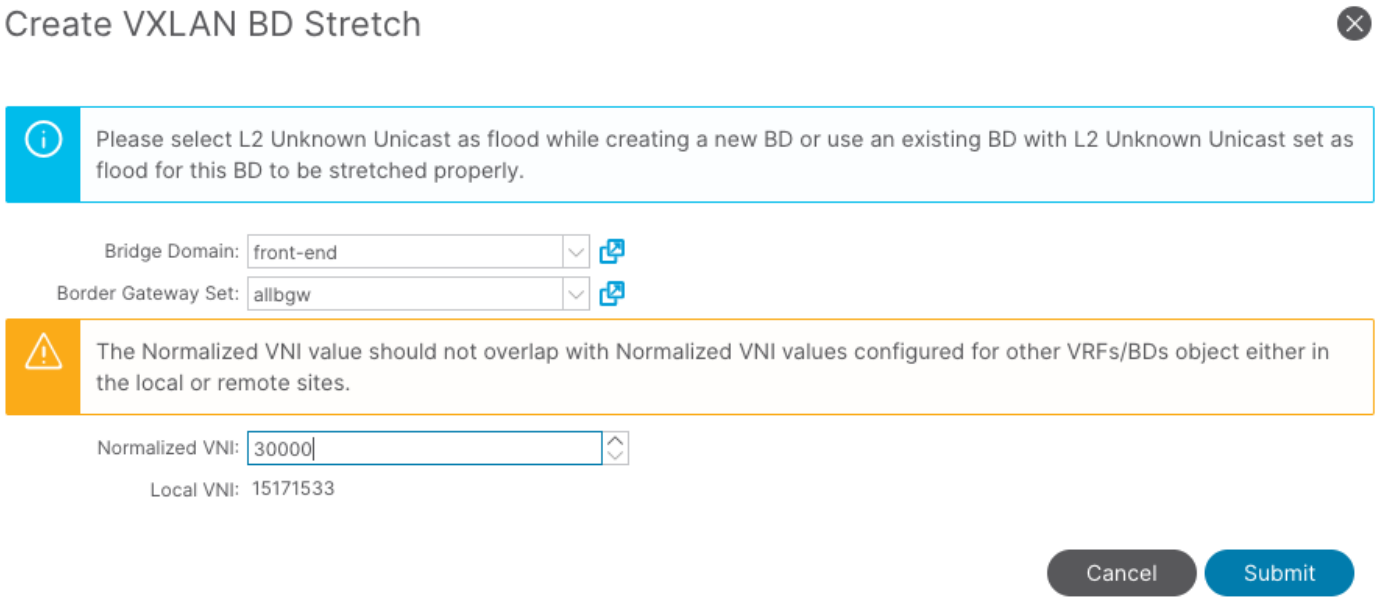

Figure 7 BD Stretch and Normalization

Note: To advertise the ACI Bridge Domain subnets to remote NX-OS BGWs, they must also be configured with the 'Advertised Externally' flag.

Note: The ACI BDs must be configured to flood all unknown unicast bridged traffic.

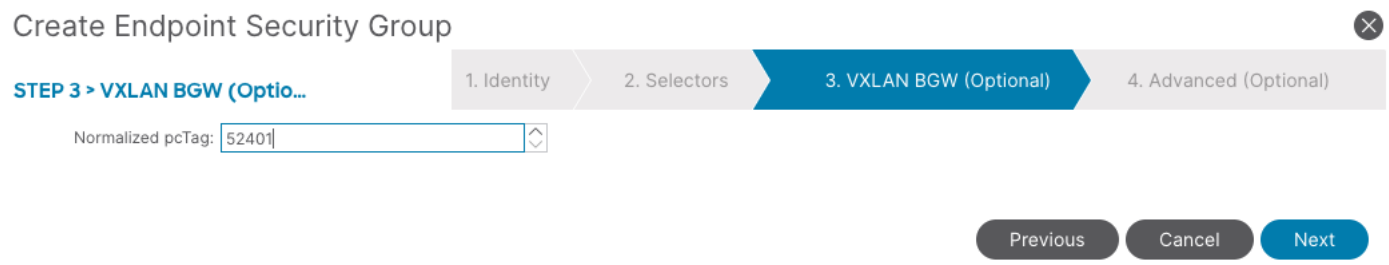

Figure 8 ESG Normalization

This mechanism ensures seamless communication and consistent policy enforcement across heterogeneous ACI and NX-OS environments, facilitating workload mobility and unified multi-tenant operations.

Policy Enforcement

This section focuses on the policy enforcement related to ACI fabrics only, assuming that remote fabrics are policy-unaware. A broader review that takes in consideration policy enforcement across ACI and NX-OS VXLAN GPO-aware fabrics is discussed into the chapter End-to-End Connectivity and Policy Enforcement

.

Policy enforcement within the Nexus One architecture is a critical aspect of securing communications between ACI fabrics and remote VXLAN EVPN domains.

This section details how traffic policies are applied, highlighting the mechanisms available prior to ACI 6.1(4) and the enhanced capabilities introduced with ACI 6.1(4).

Policy enforcement was not supported in ACI release 6.1(1), which was the first to introduce the BGW functionality. To allow the traffic to flow, the ACI VRF had to be set as Unenforced. Subsequent releases incrementally added enforcement capabilities.

In ACI 6.1(2), security enforcement could be applied on the ACI fabric after classifying remote endpoints and subnets through configuration-based methods.

ACI 6.1(4) enhanced interoperability with NX-OS policy-aware fabrics by enabling the BGW to learn, via EVPN extended communities, the Security Group Tags assigned by NX-OS fabrics during their classification processes for a more scalable way to map the remote SGTs in ACI ESGs.

Note: The security group identifier, or tag, may be referred to by different names depending on the context. In VXLAN EVPN fabrics, the identifier is also known as the SGT (Security Group Tag) or GPO id. In ACI literature, it is often referred to as *pcTag* or *sclass*. This document uses these terms according to their respective contexts.

Introduction to Policy Enforcement

The ACI Border Gateway (BGW) acts as a bridge, extending ACI's policy-driven networking to remote VXLAN EVPN fabrics. Policy enforcement ensures that only authorized communication flows between these disparate environments are allowed, adhering to the security policies defined within ACI. This involves classifying traffic based on various criteria and applying predefined actions (e.g., permit, deny, redirect) at specific points in the network.

Endpoint Classification and Policy Enforcement Evolution

- From ACI 6.1(2) to ACI 6.1(3): Users classify remote endpoints and networks by configuring traditional Endpoint Security Group (ESG) selectors and new ones introduced in ACI 6.1(2). This classification method only applies when the remote NX-OS fabric is policy-unaware.
- Starting from ACI 6.1(4): The ACI Border Gateway (BGW) will learn and reuse the SGT tags advertised via EVPN from remote VXLAN EVPN fabrics when the remote fabric is policy-aware (NX-OS 10.6(1)F or newer is required); when this happens, the BGW translates the remote SGT tag into a local ESG pcTag, enabling seamless and consistent classification and policy enforcement.
  For policy-unaware remote VXLAN EVPN fabrics, the BGW cannot rely on EVPN SGT tags because they are not advertised. In this case, ACI falls back to the local classification methods available prior to 6.1(4), using ESG selectors configured by administrators to classify remote endpoints and routes.

Note: When ACI receives remote endpoints and subnets that include an SGT extended community, this externally advertised SGT takes precedence over any SGT derived from ACI's local classification configuration for those same entities.

Note: ESGs are always mandatory on the ACI fabric, and customers who are still using EPGs as security classifiers are required to migrate to ESGs. A script to help customers transition from EPG to ESG is available here.

Policy Enforcement Points

Policy enforcement within the ACI fabric is applied at specific points based on traffic type and direction.

For bridged traffic (Layer 2) enforcement primarily occurs at the ACI leaf for bidirectional East-West flows. The ACI leaf maintains MAC classification for both locally connected and remote endpoints through data-plane learning. The Border Gateway (BGW) handles exceptions such as transient or unidirectional flows; an example is the case when the remote fabric MAC addresses are not in the ACI leaf forwarding table but only exists in the BGW forwarding and EVPN tables.

For routed traffic (Layer 3) enforcement is split between the ACI leaf and BGW depending on flow direction and type.

For East-West routed traffic originating from remote NX-OS fabrics and targeting ACI-connected endpoints, enforcement occurs at the ACI leaf. This is because ACI BGWs do not install local endpoint host routes in their forwarding tables; instead, they forward traffic to the spine-proxy despite having learned these routes via the control plane. The iVXLAN packet carries the source ESG tag, injected by the ACI BGW after classification, which the egress leaf uses to apply policy enforcement.

For North-South routed traffic originating from remote NX-OS fabrics and targeting ACI external routes learned via L3Out, enforcement occurs at the ACI BGW, which locally installs all ACI learned external routes with their respective pcTag.

Traffic originating inside the ACI fabric and targeting remote fabric connected endpoints or external networks might be enforced at the local ACI leaf or at the ACI BGW:

- ACI leaves do not proactively learn remote endpoint host routes via the control plane, nor do they perform IP-based data plane learning for remote endpoints. This design choice prevents the installation of potentially thousands of /32 host routes into leaf hardware tables, ensuring the fabric remains within maximum scale limits. Instead, ACI leaves rely on ARP-based data plane learning, populating their endpoint tables only when they process ARP packets originated from remote

endpoints and destined for an endpoint locally attached to the ACI fabric. When the remote endpoint is known to the leaf, policy is enforced locally; otherwise, enforcement occurs at the ACI BGW, which maintains full awareness about the remote endpoints and classifies them based on the ESG selectors.

- ACI leaves always install remote fabric external routes with a pcTag equal to 1. If the route is classified in an ESG, only ACI BGWs will program the FIB with the correct value.

This architecture ensures efficient, scalable, and consistent policy enforcement across East-West and North-South traffic flows between ACI and remote VXLAN domains.

The next table and figures summarize the different policy enforcement points in an ACI fabric; once more the assumption is that the remote fabric is policy-unaware.

Flow Type	Direction	Enforcement Point	Explanation
Bridged Traffic: Remains inside the same stretched Bridge Domain	NX-OS -> ACI	ACI Egress Leaf	ACI BGW installs local ACI-learned endpoint MACs with pcTag =1. Traffic is encapsulated in iVXLAN; the BGW sets the pcTag assigned to the ESG that classifies the remote NX-OS attached endpoint. The leaf, knowing the destination pcTag as a connected EP, enforces the policy
	ACI -> NX-OS	ACI Ingress Leaf	Assumption: ACI leaf has learned the NX-OS endpoint pcTag via data plane and applies the policy directly
		ACI BGW	Assumption: ACI leaf has not learned the NX-OS endpoint via data plane; it floods the frame. BGW, knowing the remote EP via EVPN, and its pcTag via ESG selector rules, applies enforcement. Unknown unicast traffic destined to a remote EP is flooded in the NX-OS fabric
Routed Traffic: Routed by ACI pervasive gateway	NX-OS -> ACI (East-West)	ACI Egress Leaf	ACI connected endpoints are not in BGW forwarding tables. BGW uses a pervasive route to spine-proxy. Traffic encapsulated in iVXLAN carries pcTag based on BGW classification. Leaf enforces policy knowing destination pcTag as connected EP
	NX-OS -> ACI (North-South)	ACI BGW	ACI BGW learns ACI L3Out external subnets and their pcTag; enforcement is applied immediately on packet receipt
	ACI -> NX-OS (East-West)	ACI Leaf or ACI BGW	IP information for remote NX-OS endpoint might not be learned via data plane by the ACI Leaf nodes. If that's the case, traffic for these endpoints is forwarded to the BGW which then uses the destination pcTag learned from the local ESG selector to enforce traffic. If the ACI Leaf knows the remote NX-OS

Flow Type	Direction	Enforcement Point	Explanation
			attached endpoint then it can directly apply the policy.
	ACI -> NX-OS (North-South)	ACI BGW	Remote NX-OS external route pcTag is not installed in ACI Leaf forwarding tables; leaves forward traffic to BGW, which maps the destination pcTag based on the local ESG selector

Figure 9 illustrates bridged traffic being forwarded from an NX-OS-connected endpoint to an ACI-connected endpoint within the same L2VNI. In this scenario, the ACI egress leaf serves as the enforcement point. The BGW installs local ACI-learned endpoint MAC addresses with a pcTag of 1. During iVXLAN encapsulation, the BGW injects the pcTag derived from the ESG selector. The egress leaf then enforces the policy, as it looks up the source pcTag from the iVXLAN header and knows the destination pcTag of the locally connected endpoint.

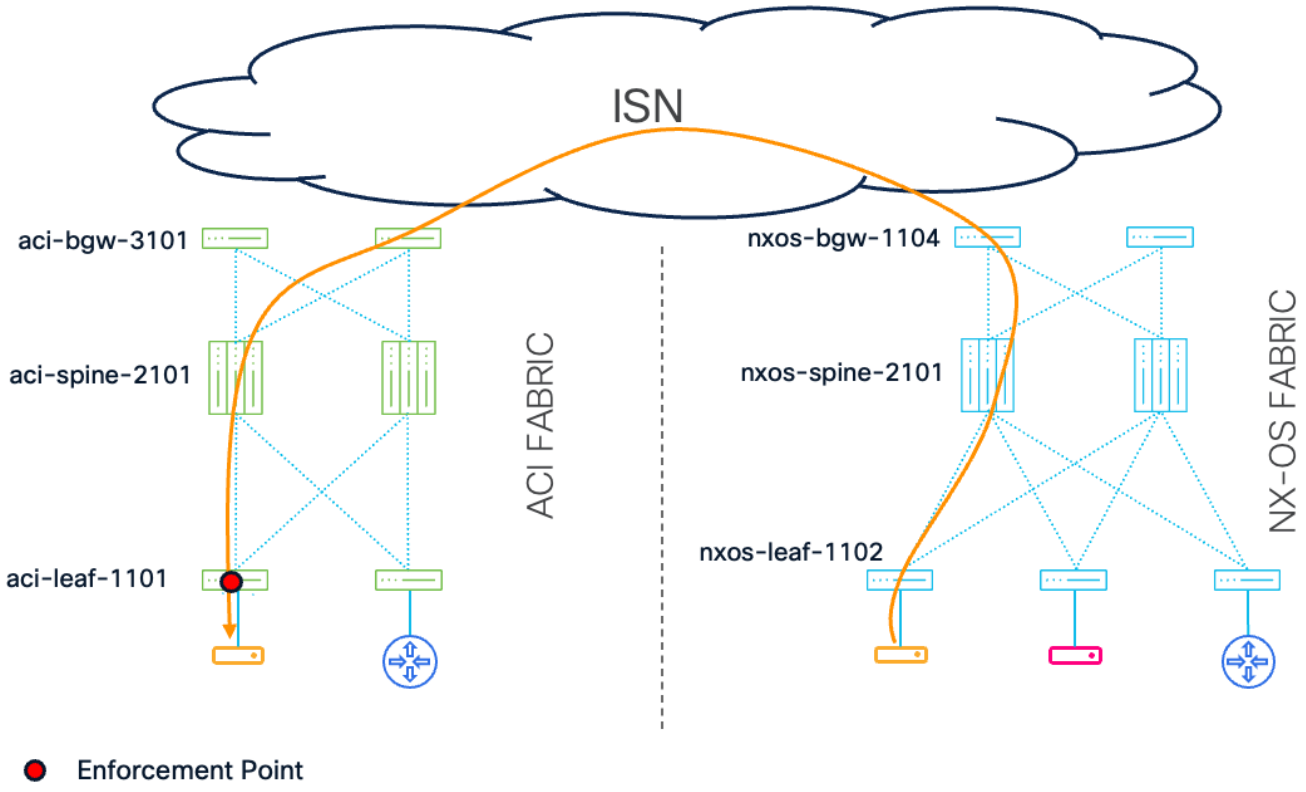

Figure 9 Bridged Traffic From NX-OS to ACI

The following figure illustrates the same bridged traffic flow in the opposite direction (ACI to NX-OS). Policy enforcement occurs in one of two locations:

- At the ACI ingress leaf: This occurs if the leaf has already learned the MAC (and associated pcTag) of the NX-OS endpoint via the data plane, thanks to a previously received frame, allowing it to enforce the policy directly.
- At the ACI Border Gateway (BGW): This occurs when the MAC of the remote NX-OS-connected endpoint is not yet known to the ACI ingress leaf. With bidirectional traffic, this typically happens only when the initial frame is sent from ACI to the NX-OS fabric.

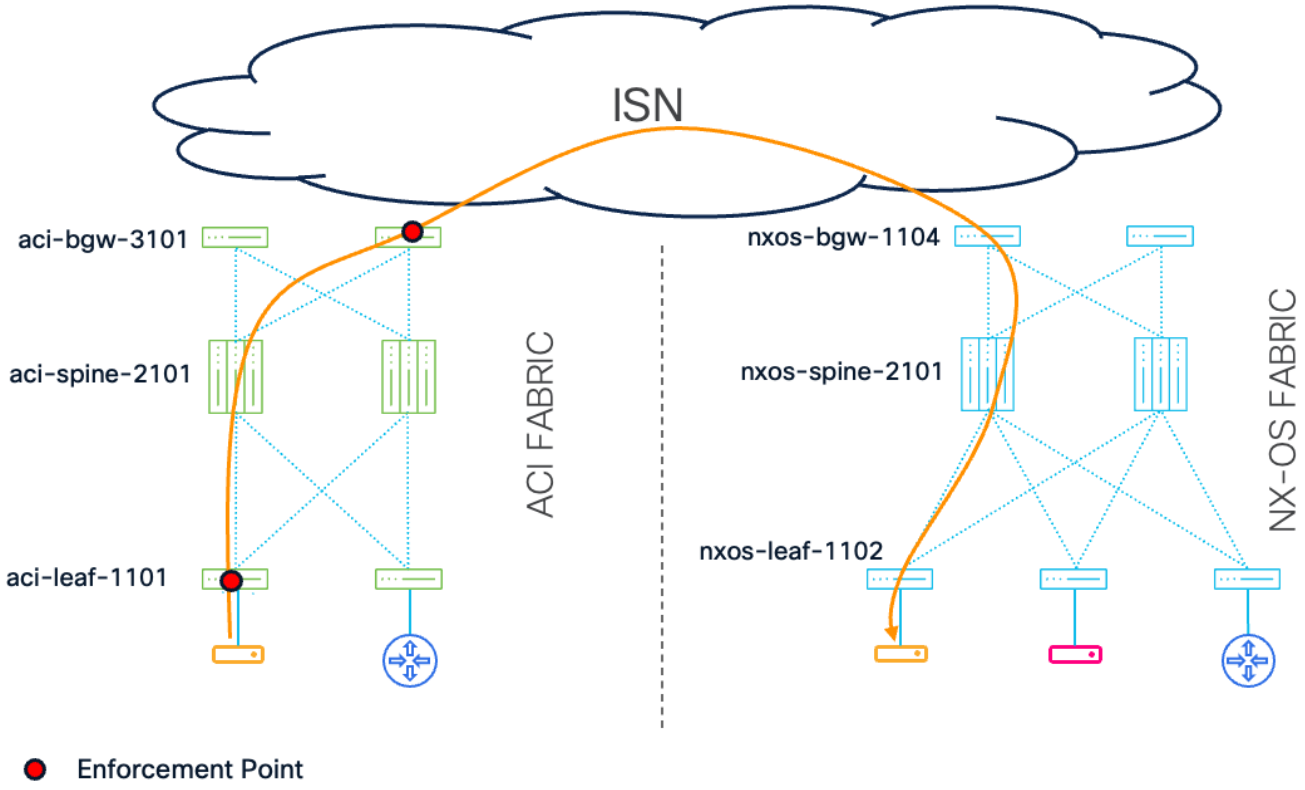

Figure 10 Bridged Traffic from ACI to NX-OS

We will now look at routed traffic moving from an endpoint connected to an NX-OS fabric to one connected to the ACI fabric. In this scenario, the ACI egress leaf serves as the enforcement point. Because ACI-connected endpoints are not present in the Border Gateway (BGW) forwarding tables, but only in the control plane, the BGW routes traffic to the spine-proxy via the pervasive route. The traffic is encapsulated in iVXLAN, and carries the pcTag determined by the BGW ESG classifiers, which allows the egress leaf to enforce the policy.

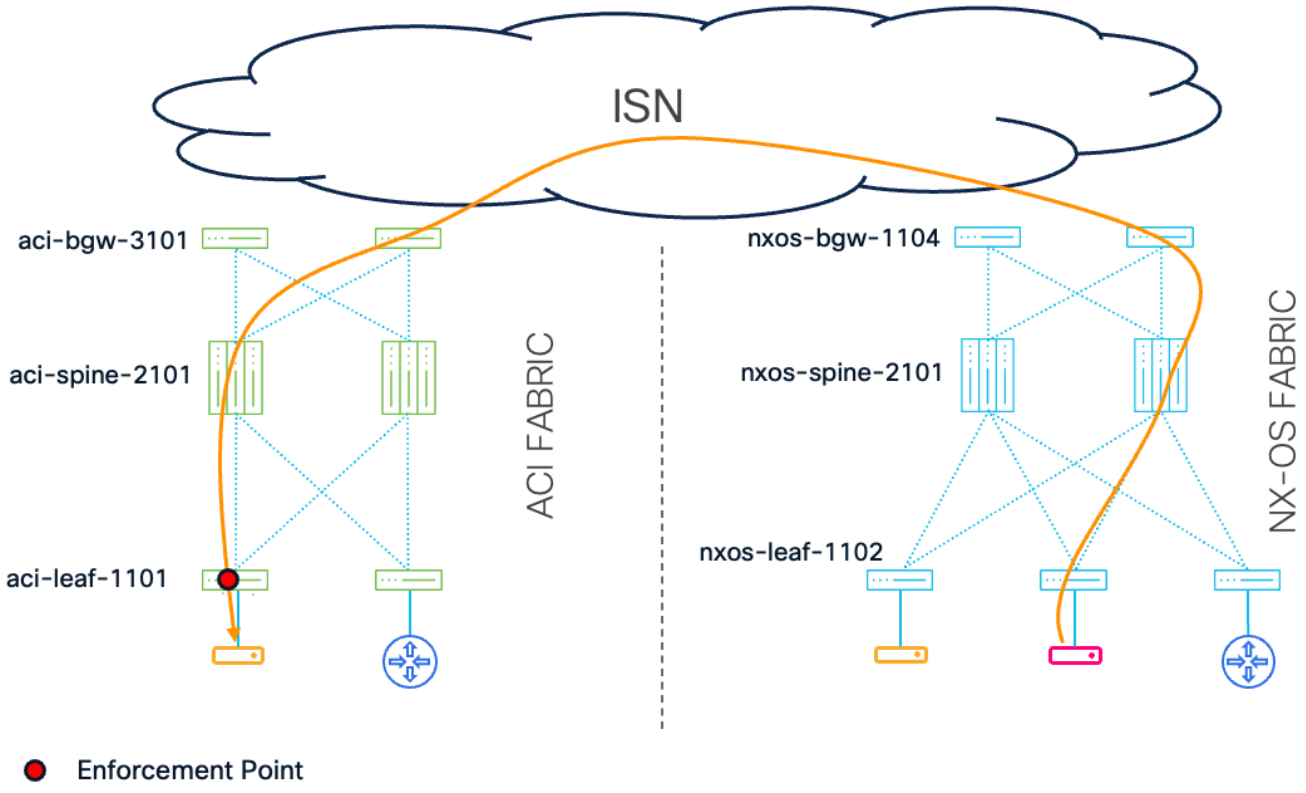

Figure 11 Routed Traffic from NX-OS to ACI (East-West)

Next we will look at the returning traffic from the ACI-connected endpoint to an NX-OS-connected endpoint. In this scenario, the policy might be enforced in two different locations:

- At the ACI leaf, if the NX-OS endpoint was discovered via an ARP packet previously sent. IP data plane packets from remote NX-OS connected endpoints instead are not used for data plane learning by design.
- At the ACI Border Gateway when ACI leaf doesn't have the remote NX-OS endpoint IP locally stored. The BGW then enforces the policy using the destination pcTag derived from the local ESG selector.

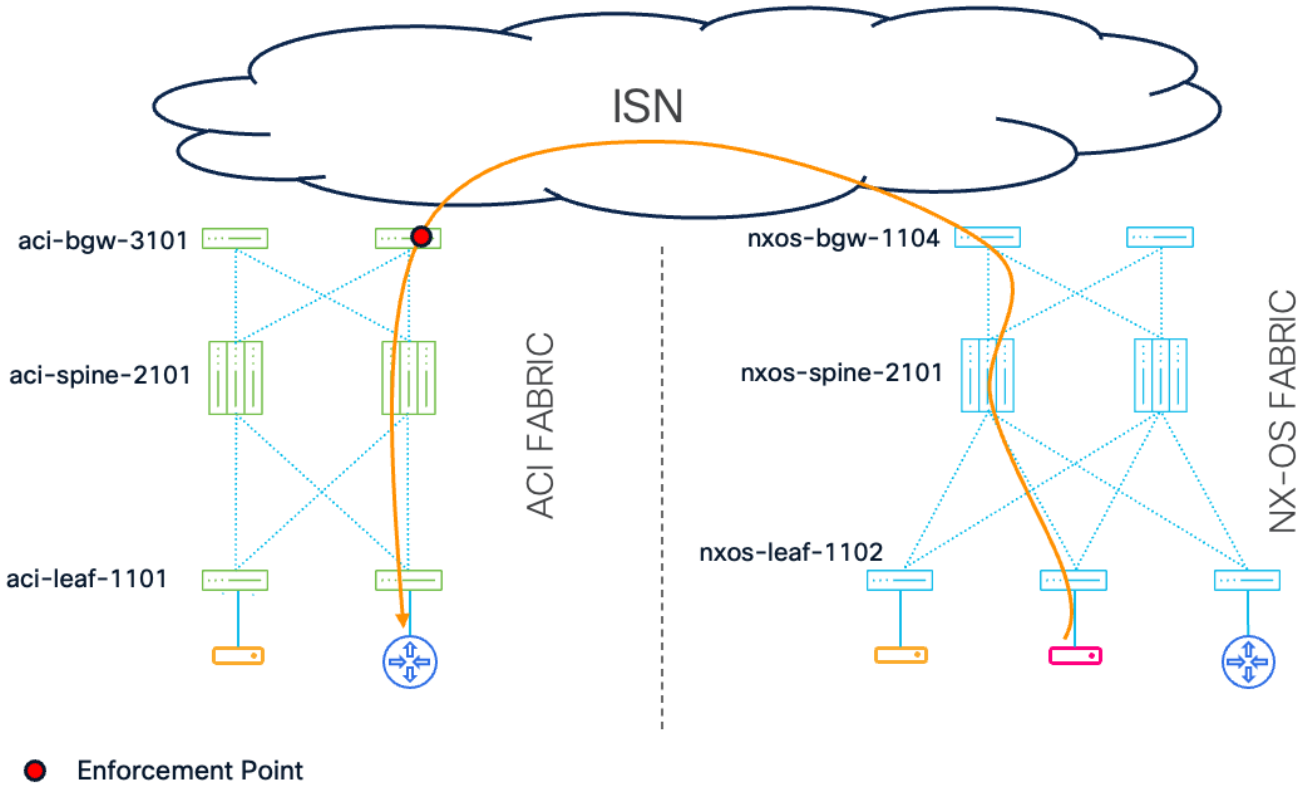

Figure 12 Routed Traffic from ACI to NX-OS (East-West)

The next example covers the scenario where an external client's traffic that ingresses the Multi-Site via the ACI fabric gets enforced. In this scenario the ACI leaf cannot apply the policy when the packet is received from the external domain as the information for the remote endpoint is not locally available. The leaf hence encapsulates the original packet into an iVXLAN and sends it toward the BGW. If the subnet is stretched the packet will be proxied via the spines. When the packet arrives at the ACI BGW this device will read the pcTag set in the iVXLAN header, get the pcTag assigned to the remote NX-OS device and apply the policy.

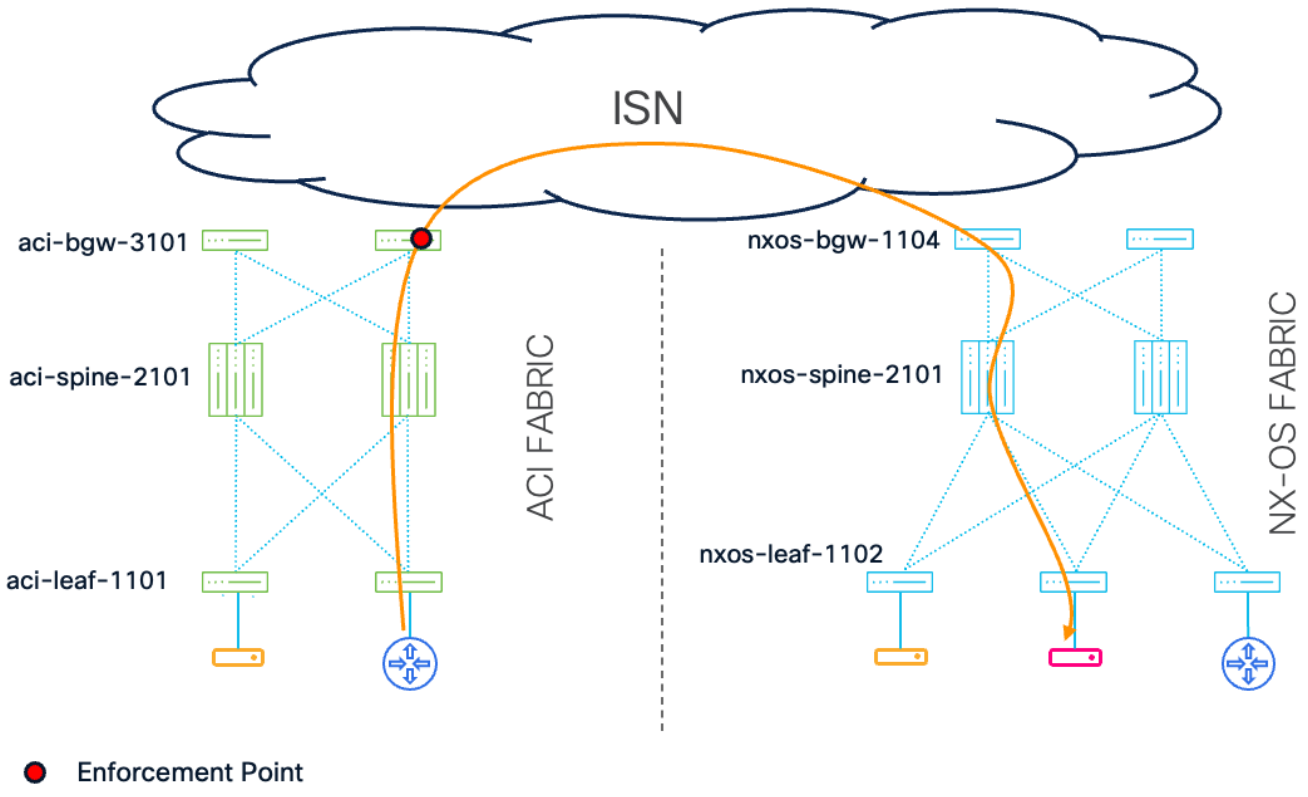

Figure 13 Routed Traffic from ACI to NX-OS (North-South)

The following figure illustrates the returning traffic from the NX-OS connected endpoint to the client connected behind the ACI L3Out. In this scenario, the ACI Border Gateway serves as the enforcement point. Because the BGW, which locally classifies the NX-OS connected endpoint, has also learned the ACI L3Out external subnets via VPNv4 advertisements and their associated pcTags via configuration policies, as normally happens for every L3Out, it is able to apply policy enforcement immediately upon packet receipt from the NX-OS fabric.

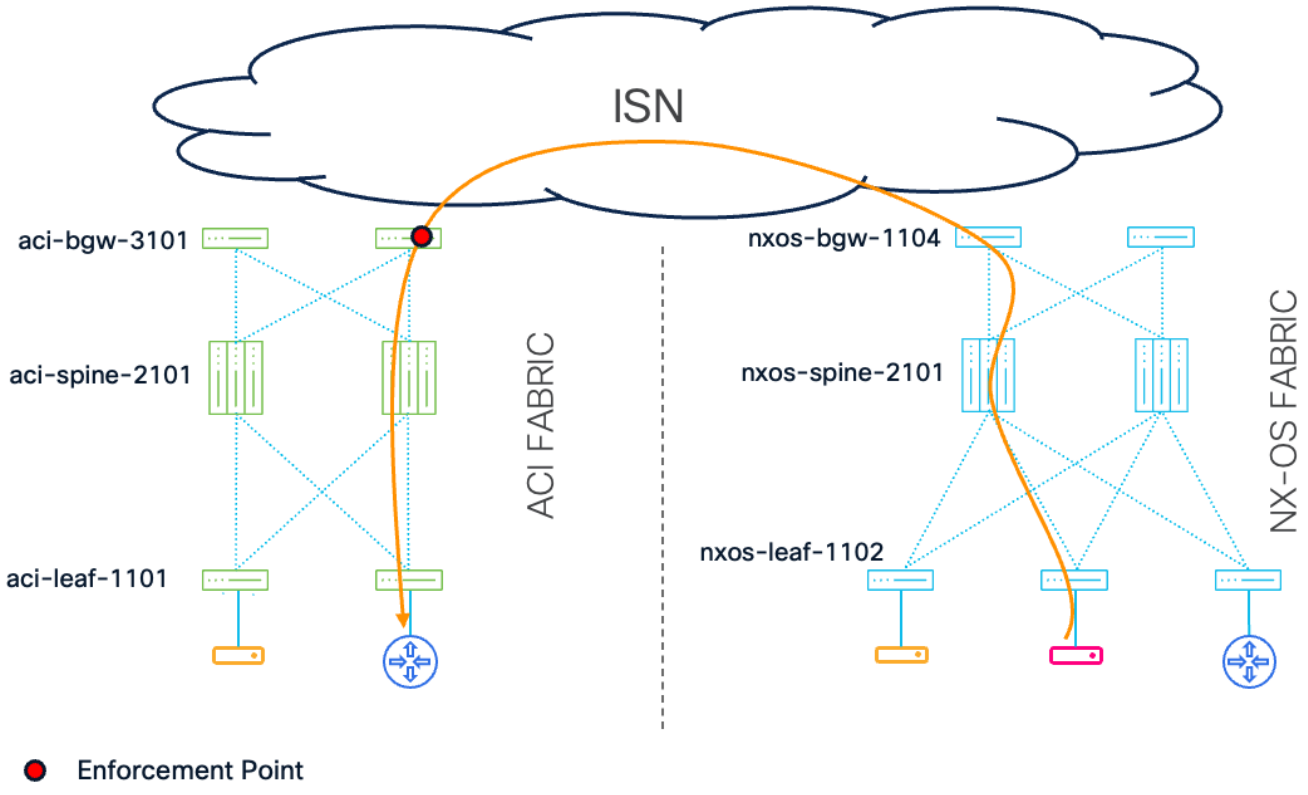

Figure 14 Routed Traffic from NX-OS to ACI (North-South)

The next scenario covers the traffic flow from a client connected behind a L3Out deployed in the NX-OS fabric, to a server attached to the ACI one. In this case the traffic enters the NX-OS fabric and is forwarded to the ACI fabric without any policy enforcement. When the packet hits the ACI BGW it gets stitched to the next iVXLAN tunnel, but the policy enforcement won't be possible. The reason is that the ACI BGW doesn't install the local ACI endpoint into its forwarding tables. Due to this the BGW sets in the iVXLAN header the NX-OS external subnet pcTag derived via the classification rules and sends it toward the egress leaf. The packet transits via the spine proxy and finally gets delivered to the egress leaf. The device at this point can apply the policy as it knows the pcTag associated with the source of the traffic and the one associated with destination.

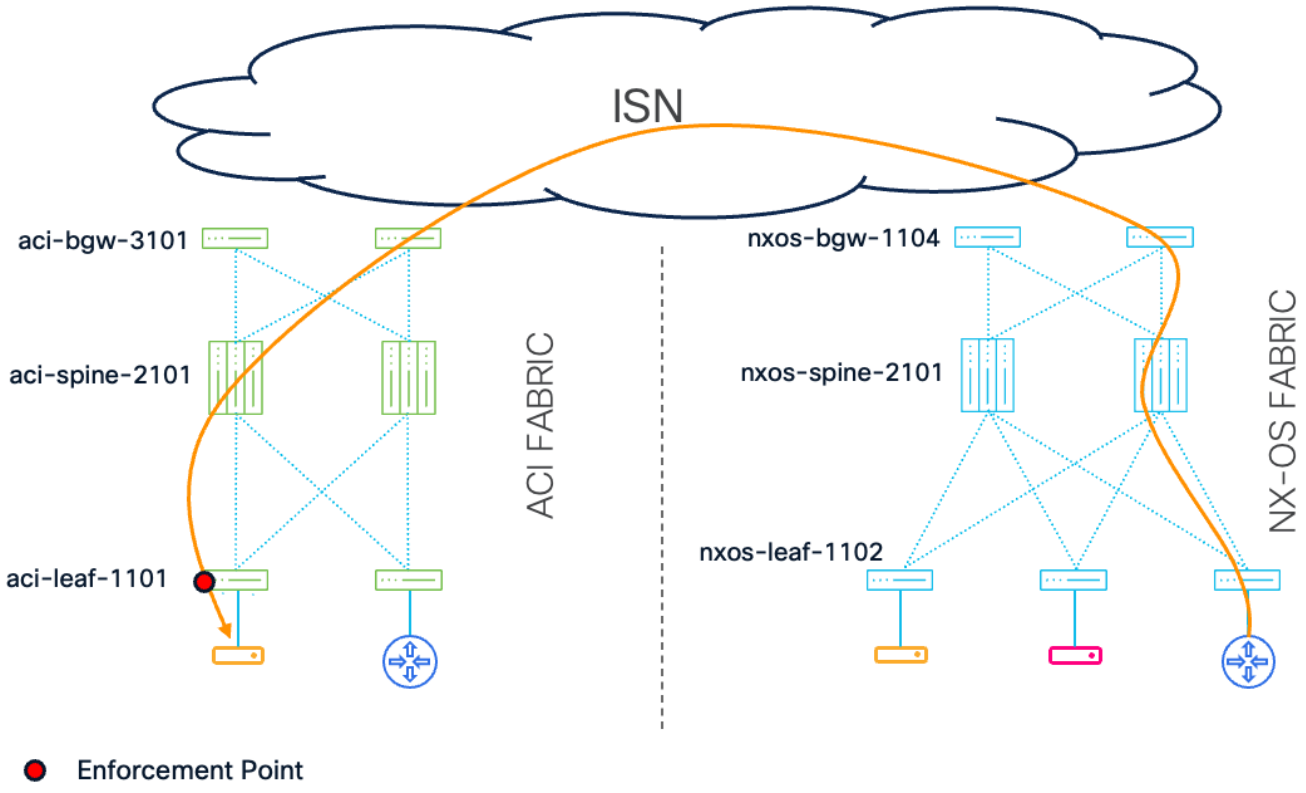

Figure 15 Routed Traffic from NX-OS to ACI (North-South).

We will now look at the returning traffic from the ACI-connected server toward the external client connected behind the NX-OS L3Out. The assumption is that the external client subnet is classified in the ACU ESG with an External Subnet Selector. In this scenario, the ACI ingress leaf serves as the enforcement point. Because the remote NX-OS external route and the related pcTag are deployed in all devices where the VRF is configured via policy means, and not with control plane, same as it happens in classic ACI external EPGs.

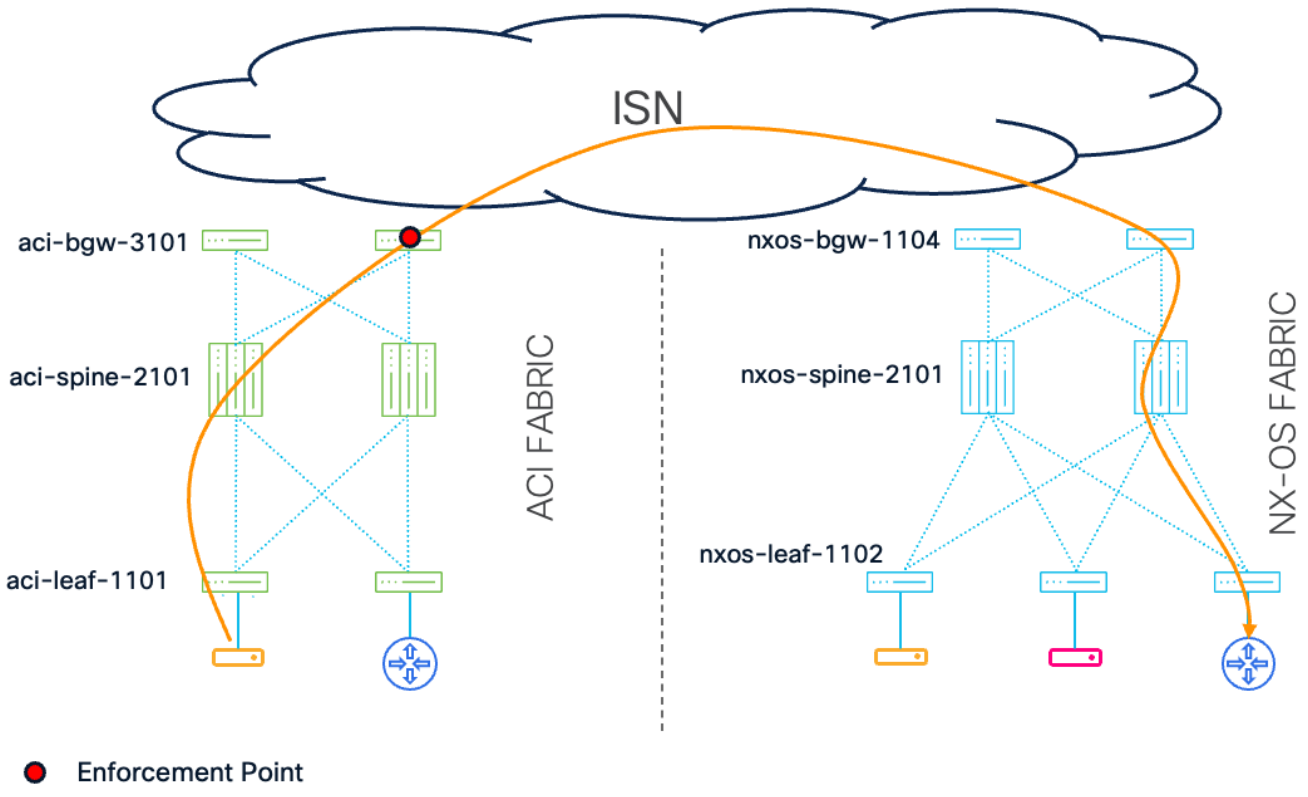

Figure 16 Routed Traffic from ACI to NX-OS (North-South).

The final scenario covers the transit routing use case, where the traffic enters a L3Out in the ACI fabric and egresses a different L3Out connected to the NX-OS devices. The ACI leaf will apply the policy for the traffic entering via the local L3Out interface. As we will discover in the next section, and as we anticipated in the previous example, External Subnet Selectors, which can be used for classifying external routes learned locally or remotely, do not require corresponding control-plane route to exist; as it happens for the External EPG subnets they get configured on the leaves by policy means. In this case the assumption is that policy will be applied because the external subnets will be classified into two different ESGs

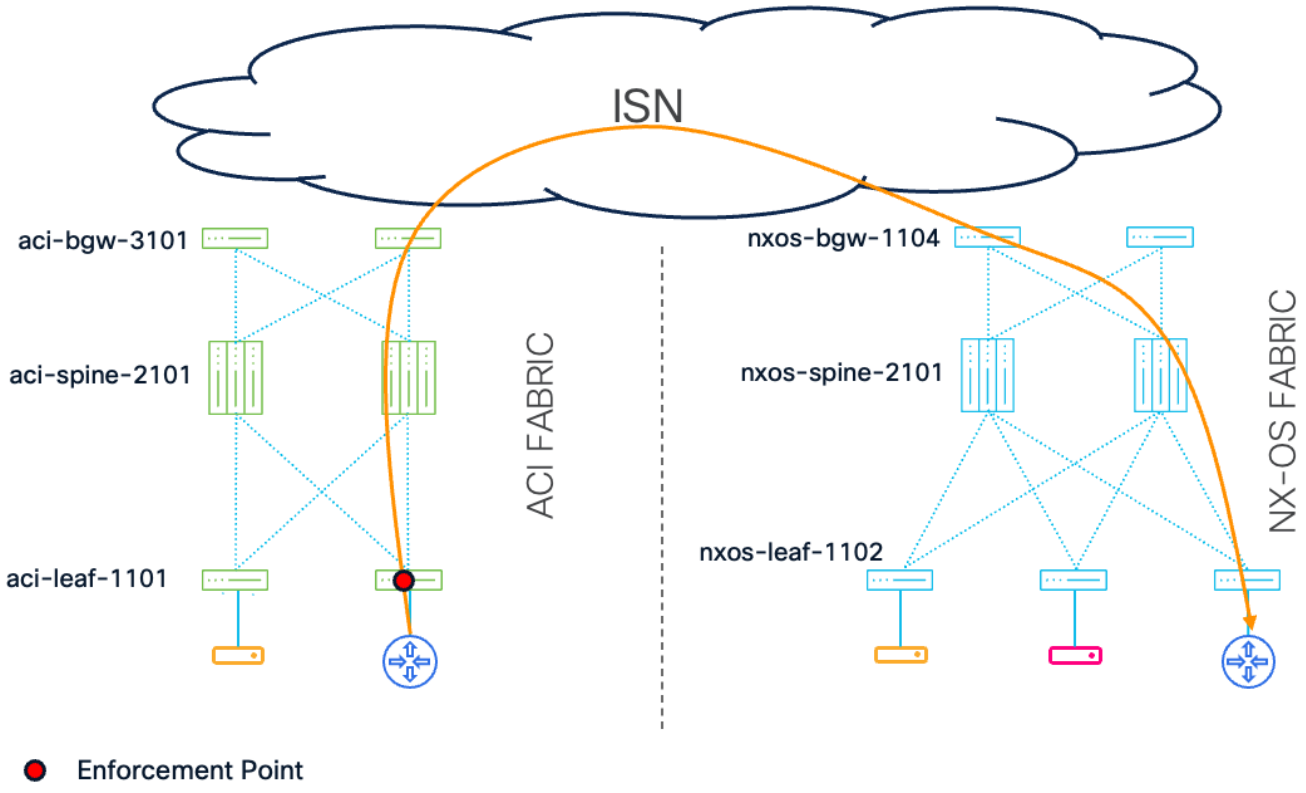

Figure 17 Transit Traffic from ACI to NX-OS

For the returning traffic, entering via the NX-OS L3out and getting out via the ACI L3Out what we have discussed before remains valid. The only difference is that this time the policy will be enforced at the ACI BGW, which will also be programmed with both ESGs External Subnet Selectors and hence will be able to directly lookup both source and destination pcTags on the local zoning-prefix table.

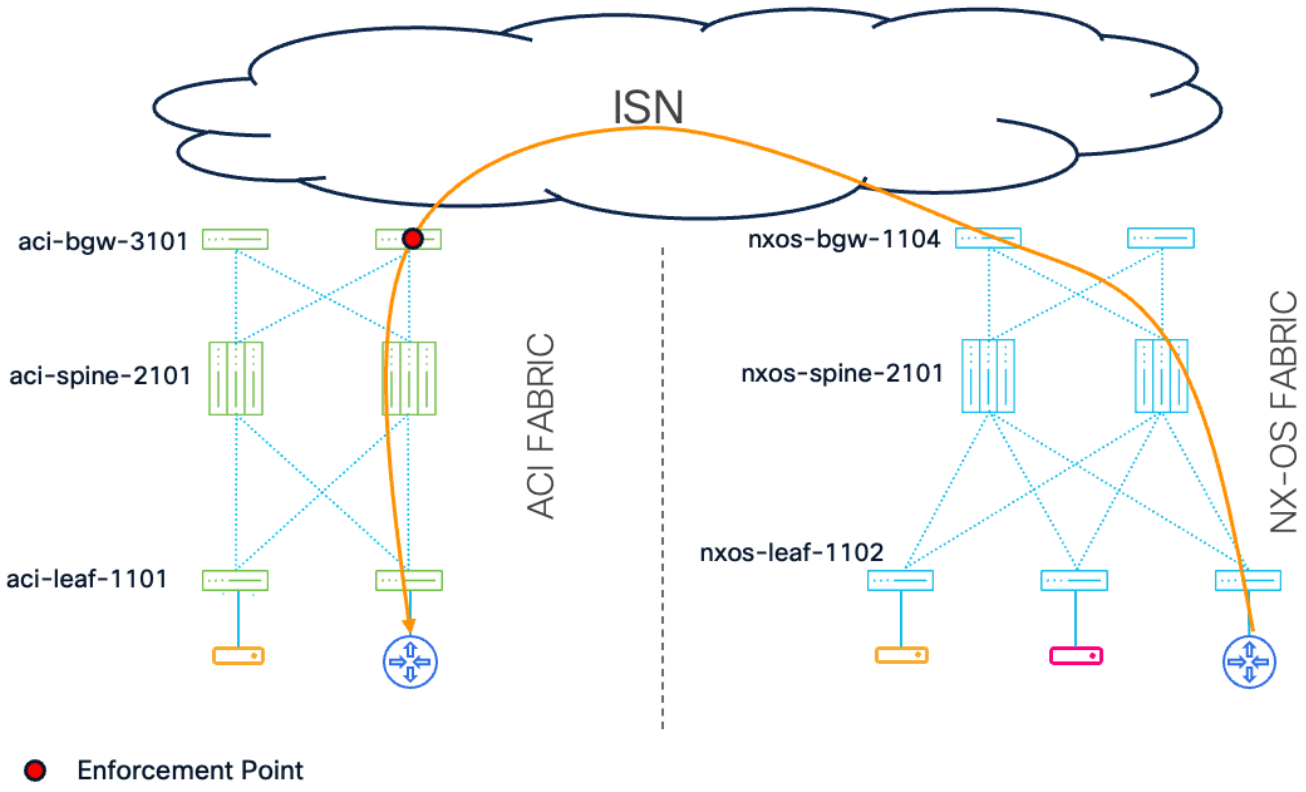

Figure 18 Transit Traffic from NX-OS to ACI

ACI enforcement for Security Groups unaware fabrics

When an ACI fabric is connected to Security Group-unaware fabrics, the ACI BGWs classify all endpoints and prefixes received via EVPN from the remote NX-OS BGWs. This classification is based on configurations made within the ACI APIC which then pushes these policies to the ACI BGW. The core idea is to assign a security class (sclass/pcTag) to incoming traffic, enabling policy enforcement.

Here are the different classification options available and when to use them:

- VXLAN BD Selector:

  This selector classifies **mac-only Type-2** routes, originating from a remote Border Gateway (BGW). Classification occurs after applying a selector that maps to one of the stretched Bridge Domains. For proper operation, the remote NX-OS L2VNI must be attached to the BGWs to ensure Type-2 routes are advertised to the ACI fabric.

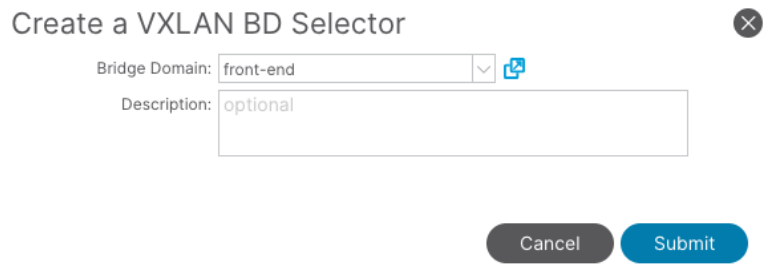

Figure 19 VXLAN BD Selector

All remote MAC addresses received over the BD L2VNI will be classified in the ESG. If more granular control is required, it can be achieved via the Tag Selector later explained. When this selector is used, policy enforcement can only be applied to bridged flows as remote EP IP addresses will not be classified and tagged. For routed traffic use either the Tag Selector or the IP Subnet Selector explained below.

The following outputs represent a common situation where ACI BGW receives a Type-2 route containing both MAC and IP addresses. Using the VXLAN BD Selector causes the IP address to be assigned with sclass 12 which means an unclassified endpoint and points to an implicit drop rule:

```
bgw-3301# show system internal epm endpoint ip 172.16.1.110


MAC : 0200.ac14.0002 ::: Num IPs : 1
IP# 0 : 172.16.1.110 ::: IP# 0 flags : sclass| ::: l3-sw-hit: No ::: sclass : 12
Vlan id : 0 ::: Vlan vnid : 0 ::: VRF name : marvel_tn:prod_vrf
BD vnid : 0 ::: VRF vnid : 2490369
Phy If : 0 ::: Tunnel If : 0x1801000f
Interface : Tunnel15
Flags : 0x10000000004400 ::: sclass : 0 ::: Ref count : 3
EP Create Timestamp : 08/29/2025 12:24:48.609484
EP Update Timestamp : 09/02/2025 08:15:17.400382
EP Flags : IP|timer|vxlan-ext-evpn-ep|
```

In contrast, the MAC address is correctly associated with the local ESG sclass 5475:

```
bgw-3301# show system internal epm endpoint mac 0050.56b6.dfbf
MAC : 0050.56b6.dfbf ::: Num IPs : 0
Vlan id : 7 ::: Vlan vnid : 14974941 ::: VRF name : marvel_tn:prod_vrf
BD vnid : 14974941 ::: VRF vnid : 2490369
Phy If : 0 ::: Tunnel If : 0x1801000f
Interface : Tunnel15
Flags : 0x10000080004800 ::: sclass : 5475 ::: Ref count : 4
EP Create Timestamp : 08/28/2025 21:31:29.982825
EP Update Timestamp : 09/02/2025 08:15:17.405233
EP Flags : MAC|sclass|timer|vxlan-ext-evpn-ep|
```

- Tag Selector

ESG Tag Selectors enable endpoint mapping based on the following attributes:

- IP address (IP Selector)
- MAC address (MAC Selector)
- VM name
- VM tag

Important: The current implementation for VM name and VM tag requires micro-segmentation on the EPGs. uSeg enables proxy-arp which, at the current stage (ACI 6.1(4) and NX-OS 10.6), cannot inter-operate with Nexus One stretched L2VNIs. Due to this these two options will be excluded from the remainder of this explanation.

To enforce the traffic to/from remote IP and MAC addresses administrators are required to map them to “key, value” tags which can be then referenced in the ESG Tag Selector. Mapping can be done manually via the Web-UI or by using automation.

To manually map IP and MAC addresses, users must navigate inside of the Tenant / Policies/ Endpoint Tags and then select the appropriate section

- IP Tag mapping:

  Used for classifying /32 IP addresses advertised by remote BGWs as Type-2 routes. IP Tag mapping can be used to achieve routed traffic policy enforcement to/from a remote endpoint.

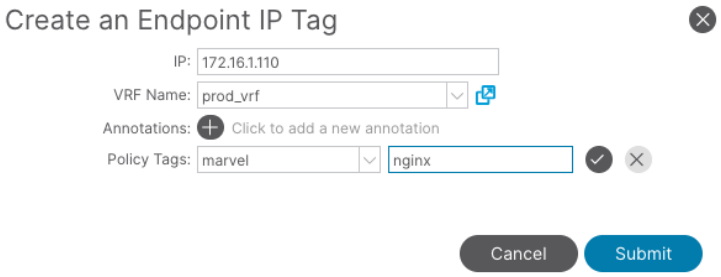

Figure 20 IP Tag mapping

In this example, we map the remote VXLAN fabric endpoint 172.16.1.110 within the prod_vrf to the tag " marvel; nginx." Only the IP address will be classified in this case. If the MAC address associated with the IP has not been classified either using a VXLAN BD Selector or a MAC Tag mapping, bridged traffic shall be dropped.

This output represents an unclassified MAC address, observe the assigned sclass 12 which points to an implicit drop rule:

```
bgw-3301# show system internal epm endpoint mac 0050.56b6.dfbf
MAC : 0050.56b6.dfbf ::: Num IPs : 0
Vlan id : 7 ::: Vlan vnid : 14974941 ::: VRF name : marvel_tn:prod_vrf
BD vnid : 14974941 ::: VRF vnid : 2490369
Phy If : 0 ::: Tunnel If : 0x1801000f
Interface : Tunnel15
Flags : 0x10000080004800 ::: sclass : 12 ::: Ref count : 4
EP Create Timestamp : 08/28/2025 21:31:29.982825
EP Update Timestamp : 09/02/2025 08:40:17.454365
EP Flags : MAC|sclass|timer|vxlan-ext-evpn-ep|
```

In contrast, the IP address is correctly associated with the local ESG sclass 5475:

```
bgw-3301# show system internal epm endpoint ip 172.16.1.110
MAC : 0200.ac14.0002 ::: Num IPs : 1
IP# 0 : 172.16.1.110 ::: IP# 0 flags : sclass| ::: l3-sw-hit: No ::: sclass : 5475
Vlan id : 0 ::: Vlan vnid : 0 ::: VRF name : marvel_tn:prod_vrf
BD vnid : 0 ::: VRF vnid : 2490369
Phy If : 0 ::: Tunnel If : 0x1801000f
Interface : Tunnel15
Flags : 0x10000000004400 ::: sclass : 0 ::: Ref count : 3
EP Create Timestamp : 08/29/2025 12:24:48.609484
EP Update Timestamp : 09/02/2025 08:40:17.449516
EP Flags : IP|timer|vxlan-ext-evpn-ep|
```

- MAC Tag mapping:

  Used for classifying MAC addresses advertised by remote BGWs as Type-2 routes. It has precedence over the VXLAN BD selector for MAC classification. MAC Tag mapping can be used to achieve bridged traffic policy enforcement toward a remote endpoint. As this provides a very granular option, it can be used when the same L2VNI is divided into multiple security-zones.

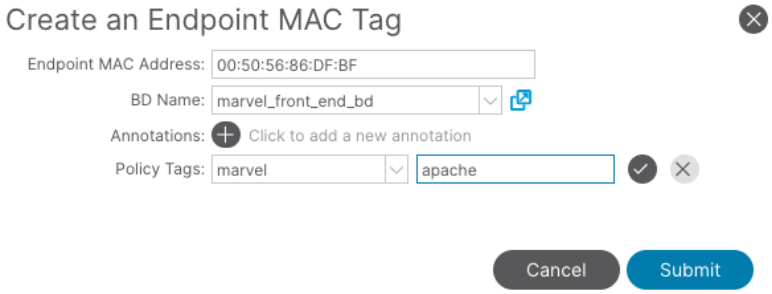

Figure 21 MAC Tag mapping

Tag keys and values can be completely customized according to user preference. It is important to apply common sense when naming tags to ensure they are easily recognizable and meaningful for future reference.

The MAC Tag classification done by the BGW works slightly differently from the MAC Tag classification done at the compute leaf. BGW classifies only the MAC address, whereas the ACI leaves classify the MAC address and all the IP addresses associated with it if no other classifiers do that already.

> Important: Both IP and MAC Tag mapping and classification require Type-2 routes to be received and imported by the ACI BGW. To achieve that the L2VNI must be configured on BGWs in both fabrics. If those are missing the ACI BGW will fail the classification and the traffic enforcement.

- IP Subnet Selector:

  This selector should be used to classify endpoints that are directly connected to the ACI or NX-OS fabric. This selector relies heavily on the routing entries and their EVPN route type. The following paragraph explains the different options and matching conditions.

Important: Although this selector might also be used for matching external remote routes, configurations and troubleshooting might become complex, hence the best practice is to use the External Subnet Selector for such use cases.

Important: This selector only classifies IP addresses, hence switched bridged traffic in a stretched subnet will not be subject to policy enforcement even if source and destination IP addresses fall inside the same ESG. The reason is that ACI, for bridged traffic, performs the EP to pcTag lookup based on the source and destination MAC addresses. This is the reason why for bridged traffic you must use the MAC Tag Selector or the VXLAN BD Selector.

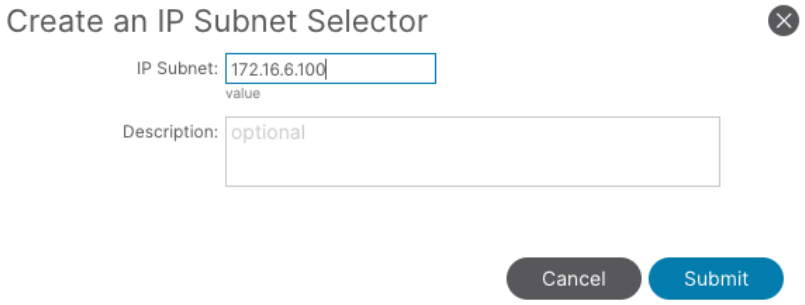

Figure 22 IP Subnet Selector

- Matching a remote connected endpoint IP advertised via a Type-2 route can be done by:
  - Using the exact /32 IP address.
  - Using any larger subnet that overlaps with the remote IP address, including the default route. Pay attention when using larger subnets as it might cause multiple Type-2 routes to end up being classified in the ESG.

Important: Type-2 /32 IP routes will only be announced by NX-OS BGWs when the L2VNI is configured on them. If only L3VNI is configured, then just the SVI/DAG subnet will be advertised as Type-5

When multiple routes and IP Subnet classifiers overlap, the system uses Longest Prefix Match (LPM), consistent with standard ACI endpoint IP matching.

Matching a remote connected endpoint IP advertised only via the Type-5 SVI/DAG route injected by NX-OS leaves can be done by using the exact Type-5. No other method will work as the assumption here is that Type-5 route will always be advertised when the VRF is extended.

The following table provides some quick reference on the remote route classification with the IP subnet selector.

Considering the scenario where the ACI BGW has received from the remote fabric and installed in a tenant VRF the following routes:

	Route	Type	Notes
	172.16.6.100/32	Type-2	EP host route
	172.16.6.200/32	Type-2	EP host route
	172.16.5.100/32	Type-2	EP host route

	Route	Type	Notes
	172.16.5.0/24	Type-5	DAG/SVI subnet
	172.16.6.0/24	Type-5	DAG/SVI subnet
	172.16.7.0/24	Type-5	DAG/SVI subnet

Let's cover some use cases and understand if they are feasible or not

- Use Case 1

Requirement: 172.16.6.100/32 and 172.16.6.200/32 must be classified into two different security groups

Solution: Create two different ESGs, and add in each of them the IP Subnet Selector that matches the exact /32 match received via the Type-2 route

Note: This works as the ACI BGW receives the specific /32 host routes via Type-2 EVPN advertisements.

- Use Case 2

Requirement: All endpoints part of 172.16.6.0/24 must be classified in a single ESG

Solution: Create one ESG and use 172.16.6.0/24 as IP Subnet Selector. This will cause all the Type-2 routes matching 172.16.6.0/24 to fall in the ESG

Notes: Additional solutions could be implemented. For example, adding all the /32 Type-2 routes would work as well, but might require too many configurations.
Adding a major subnet, like 172.16.0.0/16 will also work but it will classify inside the same ESG the other Type-2 route matching, in this case 172.16.5.100/32.

Use Case 3:

Requirement: 172.16.7.100/32 and 172.16.7.200/32 must be classified into two different security groups

Solution: With the IP Subnet Selector this cannot be achieved as the /32 Type-2 routes are not in the BGW routing table. A working solution would require either the advertisement of the Type-2 routes for the 172.16.7.0/24 network from the NX-OS BGWs, or the use of the External Subnet Selector which allows administrators to classify any arbitrary subnet without a corresponding entry in the control plane.

- External Subnet Selector:

This selector was known as "VXLAN External Subnet Selector" in ACI 6.1(2) and 6.1(3). Starting from 6.1(4) it was renamed as "External Subnet Selector" as it can also be used to classify L3Out learned routes directly in ESGs.

On the BGW, External Subnet Selector can be used to classify remote external subnets that are received as Type-5 EVPN routes; it can also be used to classify part of those subnets in case more granular control is required even if the smaller subnet is not present in the routing table. A Longest Prefix Match applies also in this case.

Default route 0.0.0.0/0 can be used as External Subnet Selector starting from ACI 6.1(5)M, for previous releases, as a workaround, the subnets 0.0.0.0/1 and 128.0.0.0/1 can be used.

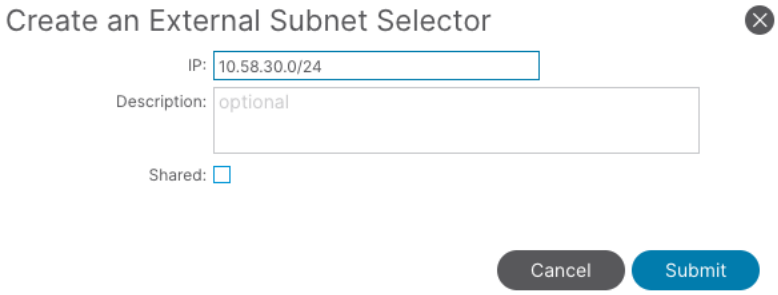

Figure 23 External Subnet Selector

External Subnet Selector, as classic L3Out External EPGs, map the CIDRs to pcTags in the zoning-prefixes table that can be displayed with the command below. The scope, which is the VRF VNI, can be optionally used to limit the command output

```
bgw-3301# show zoning-prefixes  scope 2490369
+---------+--------------------+---------------+-------+-----------+
| Vrf-Vni |      Vrf-Name      |    Address    | Class | OperState |
+---------+--------------------+---------------+-------+-----------+
| 2490369 | marvel_tn:prod_vrf |   0.0.0.0/0   |   15  |  enabled  |
| 2490369 | marvel_tn:prod_vrf |      ::/0     |   15  |  enabled  |
| 2490369 | marvel_tn:prod_vrf | 10.58.30.0/24 |  5480 |  enabled  |
+---------+--------------------+---------------+-------+-----------+
```

We have now covered all the different options available to classify remote endpoints and external subnets, so we can briefly describe how to use these classification methods. The ACI whitelist model, by default, drops all traffic within a VRF; this behavior applies to traffic flowing between ACI and NX-OS fabrics and vice-versa. To allow routed and bridged traffic between different endpoints, administrators can classify them inside the same ESG, which permits traffic by default within that group. Alternatively, endpoints can be classified into different ESGs, and contracts can be applied to allow any or specific traffic flows depending on the filters configured.

There is no exception to this model. To enable traffic between endpoints distributed across different fabrics, the administrator can either classify them inside the same ESG or apply the appropriate contracts between different ESGs to permit the desired traffic flows.

There are no limitations in having remote endpoints and external subnets classified within the same ESGs as ACI endpoints. This design is valid for customers who do not require any demarcation between groups of endpoints in different fabrics. For customers who prefer to maintain a clear separation between ACI endpoints and remote fabric endpoints, additional ESGs and contracts will be necessary.

Important: The decoupling of the data-plane and policy-plane provides significant flexibility in enforcing traffic policies between different endpoints, regardless of the connection model associated with their flows. Whether the traffic consists of bridged frames or routed packets, it is evaluated by the ASICs and either permitted or denied. Since, in this white paper chapter, all traffic inspection occurs on the ACI side, it is crucial to understand the differences between bridged and routed traffic policy enforcement and to use the appropriate Endpoint Security Group (ESG) selectors accordingly.

A broader discussion on policy enforcement between ACI and NX-OS policy aware fabrics is provided in the chapter End-to-End Connectivity and Policy Enforcement

By now let's review some use cases where traffic will only be enforced on the ACI side, as either the remote fabric is not policy aware, or ACI is running either 6.1(2) or 6.1(3) releases. This will provide the reader a better understanding of the classification methods offered by ACI BGW.

Note: Most of the examples coming next will be showing ESGs bounded within a single Bridge Domain. This is purely done to simplify the diagrams; in ACI, ESGs can classify endpoints across multiple BDs, this is also true in the context of Nexus One topologies.

It is important to understand that the below use cases are applicable when ACI runs on version 6.1(2) and 6.1(3) and/or when the NX-OS fabric is non-policy aware. For policy-aware NX-OS fabrics there won't be need to classify NX-OS resources as classification already happens there and gets propagated via EVPN. Additional information on this is available in the End-to-End Connectivity and Policy Enforcement

chapter.

- **Use Case 1**: Network-Based group segmentation with non-stretched subnets
  - **Scenario:**

    Two fabrics with local subnets; VRF extended across sites. Traffic allowed between specific subnet pairs classified into matching Endpoint Security Groups (ESGs). Only routed traffic flows across the BGWs since subnets are local to each fabric. Classification for remote subnets is done via IP Subnet Selectors.

  - Diagrams:

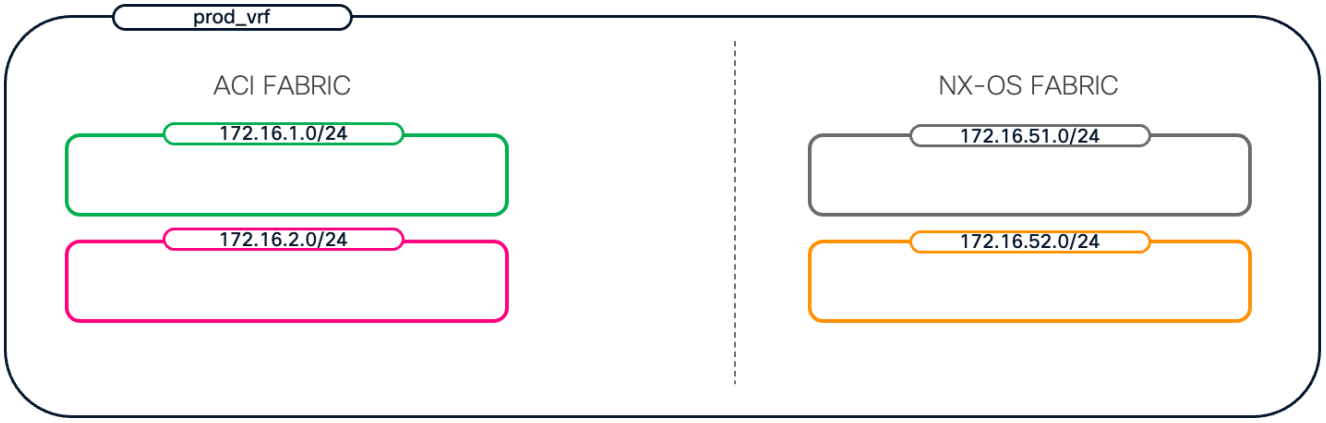

Figure 24 Forwarding

In this scenario each fabric has two subnets locally defined. Forwarding policies have already been configured by extending the VRF across sites.

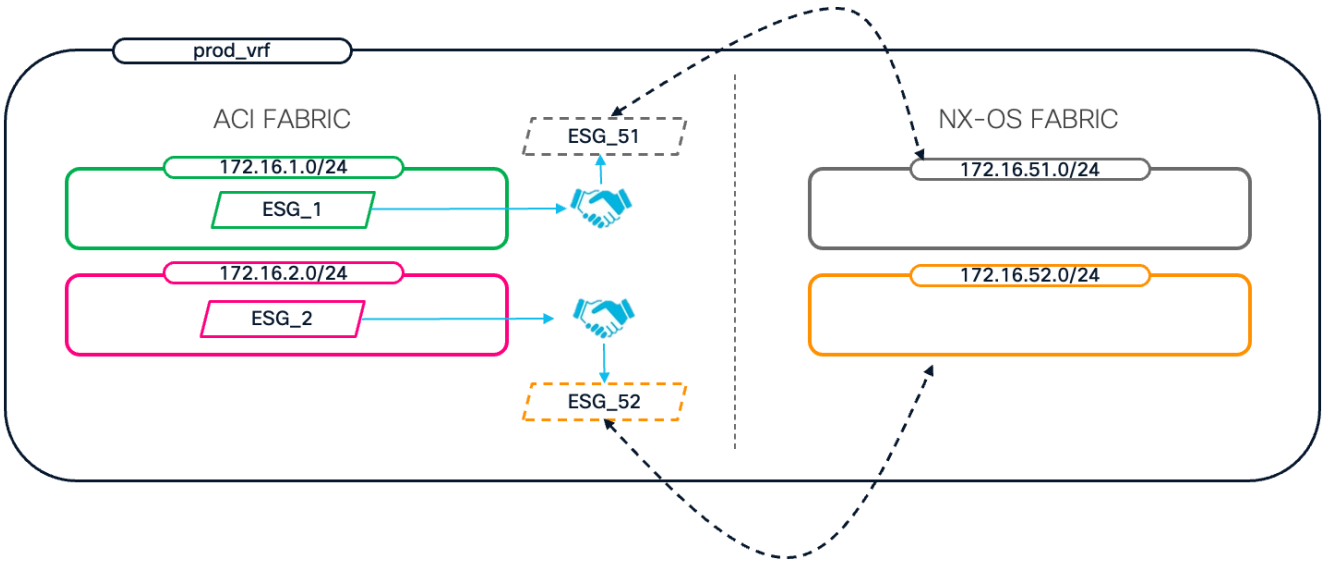

Figure 25 Forwarding and Security

The goal is to permit communications from all endpoints within the ACI 172.16.1.0/24 subnet which are classified in ACI ESG_1 to the 172.16.51.0/24 NX-OS subnet endpoints, and from the 172.16.2.0/24 subnet, classified in ACI ESG_2, to the 172.16.52.0/24 subnet. A straightforward approach to achieve this is by classifying the 172.16.51.0/24 subnet within ESG_1 using an IP Subnet Selector and similarly classifying the 172.16.52.0/24 subnet within ESG_2. This method enables unrestricted communication between the 172.16.1.0/24 and 172.16.51.0/24 subnets, as well as between the 172.16.2.0/24 and 172.16.52.0/24 subnets, since endpoints within each subnet pair are considered part of the same security group. However, if traffic must be filtered based on specific policies, administrators need to define dedicated ACI ESGs classifying the NX-OS subnets and apply the appropriate contracts and filters between them.

ACI to ACI: Traffic will be enforced. According to the scenario all traffic from ESG_1 to ESG_2 will be dropped as there is no contract.

ACI to NX-OS (and vice-versa): Traffic will be enforced

NX-OS to NX-OS: Traffic will NOT be enforced

- Prerequisites:

Endpoint Security Groups (ESGs) are defined in the APIC for the stretched VRF and use appropriate selectors for classifying ACI-attached endpoints. Intra-ACI policy enforcement is already active.

NX-OS networks (L2VNIs) may or may not be attached to the NX-OS BGWs. Attaching them triggers Type-2 route advertisement to ACI BGWs, but this is not mandatory since, in this use case, ESG classification can also be done on the Type-5 route representing the entire L2VNI subnet.

- Considerations:

Remote endpoints will be classified with their entire subnet as this will be the only EVPN advertisement sent to the ACI BGWs.

Since subnets are local to each site, no bridged traffic traverses the ACI BGW; therefore, policy enforcement relies solely on IP classifiers.

- Notes:

  Endpoints attached to remote NX-OS fabrics may communicate freely, assuming the remote fabric is not policy-aware.

- Solution:

  In ACI, configure ESG_51 and ESG_52 to classify the remote subnets into them using the IP Subnet Selector.

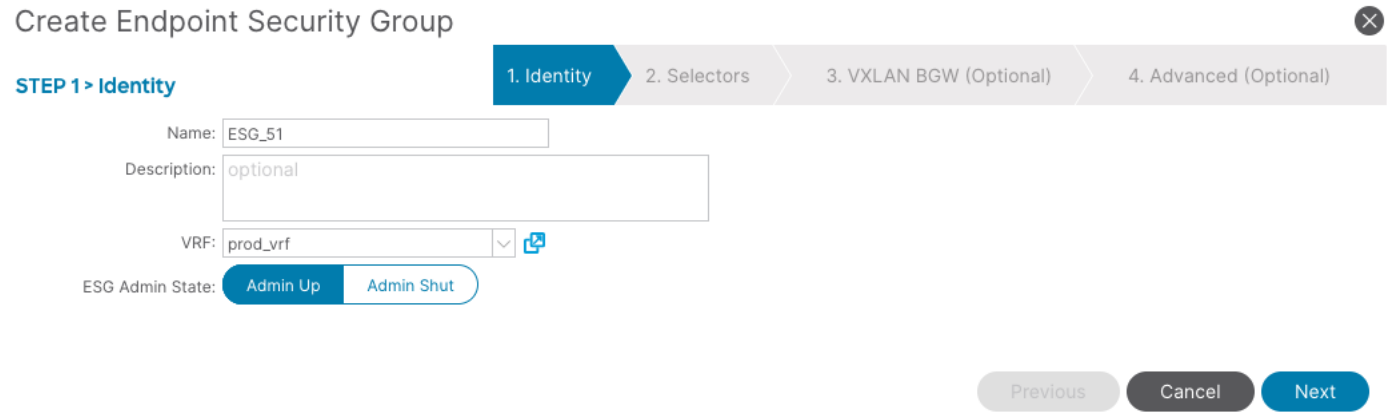

Figure 26 Define a new ESG

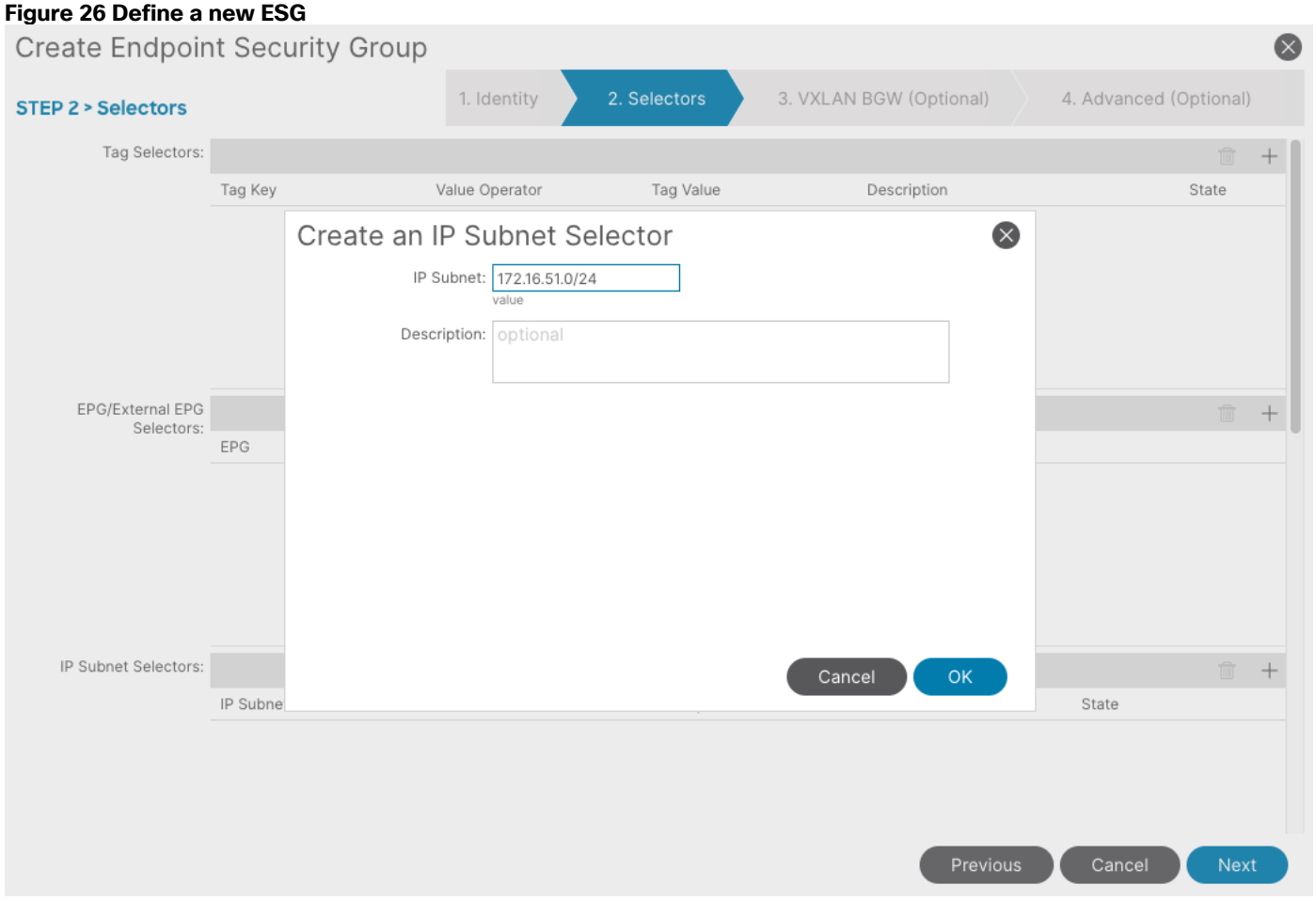

Figure 27 Classify the remote subnet as IP Subnet Selector

Define appropriate filters and contracts to permit traffic flows between different security groups as required.

Use the provider and consumer relationships for ESG_1 and ESG_51 with the first contract, and for ESG_2 and ESG_52 with the second contract.

- Solution Notes:

  Since L2VNIs are not attached to NX-OS BGWs, Type-2 routes for local endpoints are not advertised, making IP Tag Mapping infeasible.

  Using the IP Subnet Selector simplifies APIC configuration by requiring only a single entry per remote subnet, reducing administrative overhead and potential misconfigurations. As previously mentioned, an exact match between the IP Subnet Selector prefix and a Type-5 route received from remote fabric BGWs must be found.

  This approach aligns with the local subnet deployment model, ensuring scalable and manageable policy enforcement without unnecessary complexity.

- **Use Case 2**: Network-Based group segmentation with stretched and non-stretched subnets
  - Scenario:

    One L2VNI subnet is stretched between fabrics, enabling both bridged (Layer 2) and routed (Layer 3) traffic across sites. VXLAN BD Selector is used for bridged traffic and IP Subnet Selector for routed traffic, avoiding MAC and IP Tag Mapping to reduce operational overhead while maintaining enforcement.

  - Diagrams:

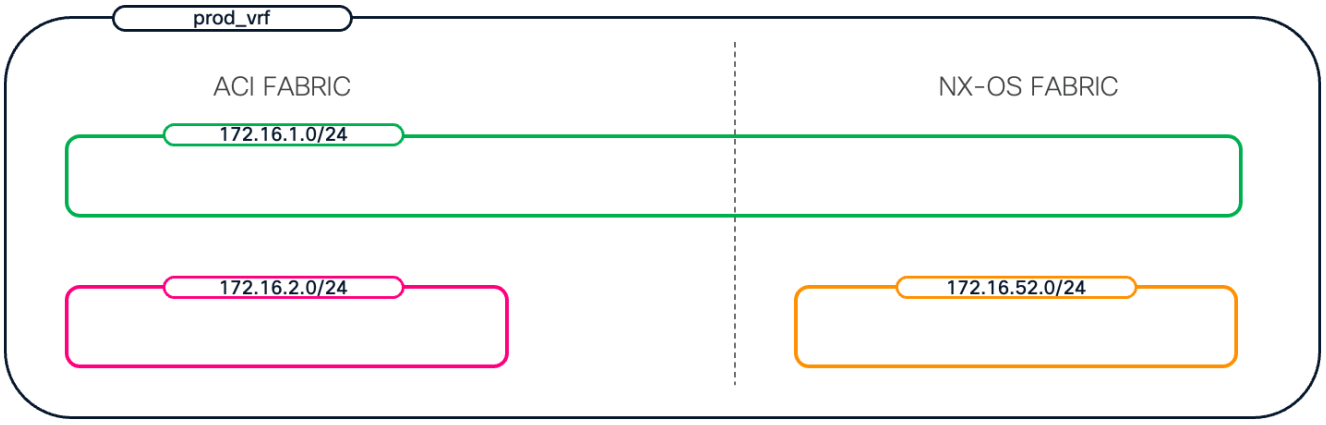

Figure 28 Forwarding

In this use case, one of the previously local L2VNIs is now stretched between the fabrics. This is achieved by locally defining the BD/L2VNI on the local BGWs. Therefore, both the ACI and NX-OS BGW fabrics re-originate the EVPN Type-2 routes for the same BD/L2VNI. Stretching the L2VNI enables bridged traffic between endpoints attached to it, regardless of their physical location.

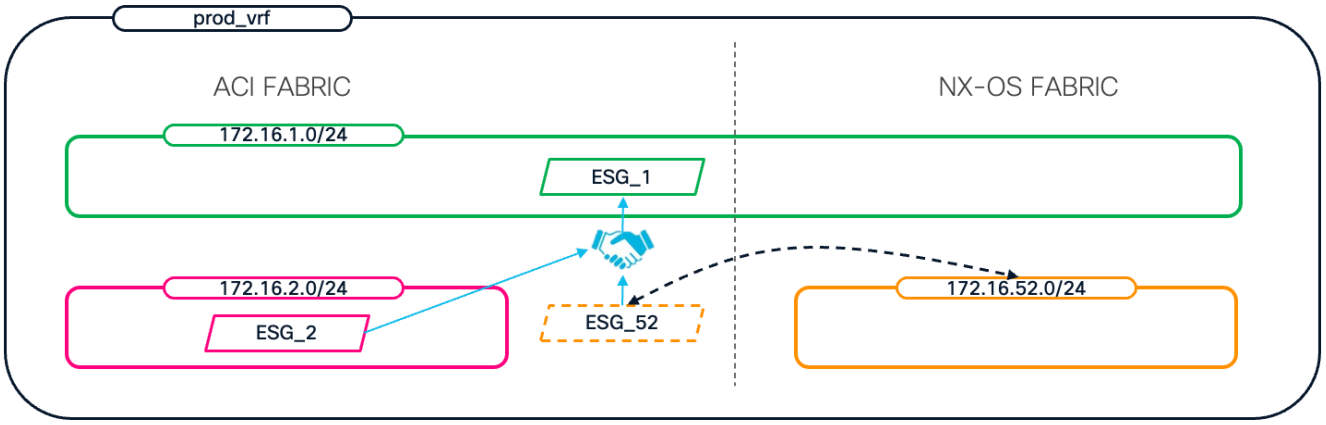

Figure 29 Forwarding and Security

All endpoints attached to the 172.16.1.0/24 network will be classified inside ESG_1, 172.16.2.0/24 into ESG_2 and 172.16.52.0/24 in ESG_52. Again, as the NX-OS is not policy-aware, ESG_52 will only be defined in ACI to enforce security policies between the ACI endpoints and the remote 172.16.52.0/24 endpoints.

The goal in this use case is to allow ESG_1 endpoints to access the ESG_2 and ESG_52 ones over a limited number of ports that will be dictated by the contract filter.

Note: This example illustrates the inconsistent behavior of policy enforcement depending on the physical location of the endpoints. Traffic flowing between endpoints where at least one endpoint is physically attached to the ACI fabric will be subject to policy enforcement. However, traffic between endpoints attached to the NX-OS fabric will not be enforced, since NX-OS in this scenario does not utilize security-groups for enforcement.

ACI to ACI: Traffic will be enforced

ACI to NX-OS (and vice-versa): Traffic will be enforced

NX-OS to NX-OS: Traffic will NOT be enforced

- Prerequisites:

Required VRFs and Bridge Domains are stretched across fabrics. This is achieved by deploying the VRFs/BDs on the local BGW nodes (in the ACI and NX-OS domains).

ACI ESGs are defined and used for endpoint classification within the ACI fabric.

ACI endpoints are classified as part of ESG_1 and ESG_2 using the EPG selector.

Note: For bridged traffic policy enforcement, ACI endpoints must be classified using either the EPG selector or the MAC Tag selector. Please note that, in ACI fabrics, IP-based classification for bridged traffic with the help of micro-segmentation feature is currently unsupported in networks stretched via BGWs. Future releases will introduce consistent proxy-ARP functionality across architectures, which will enable the use of IP-based classifiers for bridged traffic.

- **Considerations:**

Endpoints within the same stretched subnet are distributed across multiple sites but classified into the same ACI security group.

As one of the networks is stretched, BGWs advertises MAC-only and MAC/IP Type-2 EVPN routes to their remote counterparts.

- **Solution:**

For bridged traffic within the 172.16.1.0/24 subnet, use VXLAN BD Selector to ensure traffic between endpoints is correctly evaluated based on MAC address. Not classifying remote MAC addresses will block bridged traffic. MAC Tag Mapping is not required here since endpoints in the same subnet (L2VNI) are not distributed across multiple security groups.

For routed traffic, use IP Subnet Selector matching the L2VNI subnet to ensure traffic between endpoints in different subnets is correctly evaluated based on IP to pcTag mapping. More specific /32 IP Subnet Selectors or IP Tag Mapping are unnecessary as endpoints in the same subnet are not divided into different security groups.

Define appropriate filters and contracts to permit traffic flows between different security groups as required.

Use the provider and consumer relationships for contract between ESG_2, ESG_52 and ESG_1.

- Solution Notes:

VXLAN BD Selector supports bridged traffic policy enforcement across stretched subnets by mapping MAC addresses to security classes, ensuring consistent policy enforcement.

IP Subnet Selector provides efficient routed traffic classification without excessive granularity, simplifying management. Especially in this case where one security group spans the entire subnet the selector is highly efficient.

Avoiding MAC Tag Mapping and IP Tag Mapping reduces complexity while meeting security requirements for this scenario.

This approach balances enforcement precision and operational simplicity in stretched subnet environments where endpoints in the same subnet share the security group.

- **Use Case 3**: Network-Based group segmentation with non-stretched subnets with additional remote external subnets
  - Scenario:

  Builds on Use Case 1 by adding external connectivity via the NX-OS fabric with remote external subnets propagated to the ACI fabric. Traffic includes routed flows to external subnets. External Subnet Selectors classify remote external subnets, enabling granular control and consistent policy enforcement across local and remote domains.

  - Diagrams:

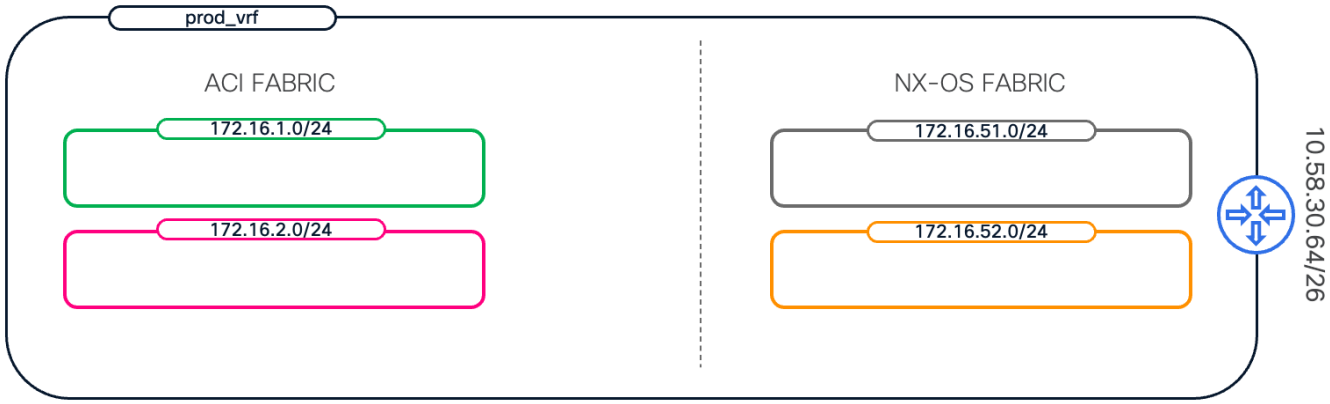

Figure 30 Forwarding

Compared to Use Case 1, in this topology we are adding some external connectivity via the NX-OS fabric. As the VRF is extended over the Nexus One architecture the subnet 10.58.30.64/26 will be automatically propagated to the ACI fabric.

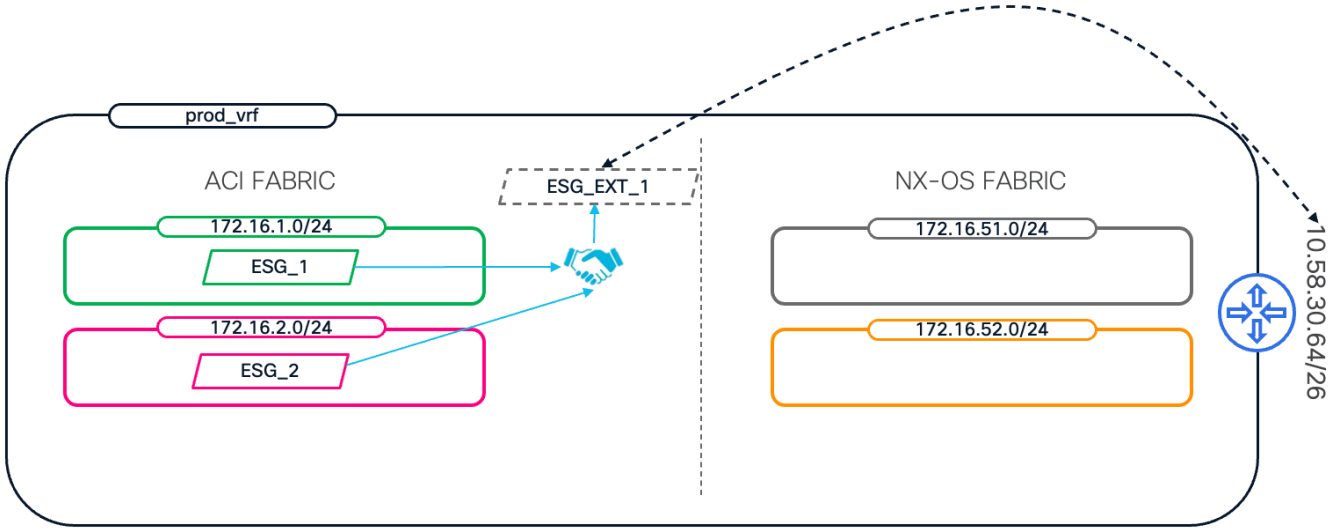

Figure 31 Forwarding and Security

The intention this time is to selectively allow traffic between ACI ESG_1 and ESG_2 and the external network which in this case will be seen by ACI as an additional ESG (ESG_EXT_1) which will classify the 10.58.30.64/26 using the External Subnet Selector.

- Considerations:

This use case continues from Use Case 1, adding classification and enforcement for remote external subnets.

External subnets require classification to ensure consistent policy enforcement across local and remote domains.

- Solution:

Create a new ESG named ESG_EXT_1 and classify the 10.58.30.64/26 subnet using an External Subnet Selector.

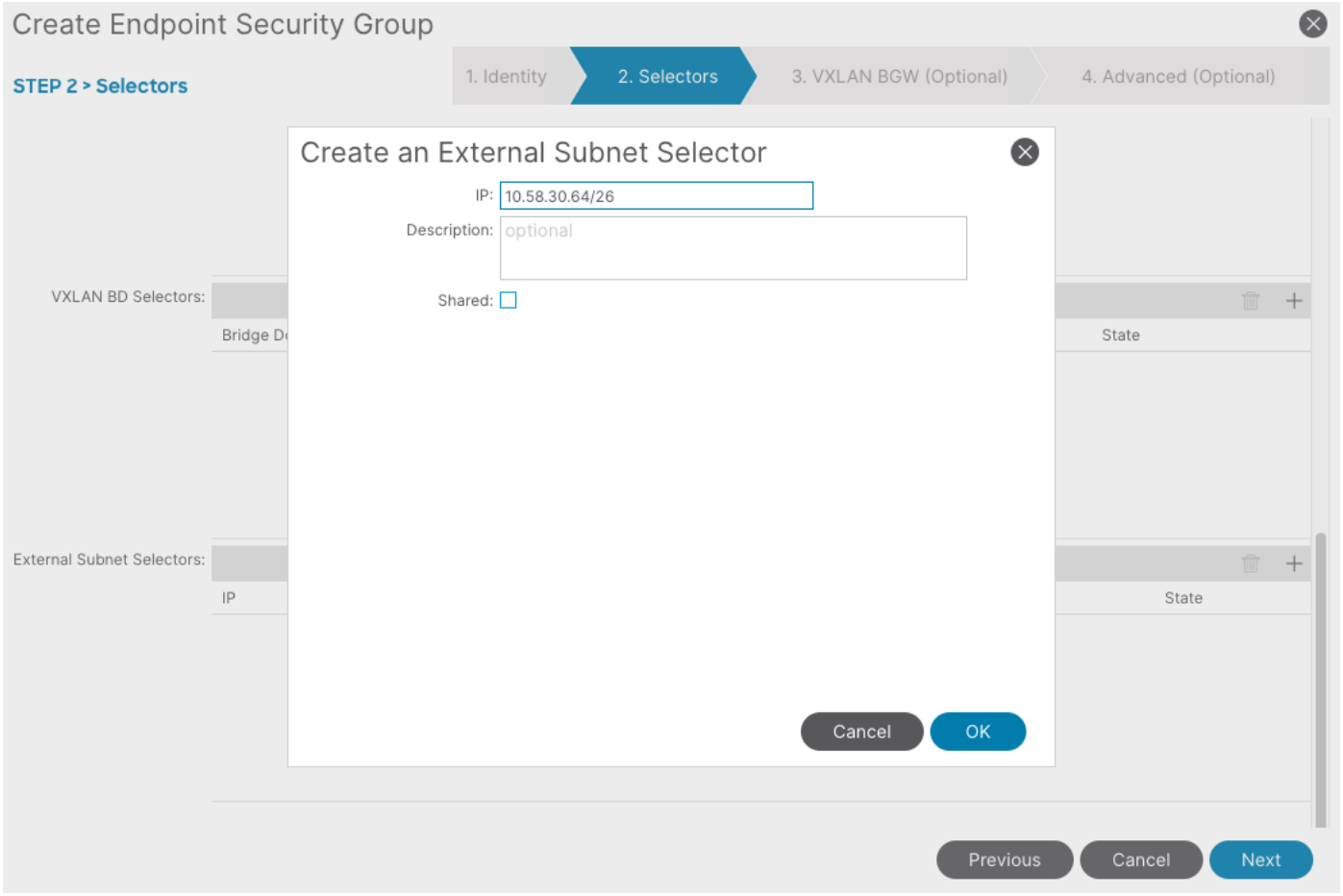

Figure 32 Classify the remote external subnet with an External Subnet Selector

Define appropriate filters and contracts to permit traffic flows between different security groups as required.

Use the provider and consumer relationships for contract between ESG_1, ESG_2 and the ESG_EXT_1.

- Solution Notes:

  The External Subnet Selector provides granular control over external routes, enabling precise segmentation and security posture. Subnet prefixes configured as External Subnet Selector do not need to have exact matches in the routing table which means that administrators can define smaller subnets or even super-nets, excluding the default routes as previously mentioned.

  This extension maintains the simplicity of Use Case 1 while broadening the scope to include external connectivity.

- **Use Case 4:** Host-Based group segmentation within stretched and non-stretched subnets.
  - **Scenario**:

Endpoints within the same stretched subnet are assigned to different security groups, requiring endpoint-level differentiation. Both bridged (Layer 2) and routed (Layer 3) traffic must be granularly classified and enforced to maintain strict micro-segmentation.

This use case involves higher complexity in policy enforcement and classification compared to previous use cases, requiring careful planning and management.

- **Diagrams**:

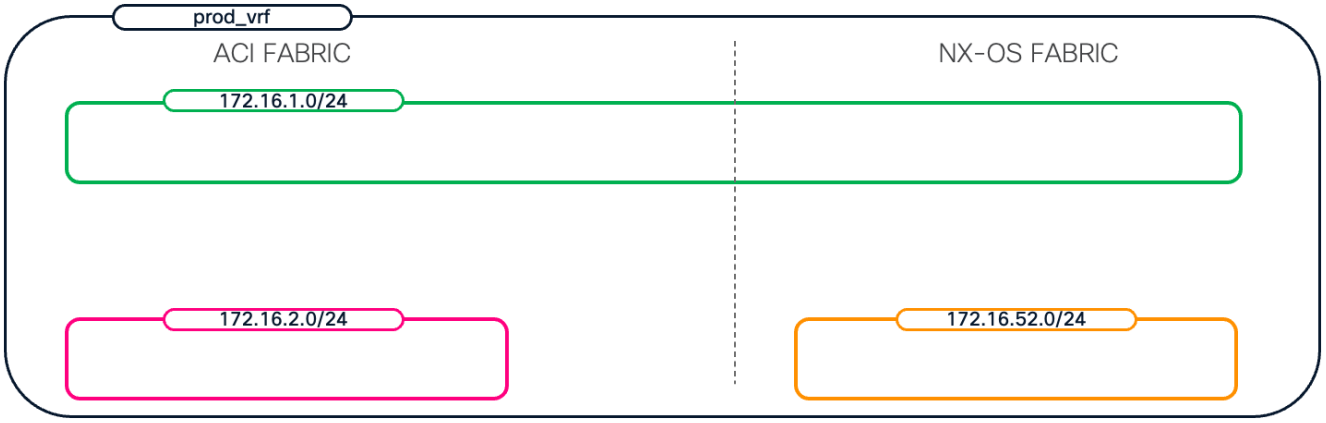

Figure 33 Forwarding

Forwarding scenario is the same as the previous use cases.

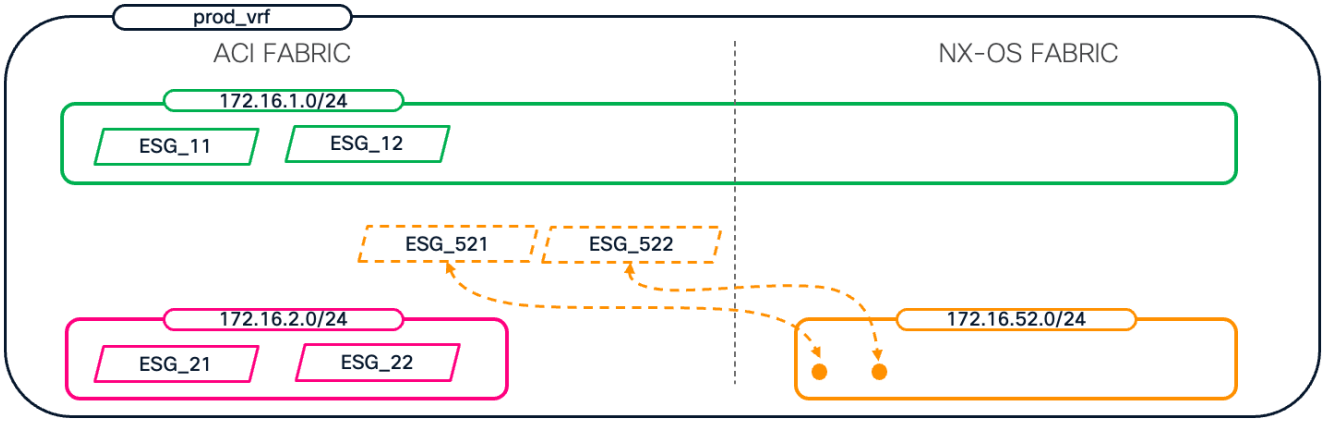

Figure 34 Forwarding and Security Example 1

The security diagram omits contract associations to streamline understanding and focus on classification. The workflow for attaching contracts between different ESGs remains consistent with other use cases and follows the same process as in a classic ACI-only deployment.

Focus should be placed on the different ESGs drawn on the diagram, as each subnet might contain at least two of them.

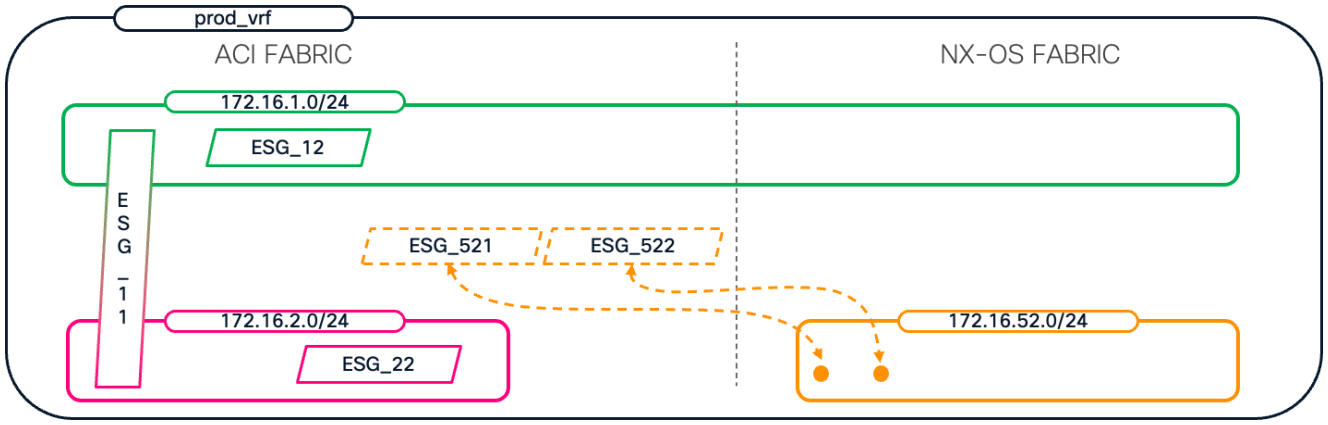

Figure 35 Forwarding and Security Example 2

A slightly different scenario is depicted in the figure above, as in this case ESG_11 is used to classify endpoints part of different BDs. There will not be any difference in this case, and the scenario is totally supported.

- Prerequisites:

Required VRFs and Bridge Domains are stretched across fabrics.

Important: Administrators must ensure that Type-2 routes for network 172.16.52.0/24 are advertised from NX-OS BGWs to classify the /32 IP addresses within ESG_521 and ESG_522. This would not be possible if only a Type-5 route for the 172.16.52.0/24 subnet was advertised from the NX-OS BGWs and requires administrators to locally provision on the NX-OS BGWs that specific L2VNI.

Endpoint Security Groups (ESGs) are defined and actively used for ACI endpoint classification.

Due to subnet stretching, Type-2 EVPN routes for MAC-only and MAC/IP addresses part of the subnet 172.16.1.0/24 are advertised by NX-OS BGWs to ACI BGWs.

- **Notes**:

This use case represents the most granular segmentation scenario, where endpoints within the same stretched subnet are assigned to different security groups, requiring endpoint-level differentiation.

Both bridged and routed traffic must be granularly classified and enforced to maintain strict micro-segmentation.

The complexity of classification is significantly higher than in previous use cases, necessitating careful planning and management. On the policy enforcement side, filters, contracts and their association to the ESGs does not introduce additional complexity.

- **Solution**:

Use MAC Tag Mapping to associate individual MAC addresses with specific ESGs, enabling precise micro-segmentation of intra and inter site bridged traffic within the same stretched subnet. The MAC Tag Mapping should include MAC addresses for both endpoints in the ACI and NX-OS fabrics.

Use IP Tag Mapping or specific IP Subnet Selectors (e.g. /32 prefixes) to classify routed traffic at the endpoint level, ensuring that routed flows are enforced according to endpoint-specific policies. The IP Tag Mapping should include IP addresses for both endpoints in the ACI and NX-OS fabrics.

Do not use the VXLAN BD Selector, as network-level classification is insufficient when endpoints in the same subnet belong to different security groups across remote networks. The same would be true for routed traffic if using an IP selector matching the whole subnet network.

Define contracts and filters to enforce policies between security groups across both stretched and external subnets.

- **Solution Notes**:

When using an EPG selector, ACI attached endpoints MAC and IP are classified only within the local fabric. When the endpoint moves to a NX-OS remote fabric, there won't be classification rule that will place these addresses (received via EVPN updates from the NX-OS BGWs) in the right ESG, and this will cause traffic blackholes. Using MAC and IP tagging ensures that both MAC and IP addresses remain consistently classified independently of the endpoint location, maintaining continuous policy enforcement and preventing security lapses during endpoint mobility. This approach is essential in multi-fabric environments where endpoints frequently move between sites in a host-based group segmentation environment.

Distributed Anycast Gateway Across Domains

For seamless workload mobility (e.g., vMotion) across fabrics, a consistent default gateway must be available in all locations. Nexus One supports a distributed anycast gateway across ACI and NX-OS domains.

For optimal fabric integration and seamless endpoint mobility, the preferred design is to configure the VXLAN NX-OS Distributed Anycast Gateway (DAG) to match the ACI default Bridge Domain virtual MAC address. This approach works perfectly in a scenario where the VXLAN EVPN fabric is newly deployed. ACI, by default, uses the same MAC address on all the BD SVIs and aligning it in the NX-OS side requires a single command:

```
leaf-1103# show run fabric forwarding  | grep anycast-gateway-mac
fabric forwarding anycast-gateway-mac 0022.bdf8.19ff
```

In brownfield scenarios where the NX-OS fabric already has production networks and hence changing the DAG MAC might be difficult, the ACI Bridge Domain can be customized to align with the existing DAG MAC to ensure a transparent transition without requiring endpoints to re-ARP

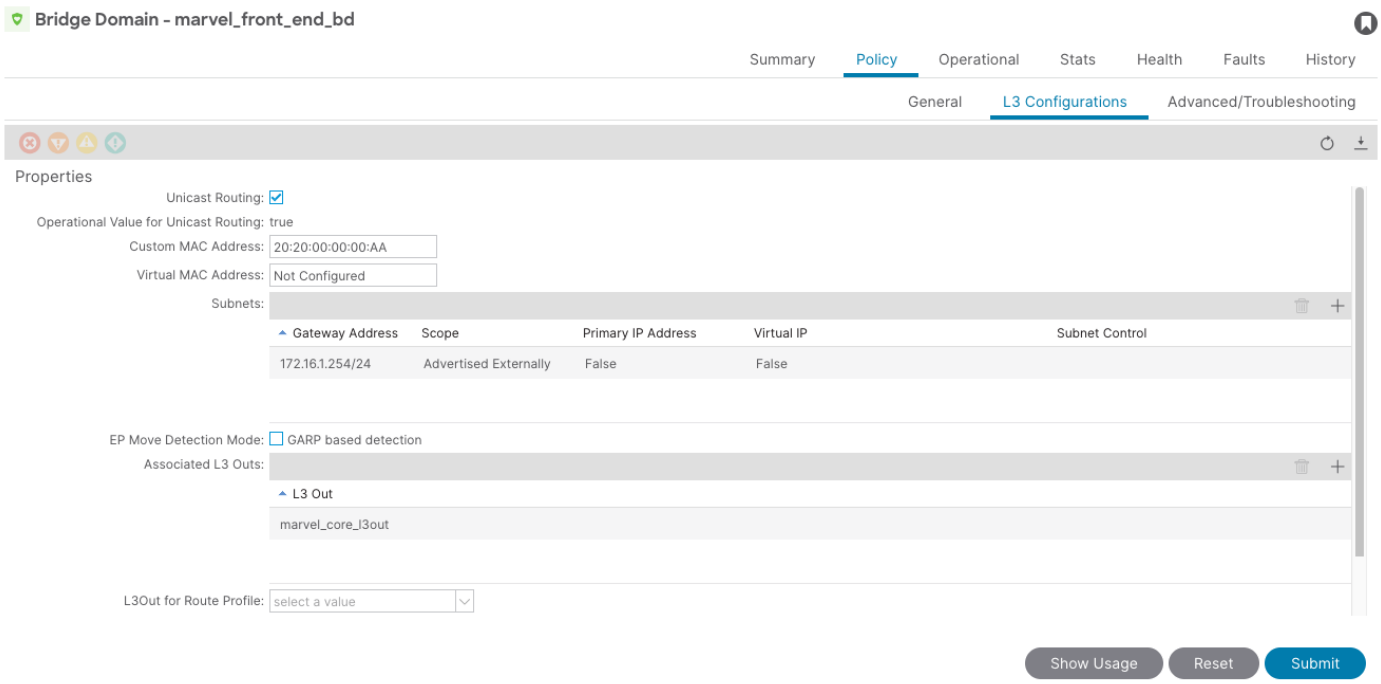

Figure 36 ACI BD Custom MAC address

NX-OS VXLAN EVPN with Security Groups

Modern data centers are the backbone of organizational success, hosting critical applications and services. Consequently, securing these environments, from individual endpoints to vast resources, is paramount. Traditionally, application and workload security were often concentrated at the perimeter, safeguarding the north-south boundary with firewalls and other inspection solutions. However, this approach has proven increasingly ineffective against the sophisticated nature of contemporary cyber threats. The attack surface within a data center extends comprehensively, encompassing both north-south and, critically, east-west traffic flows.

Perimeter-based security solutions face inherent limitations. They are primarily designed for north-south traffic and struggle with scalability when extended to the high-volume, dynamic east-west communications prevalent in modern, virtualized and microservices-based data center fabrics. The cost and complexity associated with scaling traditional security services to inspect every east-west flow become prohibitive, necessitating a more integrated and fabric-native approach to segmentation.

The introduction of Group Policy Option (GPO) in NX-OS is a cornerstone of the Nexus One architecture, as it establishes a common policy framework for VXLAN EVPN fabrics that natively integrates with ACI native segmentation capabilities. This chapter will only provide a quick overview of the GPO feature. Further information, including configurations and verification steps, can be found in the Securing Data Centers with Microsegmentation using VXLAN GPO White Paper, link also available in the references section.

The Emergence of VXLAN Group Policy Option

Segmentation, the practice of dividing a network into smaller, isolated segments to control communication, is not a new concept. Historically, it was achieved through mechanisms like Private VLANs (PVLANs) for intra-broadcast domain isolation (Microsegmentation) and Access Control Lists (ACLs) for inter-broadcast

domain control. However, with the widespread adoption of virtualization, microservices and the evolution of data centers into flexible, scalable VXLAN EVPN fabrics, a more robust and fabric optimized segmentation solution became essential.

This demand led to the introduction of VXLAN Group Policy Option. VXLAN GPO is a powerful, fabric-level solution that addresses the shortcomings of traditional security models by embedding security directly into the network infrastructure. It operates by classifying network resources, both internal and external to the fabric, into logical entities known as Security Groups (SGs). Once endpoints are grouped, specific security policies, named Security Group Access Lists (SGACLs) or Contracts, can be defined and applied between these SGs, meticulously governing communication flows between them.

Core Use Cases of VXLAN GPO

VXLAN GPO can be primarily adopted for two use cases:

- Network Segmentation: VXLAN GPO empowers administrators to define granular policies for network segmentation, enabling the creation of isolated segments within the network. This capability provides superior control over traffic, ensuring that communication between endpoints occurs only when explicitly permitted. Segmentation policies can be unidirectional or bidirectional, offering flexibility in security design. The most implemented security models are:
    - VRF as a Security Zone: While Virtual Routing and Forwarding (VRF) instances provide robust logical isolation between routing domains, GPO enhances this by allowing dynamic segmentation and classification of resources within a VRF. This is particularly valuable for isolating or quarantining specific endpoints (e.g., an infected server) without affecting other resources in the same VRF, thereby preventing lateral movement of threats.
    - Subnet as a Security Zone: In modern data centers, enforcing security policies between different subnets within the same VRF is a common requirement. Traditional methods often involve hair-pinning traffic through external firewalls, leading to performance bottlenecks and scalability issues. GPO offers a more efficient alternative by allowing each subnet to be assigned to its own security group. This enables stateless security policies, functionally like ACLs, to be enforced natively and efficiently on network devices, significantly improving scalability and network performance while maintaining logical isolation.
    - Application as a Security Zone: GPO provides a powerful framework for implementing security and segmentation policies once application components and their interactions are identified. By grouping related components of an application into distinct security groups, organizations can achieve logical isolation between different applications, prevent unauthorized interactions, and manage access to shared infrastructure resources (like backup servers or DNS). It also facilitates user-based access controls, enhancing overall security and compliance.
- Service Chaining (Traffic Redirection): Beyond segmentation, VXLAN GPO simplifies the insertion of network services into traffic flows. It can be used to redirect north-south or east-west traffic to specific network functions, such as firewalls, load balancers, or intrusion detection systems, based on defined policy criteria. This capability ensures that traffic requiring advanced security inspection or load balancing is automatically redirected through the appropriate service functions, optimizing resource utilization and enhancing overall network resilience while keeping simplicity as paramount.

Fundamental Constructs of VXLAN GPO

To enable its powerful capabilities, VXLAN GPO introduces several key constructs:

- Security Group (SG): An SG is a logical container representing a collection of network endpoints or external prefixes that share a common security posture or traffic profile. Endpoints within the same SG can communicate freely, while communication between different SGs is strictly governed by defined contracts.
- Security Group Tag (SGT): Each Security Group is assigned a unique 16-bit SGT, which is essentially the Group Policy ID. These SGTs are fundamental to defining security policies; the SGT of the source is referred to as S-SGT, and that of the destination as D-SGT. SGTs are fabric-wide constructs and must be unique. A range of SGTs (0-15) is reserved for specific purposes.
- Traffic Filter: This construct provides granular control over policy enforcement by allowing users to define the specific types of traffic that should be subject to a security policy between two SGs. A traffic filter can specify criteria such as IP Protocol, L3-L4 protocols, and L4 source/destination ports.
- Security Group Access List (SGACL): SGACLs, or security contracts, are the core of the security policy. They bind source and destination Security Groups with traffic selectors and define the enforcement action (permit, deny, or redirect). They can be configured as bidirectional (enforcing traffic in both directions) or unidirectional (enforcing traffic in only one direction).

VRF Enforcement Modes

VXLAN GPO enables policy enforcement at the VRF level, ensuring full separation between security policy definition and network connectivity provisioning. A newly defined Tenant VRF starts in an Unenforced mode, meaning no security policies are applied. To activate policy enforcement, the VRF mode must be changed to Enforced, with two primary options:

- Default-Permit: In this mode, all unicast traffic flows between security groups are permitted by default, unless explicitly denied by a contract. This model is suitable for environments where most communication is trusted, and specific flows need to be blocked.
- Default-Deny: This mode embodies a "Zero Trust" security model, where all unicast traffic flows are denied by default. Communication is only allowed if explicitly permitted by a contract. This is the most secure approach, requiring explicit authorization for all inter-SG communication.

When transitioning a brownfield environment to GPO, a common best practice is to start with **default-permit** mode, define policies with logging enabled for verification, and then switch to **default-deny** mode once all necessary permit contracts are validated. This minimizes disruption and downtime.

Operational Principles: Control and Data Plane

The effectiveness of VXLAN GPO relies on seamless integration between control and data planes:

- Control Plane Operation: Group Policy ID (SGT) information is propagated across the fabric via MP-BGP EVPN routes. SGTs are carried as an Extended Community attribute in BGP Update messages, specifically with EVPN Type-2 (MAC/IP) and Type-5 (IP Prefix) routes. This propagation allows VTEPs to learn the SGTs of remote endpoints, enabling early policy enforcement at the ingress VTEP.
- Data Plane Operation: The VXLAN header itself is extended to carry the SGT of the source endpoint. This is crucial for scenarios where the ingress VTEP may not know the destination SGT, allowing the egress VTEP to perform policy enforcement. Additionally, a "Policy Applied (PA)" bit within the VXLAN header signals whether the policy has already been enforced by the ingress VTEP. If the PA bit is set, the egress VTEP simply forwards the packet; if not, it applies the policy.

- SGT Derivation and Policy Lookup: NX-OS offers flexible and evolving policy lookup mechanisms. In the initial implementation, policy enforcement relied on an IP-based lookup for both bridged and routed traffic, necessitating that VTEPs learn the IP addresses of connected endpoints to determine source and destination SGTs. However, a new option introduced in NX-OS 10.6(1)F is the capability to enforce Layer 2 traffic policies based on MAC addresses. This MAC-based segmentation allows Security Group Access Control Lists (SGACLs) to be applied to bridged frames by classifying traffic using source and destination MAC addresses. This is a crucial distinction, as it enables policy enforcement even when routing is not configured on the VLAN, or for non-IP L2 traffic. This mode supports VLAN micro-segmentation without requiring an SVI configuration, unlike IP-based segmentation, making it particularly useful in deployments where anycast gateways (such as a firewall) act as the default gateway outside the VXLAN fabric. This MAC-based segmentation is mutually exclusive with IP-based segmentation and is enabled via a system-level CLI command, representing a fabric-wide operation that is traffic disruptive but does not require a device reload. While this enhancement is optional within homogeneous NX-OS fabrics, it should be considered mandatory when extending the policy plane between ACI and NX-OS domains to ensure consistent functionality for bridged traffic.

In its native NX-OS implementation VXLAN GPO provides flexible policy enforcement capabilities across different data center topologies:

- East-West Enforcement (Single Site): Within a single VXLAN fabric, policy enforcement can occur at either the ingress or egress VTEP. Ingress enforcement is preferred for efficiency, requiring both source and destination SGTs to be known by the ingress VTEP. If the destination SGT is unknown, the ingress VTEP signals this (by setting the PA bit to 0), and the egress VTEP performs the enforcement.
- East-West Enforcement (Multi-Site): GPO extends its capabilities across multiple fabrics within a Multi-Site Fabric-Group.
    - Policy-Aware to Policy-Aware: In a Multi-Site deployment where all sites are GPO-enabled, SGT information is exchanged between Border Gateways (BGWs) and propagated to local site VTEPs. Policy is typically enforced at the ingress VTEP of the originating site.
    - Policy-Unaware to Policy-Aware: VXLAN GPO supports coexistence with non-GPO-enabled sites. In such cases, policy-aware BGWs automatically assign a special SGT 15 (default policy-unaware tag) to EVPN routes received from policy-unaware sites. Contracts are then defined in the policy-aware sites to manage communication with SGT 15. Enforcement occurs on the ingress leaf of the policy-aware site (for traffic originating from the policy-aware site) or on the BGWs of the policy-aware site (for traffic originating from the policy-unaware site).

  A key best practice for Multi-Site deployments is to ensure all defined L2VNIs are consistently configured on the BGWs of all fabrics. This guarantees the propagation of detailed MAC/IP endpoint information and associated SGTs via EVPN Type-2 advertisements, enabling optimal policy enforcement on ingress leaf nodes.
- North-South Policy Enforcement: GPO facilitates securing traffic flows between the VXLAN fabric and external networks. The Border Leaf device, which performs the VXLAN-to-IP handoff, classifies and assigns SGTs to external prefixes. These SGTs are then propagated within the VXLAN fabric via EVPN Type-5 updates. Policy enforcement for north-south traffic occurs at the ingress VTEP (for traffic from VXLAN to external) or at the Border Leaf (for traffic from external to VXLAN).

VXLAN GPO represents a significant advancement in data center security, enabling organizations to build more secure, resilient, and agile network infrastructures. Furthermore, the introduction of GPO enhances

the Nexus One architecture by allowing a common policy domain to be extended between ACI and NX-OS environments. This capability significantly enriches the use cases described in the ACI BGW Policy Enforcement section of this document.

The Unified Nexus Dashboard

In a multi-fabric data center environment, the key challenges are not only traffic forwarding and enforcement, but consistent operations across fabrics. As ACI and NX-OS-based domains must function as part of a single architectural model, a unified approach to management, and day-2 operations becomes essential.

Beginning with release 4.2, the Unified Nexus Dashboard assumes this strategic role within the Cisco Nexus One architecture. It is positioned as the central intelligence and operational hub for the multi-fabric domain, providing a common management plane for fabrics that would otherwise be administered as separate environments. In this role, Nexus Dashboard is intended to serve as both the controller and the central source of truth for the domain, unifying configuration, orchestration, policy deployment, and operational visibility across interconnected fabrics.

A foundational principle of the Unified Nexus Dashboard is abstraction. Rather than forcing operators to work directly with the native constructs of each individual fabric, the platform provides a fabric-agnostic GUI and API model through which common logical objects can be defined once and then translated consistently across the environment. Constructs such as VRFs, networks, and security groups, along with their associated contracts, can therefore be expressed centrally and deployed across both ACI and NX-OS fabrics. This abstraction is critical to simplifying operations and establishing a more uniform architectural model across the data center.

Automation is equally important. Multi-fabric connectivity depends on a series of tightly coordinated configuration tasks that are often complex and error-prone when performed manually. These include Border Gateway onboarding, ISN underlay peering, overlay EVPN session establishment, and namespace normalization mapping. By automating these workflows, the Unified Nexus Dashboard is designed to reduce operational overhead, minimize human error, and accelerate the deployment of both connectivity and security services across fabrics.

Another key function of the platform is centralized policy management. In a distributed data center environment, policy consistency is not simply an operational benefit; it is a requirement for security, compliance, and predictable service delivery. The Unified Nexus Dashboard provides a common platform for defining, deploying, and auditing network and security policies that span multiple fabrics. Centralizing these functions helps ensure that policy intent is applied consistently, remains auditable, and can be managed as part of one cohesive operational model rather than as a collection of isolated configurations.

Beyond provisioning and policy deployment, the Unified Nexus Dashboard is intended to evolve into the day-2 operations center for Nexus One. A unified and consistent operational view across all connected sites and fabrics is essential in environments where health, performance, and troubleshooting must be understood end to end. To address this need, the platform is designed to provide consolidated health scores, anomalies, and key performance indicators across the entire domain. This single-pane operational view gives administrators a more complete understanding of both infrastructure health and security posture, while helping them identify issues that may emerge across the entire heterogeneous deployment.

These architectural principles extend across the full lifecycle of data center operations. Fabric onboarding is simplified through guided workflows that bring new ACI or NX-OS fabrics into the Nexus One domain and integrate them into the unified management framework. Network and VRF extension capabilities are intended to enable Layer 2 and Layer 3 services to span selected fabrics, supporting consistent addressing, routing, and access control for applications distributed across sites. This becomes particularly relevant for Multi-Site deployment models and disaster recovery use cases.

End-to-end visibility is another important differentiator. In a multi-fabric environment, understanding how traffic moves between endpoints and which policies affect it at each stage is critical for both troubleshooting and policy validation. The Unified Nexus Dashboard is expected to extend the existing Connectivity Analysis capability so that a traffic flow can be traced from an endpoint in an ACI fabric, through the border gateways, to an endpoint in an NX-OS fabric. By exposing the network and security policies applied at each hop, this capability can significantly improve operational assurance, accelerate root-cause analysis, and support compliance validation in complex environments.

In this way, the Unified Nexus Dashboard is more than a management interface. Within the Nexus One architecture, it is being developed as the operational layer that brings together abstraction, automation, policy consistency, and end-to-end visibility across the multi-fabric domain. Its role is to simplify how modern data center networks are built, extended, secured, and operated, while giving administrators a single and coherent point of control over an otherwise diverse and distributed infrastructure.

One important note about the role of Nexus Dashboard in the overall architecture, is the fact that users have a choice in selecting it or not. Unlike the classic ACI Multi-Site architecture where the Orchestrator was mandatory, with Nexus One, teams will be allowed to choose between different operational models depending on their design choices and preferences. If centralized lifecycle management via Nexus Dashboard is chosen, administrators will still be able to deploy the configurations independently in each fabric via the classic UI or via multiple automation options including the well-known and powerful Cisco Network as Code.

End-to-End Connectivity and Policy Enforcement

The previous chapters examined the main components that enable Nexus One as a solution by analyzing them independently. This chapter provides a holistic explanation that encompasses the control, data and security planes in a scenario where both ACI and NX-OS fabrics are policy-aware, enabling the reader to understand the complete end-to-end picture.

For consistent and optimized forwarding and enforcement, the following rules must be followed:

All devices in the NX-OS fabric should be configured with GPO MAC based segmentation, a feature available starting with NX-OS 10.6(1)F. To learn more about this consult the NX-OS VXLAN Configuration Guide, specifically the “Micro-segmentation for VXLAN fabrics using group policy option (GPO)” chapter.

Due to the above requirement, when traffic needs to be bridged, it is mandatory to ensure that Type-2 MAC-only routes will carry the required SGT. This can be achieved when using the following selectors to classify resources in both ACI and NX-OS domains:

- ACI:
  - EPG selector
  - MAC Tag policy and selector
- NX-OS:

- VLAN classifier
- MAC address classifier
- (Port, VLAN) classifier

L2VNIs for stretched VRFs should be defined on the BGW nodes even when networks are not stretched and only routed connectivity is required. Although this may seem counterintuitive, it is essential to ensure that no packets are dropped. If L2VNIs are not attached to the BGWs, type-2 routes will not be advertised, preventing the remote fabric from receiving any EVPN route that maps the host IP to its Security Group Tag (SGT). In this case, the remote fabric will only receive the type-5 route associated with the L2VNI default gateway, which is not linked to a useful SGT (in the current implementation, Type-5 prefixes carry the SGT 1 value). Attaching the L2VNI to the local BGW guarantees that type-2 routes are propagated to remote sites, preventing this issue.

This concept is further represented in the diagrams below where a stretched VRF (prod_vrf) contains networks that are only locally deployed in each site.

The first diagram shows the initial state, where BD1 and BD2 L2VNIs, both part of the stretched prod_vrf, are locally defined only on the aci-leaf and nxos-leaf, respectively. None of the bridge domains (BDs) are configured on the border gateways (BGWs). As a result, Type-2 EVPN routes are not advertised by the BGWs to the remote fabrics unlike Type-5 routes representing the L2VNIs subnets. The administrator has defined a contract between security-group 11111 and security-group 22222 that allows all the traffic. The behavior described next would not change if a different action, like deny or redirect, was set.

To simplify this discussion, we are abstracting the distinction between ACI local and normalized pcTags. For the purposes of this section, we treat 11111 and 22222 as normalized pcTags throughout the ACI fabric. Please note that in a production environment, the ACI fabric utilizes internal pcTags that are transparently translated by the BGW in both control and data plane; we are omitting this translation layer here to focus on the end-to-end flow.

Note: Type-5 routes will be announced from the ACI BGWs only if the BD subnet has the “Advertised Externally” flag set. In contrast, NX-OS BGWs will announce Type-5 routes as soon as the VRF is configured on the BGW.

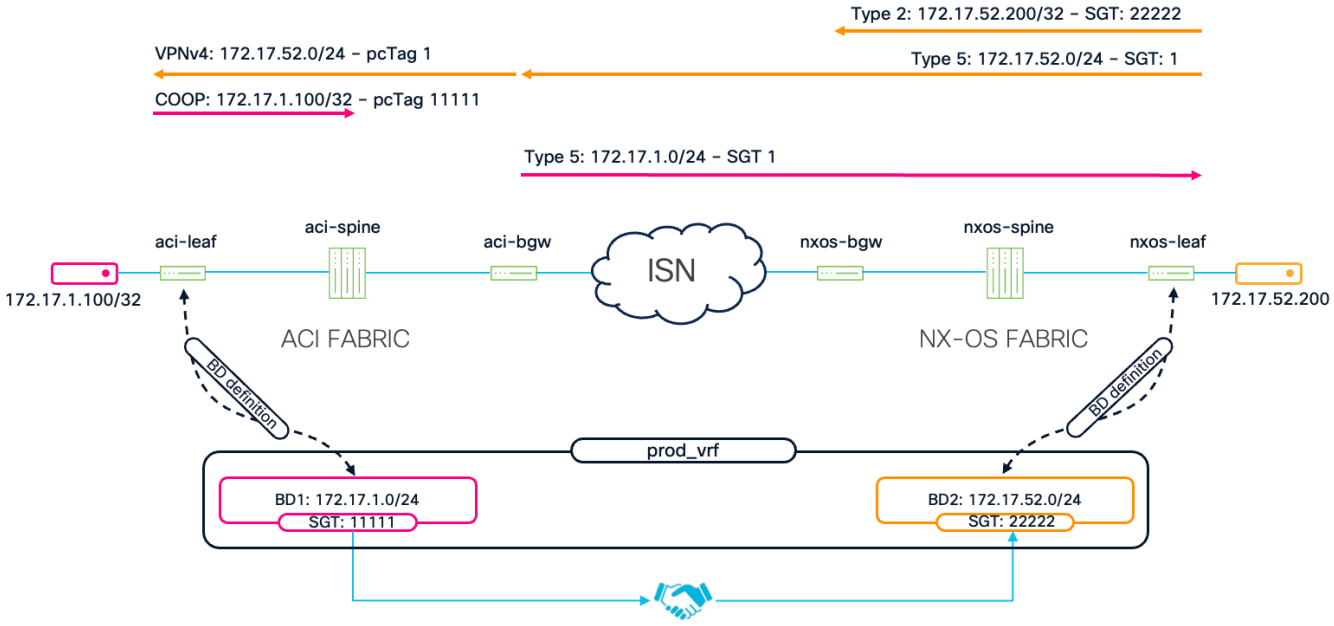

Figure 37 Control Plane advertisements when BDs only attached to the local leaf

While this configuration works for routed connectivity scenarios, it does not meet security requirements, hence it should not be considered a valid solution. Primarily, it lacks the correct classification needed to determine the Source SGT and Destination SGT of the traffic.

The above control plane scenario would translate into the following data-plane packet walk.

Notice how the packet is allowed all the way from the original NX-OS ingress leaf to the ACI egress leaf. This can occur even if there is no contract permitting the traffic, if a contract explicitly denies the traffic, or if there is a contract that allows communication between the two security groups but with filters that do not match the actual payload. This happens because the ingress NX-OS leaf applies a contract between source SGT 22222 and destination SGT 1 and NX-OS implicitly permits traffic with destination SGT 1.

The same will happen when the packet reaches the NX-OS BGW. The packet is then forwarded to the ACI BGW without GPO information, so the ACI BGW must re-lookup the source of the traffic and map it to the right pcTag. Since the ACI BGW only stores the EVPN Type-5 route for the entire BD2 subnet, the mapping derived is pcTag 1. The ACI BGW allows the packet after applying the policy, as the source pcTag and destination pcTag are 1, and sets the PA bit to 1.

Finally, the packet is encapsulated again in an iVXLAN tunnel and sent to the ACI leaf, via the Spine-Proxy. As the policy was already applied in the previous hop, the packet will be allowed.

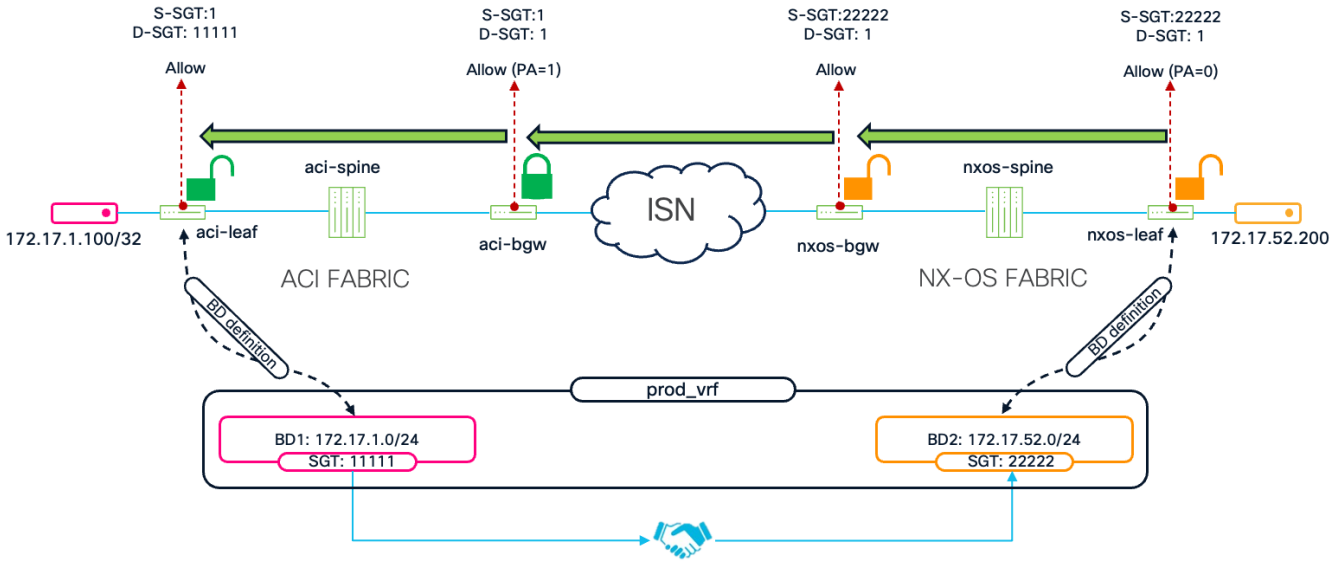

Figure 38 NX-OS to ACI traffic w/o Type-2 routes

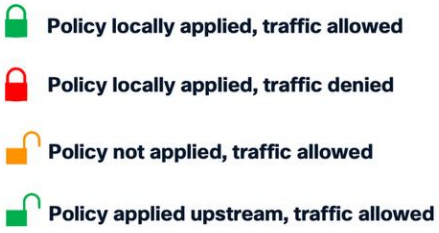

Figure 39 Policy enforcement legend

However, the return packet meets a different fate, and the outcome still fails to match the intended behavior.

In this reverse scenario, the traffic flow originates from the ACI side and is destined for the NX-OS side. The process begins at the ACI leaf, which only knows the /24 subnet and identifies the packet's destination pcTag as 1. Because ACI's design permits all traffic to and from pcTag 1, the leaf does not enforce any specific contract. Instead, it simply encapsulates the packet into an iVXLAN tunnel and forwards it to the ACI BGW.

Following the same logic, the ACI BGW also recognizes the destination as pcTag 1. Consequently, it stitches the packet for forwarding to the remote fabric without applying any security policy.

However, the packet is dropped at the NX-OS BGW. Upon arrival, the BGW inspects the packet and, after a re-lookup, identifies its source as SGT 1 and its destination as SGT 22222. Unlike ACI, NX-OS policy explicitly blocks traffic originating from SGT 1, and as a result, the packet is dropped.

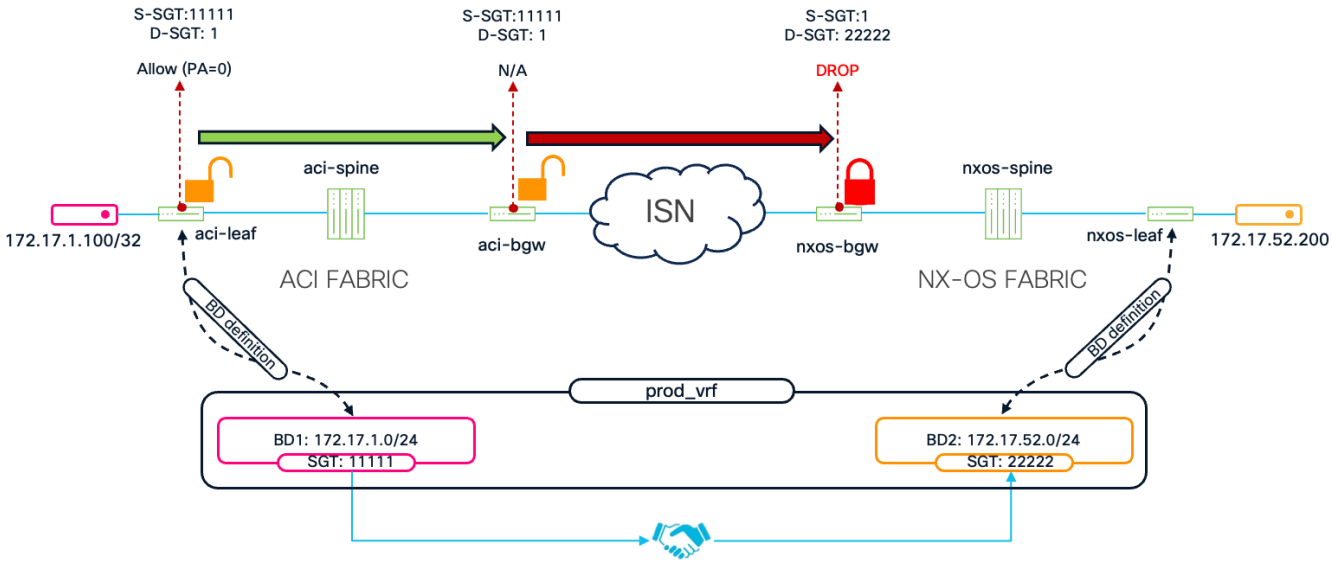

Figure 40 ACI to NX-OS traffic w/o Type-2 route

As demonstrated in the above sections, missing the BD/L2VNI configuration on the BGWs leads to a non-deterministic behavior that would not reflect the intended policy enforcement. The next section will demonstrate the best practice configurations.

The solution to these communication failures lies in ensuring the BD is also configured on the local BGWs. This configuration change is the critical step that triggers the advertisement of EVPN Type-2 routes from the local BGW to the remote fabric's BGWs. Unlike Type-5 routes, which represent the entire subnet, Type-2 routes carry specific host information, including the MAC and IP addresses and their associated security groups.

Note: Configuring the BD on the BGWs of a single fabric doesn't stretch the broadcast domain across the other ones. For this to happen the same BD should also be configured on the remote BGWs.

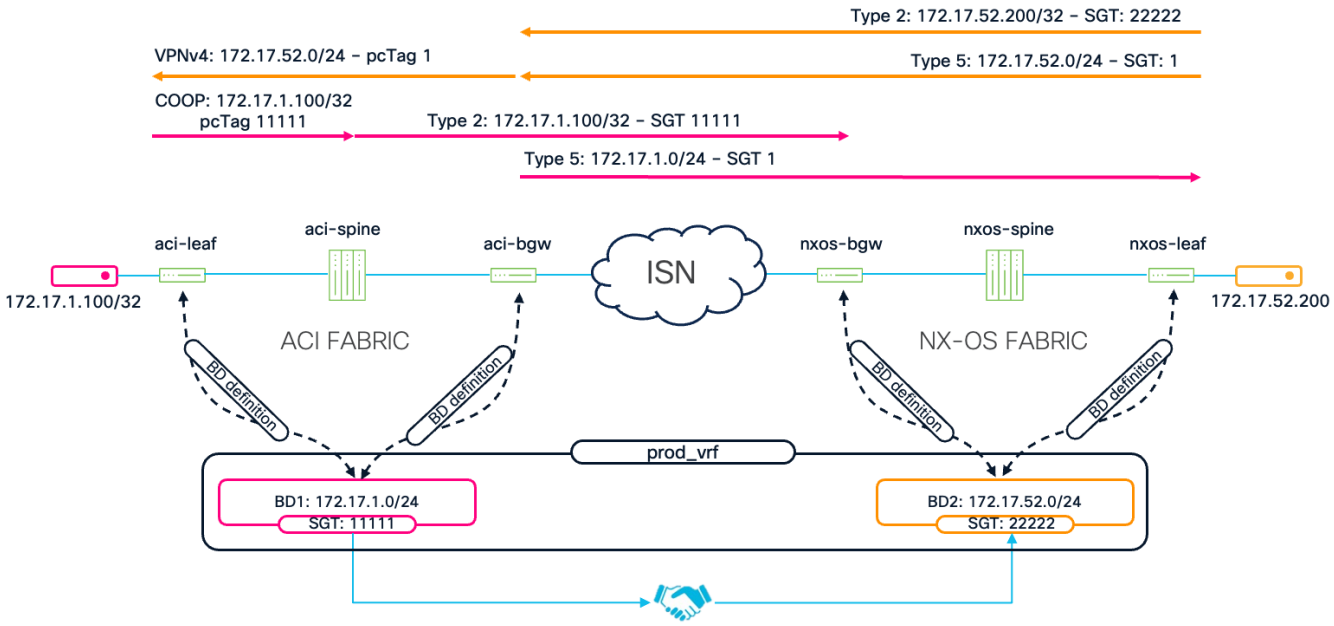

Figure 41 Control Plane advertisements when BDs attached to the local leaf and BGW

Upon receiving type-2 routes, the remote BGWs install the /32 (or /128 in case of IPv6) host routes into their forwarding tables. The MAC addresses are not installed as the BD is not configured on the remote BGWs. This enables correct classification. When a packet arrives from the remote fabric, the BGW can now perform a precise lookup based on the source IP address, mapping it directly to its correct, learned SGT instead of defaulting to the SGT 1 associated with the entire BD. As a result, security policies can be correctly applied, either by the BGW or the egress leaf, based on true source and destination SGTs.

The above configuration, with both networks stretched on the local BGWs translates in the below packet walks.

The assumption is that the traffic is allowed by the policy. If the policy does not allow it, the the packet would be dropped by the nxos-bgw for NX-OS to ACI flows and by the aci-bgw for ACI to NX-OS flows.

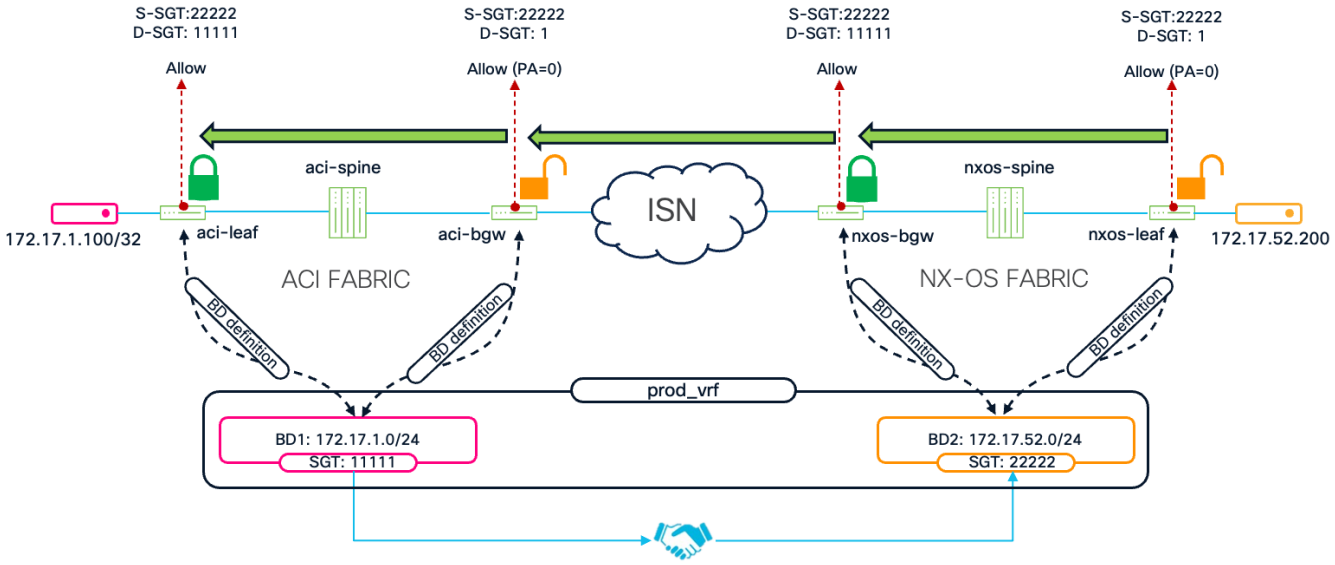

Figure 42 NX-OS to ACI traffic with Type-2 routes

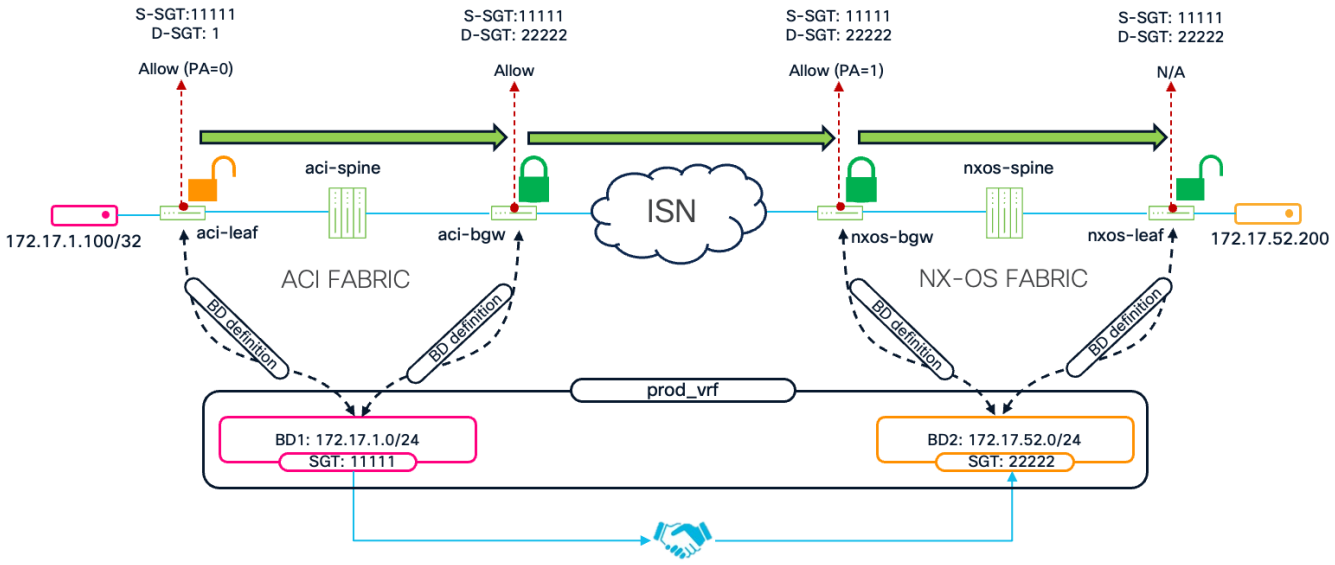

Figure 43 ACI to NX-OS traffic with Type-2 routes

While the current configuration ensures policy enforcement at the NX-OS BGW for flows destined to the remote ACI domain, a further optimization can shift this enforcement closer to the source, at the ingress NX-OS leaf itself. As it stands, the ingress NX-OS leaves are not yet able to apply the policy locally. This is because the local NX-OS BGWs do not propagate the newly learned remote Type-2 routes downstream into their own fabric, as the corresponding ACI BD is not configured on them.

To enable this more efficient, localized enforcement, an administrator can configure the remote ACI BD on the NX-OS BGWs. For instance, BD1 would be configured on the nxos-bgw. This action ensures that the Type-2 routes are propagated downstream to the NX-OS leaves, creating a pervasive endpoint-to-SGT mapping across the entire NX-OS fabric. With this complete mapping available locally, the ingress NX-OS leaf can enforce the security policy immediately, without needing to forward the packet to the BGW first.

The drawback of this implementation is that it requires allocating one additional L2VNI and VLAN on the VXLAN EVPN Border Gateways even if there aren't endpoints attached to that subnet in the fabric.

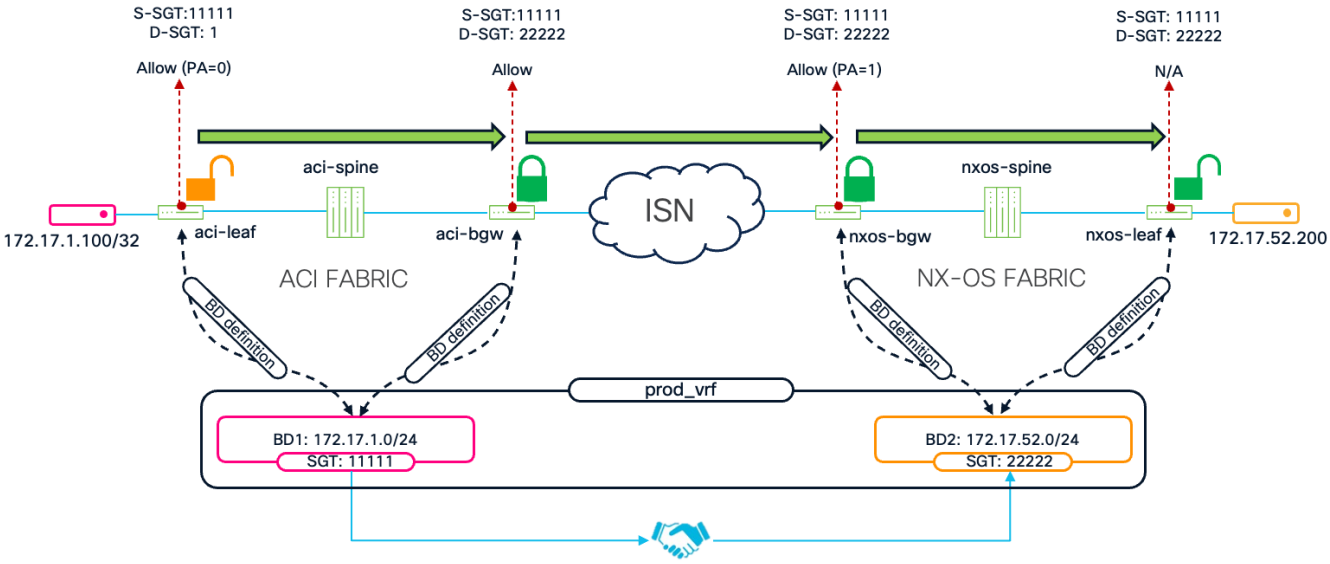

Figure 44 Control Plane advertisements when BDs attached to the local leaf, BGW and remote (NX-OS) BGW

The above configuration will allow the NX-OS leaves to apply the policy as soon as the packet arrives at the ingress leaf.

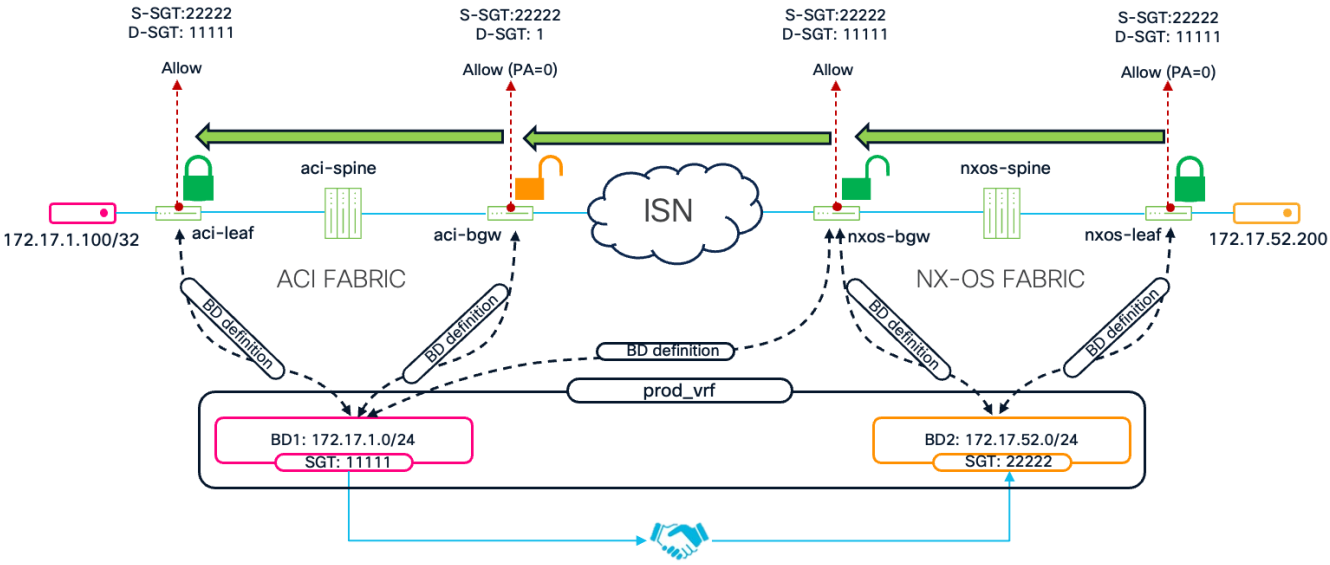

Figure 45 Optimized ingress enforcement happening at the NX-OS leaf

With this correct configuration in place, the scenarios we covered are now handled as intended, and security policies are properly enforced.

Note: It is important to highlight that this last optimization is currently only possible for NX-OS fabric, which implements remote endpoint learning via the control plane. The same behavior is not achievable in ACI

environments (as of release 6.1(5)M), because ACI leaves do not install remote endpoints learned from the BGW control-plane advertisements into their endpoint tables. Due to this, traffic sourced in an ACI fabric targeting a remote NX-OS network won't see any improvements.

Control Plane Reachability Advertisement – Step by Step

Understanding how reachability information for hosts and external networks is propagated across Nexus One architecture when both domains are policy-aware is fundamental. Having covered the foundational concepts, we will now examine this process in action.

We will use a simple topology to trace connected endpoints and external routes advertisements from NX-OS environments to ACI and vice-versa. The following diagram represents the location and details about them.

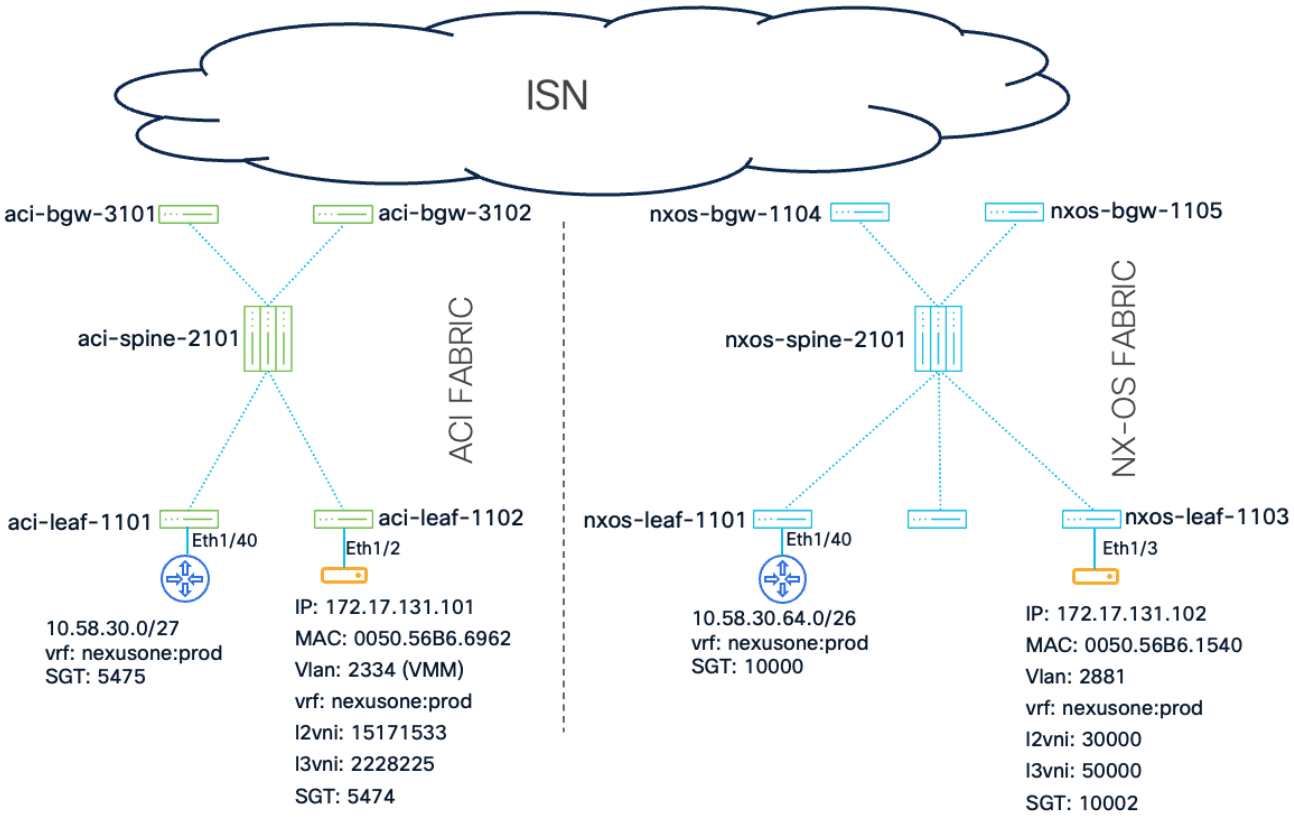

Figure 46 Global endpoints and external route diagram

As explained in a previous section, each VTEP participating in the control plane uses unique or shared loopbacks, Figure 47 Overlay Control Plane is used again to represent that from a general point of view, whereas Figure 48 Topology with infrastructure IPs, provides the key infrastructure IP addresses.

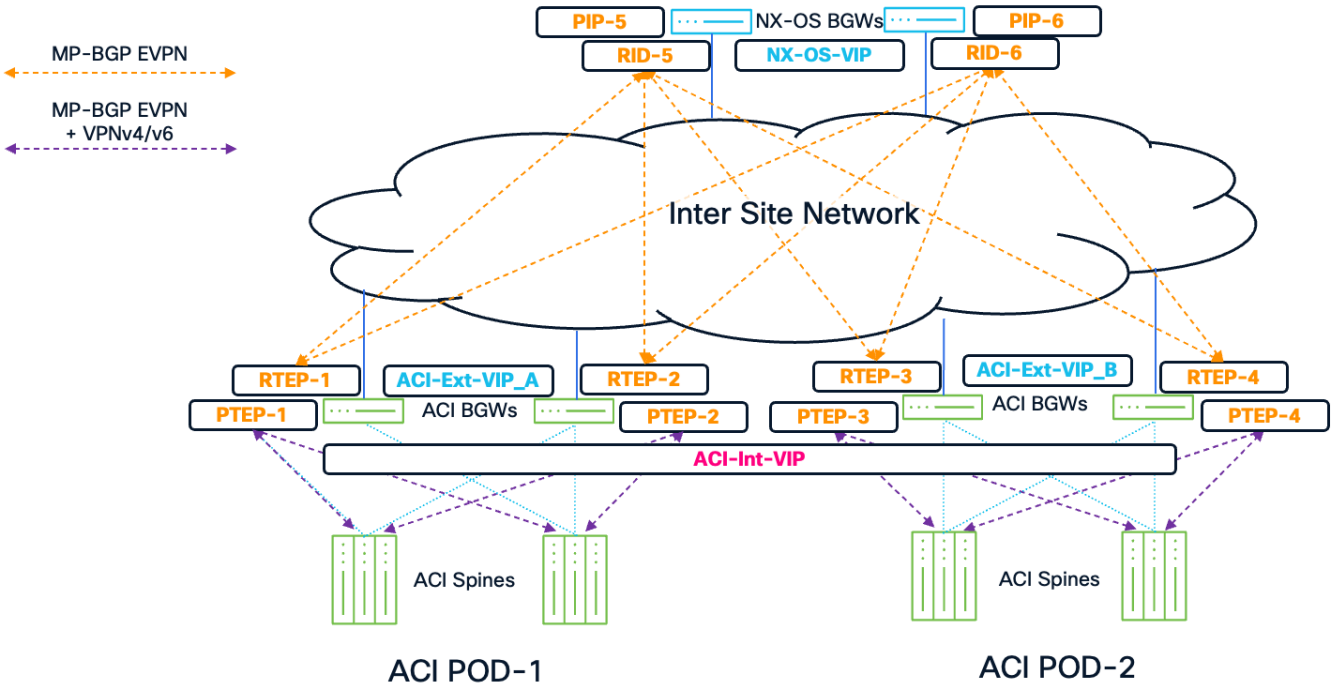

Figure 47 Overlay Control Plane

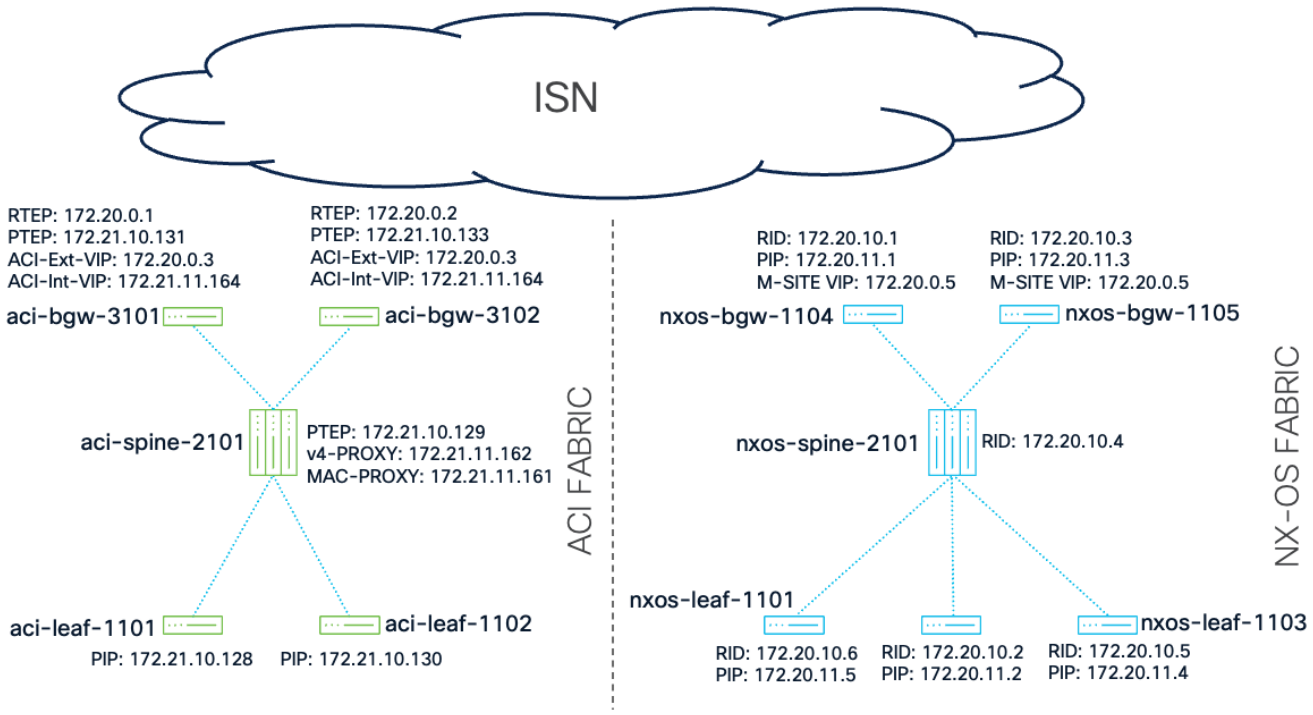

Figure 48 Topology with infrastructure IPs

Note: You might have noticed that for the NX-OS devices we are using the traditional names to represent the loopbacks:

- RID: Loopback used to establish the overlay (BGP) adjacencies. It is also used as router-id for the IGP and other functions
- PIP: Primary IP, configured on a different loopback it acts as source or destination for the VXLAN packets
- M-SITE VIP: Shared between all NX-OS Border Gateways acts as destination for the majority of VXLAN packets going across sites

Note: Additional fabrics are used into this lab but are not part of the testbed; you might see their BGWs IPs (172.20.20.1, 172.20.20.5, 172.20.30.3) appearing from time to time in the outputs.

The control plane journey for endpoints and external networks learned in the NX-OS fabric

The following endpoint has been connected and discovered by a leaf node in the NX-OS fabric:

IP Address	MAC Address	VRF	L3VNI	VLAN	L2VNI	SGT
172.17.131.102	0050.56B6.1540	nexusone:prod	50000	2881	30000	10002

The external route is also learned:

IP Address	Next-Hop	VRF	L3VNI	SGT
10.58.30.64/26	172.20.14.22	nexusone:prod	50000	10000

Both VRF and Network (L2VNI) have been locally provisioned on the NX-OS BGWs, consequently the following EVPN routes will be advertised from these devices to the ACI fabric:

- MAC-only Type-2
- MAC/IP Type-2
- DAG SVI Type-5
- External Route Type-5
- Inclusive Ethernet Multicast Tag (IMET) Type-3 for BUM Ingress Replication

Let's start our verifications from the NX-OS leaves that discover our end-host and the external network.

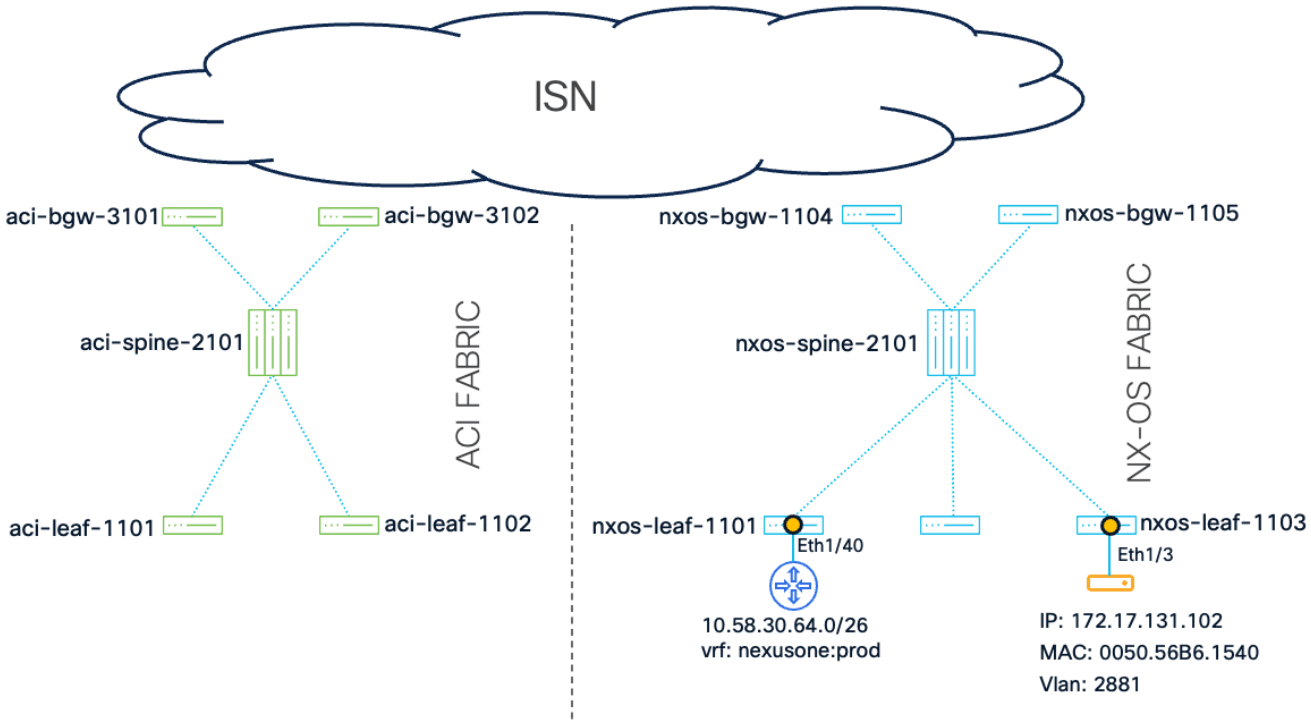

Figure 49 Control Plane verifications on nxos-leaf-1101 and nxos-leaf-1103

First, the leaves learn the necessary reachability information. nxos-leaf-1103 discovers the local host's MAC address (0050.56b6.1540) through data-plane activity and its corresponding IP address (172.17.131.102/32) via ARP. nxos-leaf-1101 installs a route for the external subnet 10.58.30.64/26 with a next-hop of 172.20.14.22 (representing the external device that advertises the prefix).

While learning the reachability information, the VTEPs also classify the host and external route in their security groups. SGT 10002 is correctly mapped to both the MAC and IP addresses of the local endpoint, as GPO classification is configured on a per-VLAN basis. SGT 10000 instead is assigned to the external subnet using an external-subnets IP classifier.

```
nxos-leaf-1103# show mac address-table esg address 0050.56b6.1540
   VNI      MAC Address      Type      age     Secure NTFY Ports     ESG Tag
---------+-----------------+--------+---------+------+----+------------------
* 2881     0050.56b6.1540   dynamic  NA         F      F    Eth1/3 10002

nxos-leaf-1103# show ip route 172.17.131.102 detail vrf nexusone:prod
172.17.131.102/32, ubest/mbest: 1/0, attached
    *via 172.17.131.102, Vlan2881, [190/0], 5d18h, hmm
         xri tag type: Security Group; value: 10002

nxos-leaf-1101# show ip route 10.58.30.64/26 detail vrf nexusone:prod
10.58.30.64/26, ubest/mbest: 1/0
```

```
    *via 172.20.14.22, [20/44], 00:17:08, bgp-65101, external, tag 65100
         client-specific data: 1c
         recursive next hop: 172.20.14.22/32
         xri tag type: Security Group; value: 10000

nxos-leaf-1103# show security-group id 10002
Security Group ID 10002 , Name nexusone:front-end
 Selector Type : Vlan
    VLAN-Id                         MAC-Address                     Interface
    Vlan2881                        *                               *

nxos-leaf-1101# show security-group id 10000
Security Group ID 10000 , Name nexusone:external_subnet
  Selector Type : External IPv4 Subnets
    VRF-Name                        IPv4-Address/mask-len
    nexusone:prod                   10.58.30.64/26
```

Note: An IP classifier would prevent the endpoint MAC address from being mapped to the SGT, consequently bridged communications would be broken, this is true assuming that MAC-based segmentation is enabled in the VXLAN EVPN fabric (as per best practice recommendation in Nexus One deployments).

These addresses are then advertised in EVPN, nxos-leaf-1103 generates Type-2 MAC-only and MAC/IP NLRIs, plus a Type-5 route as the SVI is locally deployed:

```
nxos-leaf-1103# show bgp l2vpn evpn 0050.56b6.1540
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:35648    (L2VNI 30000)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 113
Paths: (1 available, best #1)
Flags: (0x000102) (high32 00000000) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop
  AS-Path: NONE, path locally originated
    172.20.11.4 (metric 0) from 0.0.0.0 (172.20.10.5)
      Origin IGP, MED not set, localpref 100, weight 32768
      Received label 30000
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:0:10002 MAC Mobility Sequence:00:1

  Path-id 1 advertised to peers:
    172.20.10.4
```

```
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 116
Paths: (1 available, best #1)
Flags: (0x000302) (high32 00000000) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop
  AS-Path: NONE, path locally originated
    172.20.11.4 (metric 0) from 0.0.0.0 (172.20.10.5)
      Origin IGP, MED not set, localpref 100, weight 32768
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:0:10002
          MAC Mobility Sequence:00:1 Router MAC:ecc0.182a.7173

  Path-id 1 advertised to peers:
    172.20.10.4

nxos-leaf-1103# show bgp l2vpn evpn 172.17.131.0
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:4    (L3VNI 50000)
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 214
Paths: (2 available, best #2)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn
  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: NONE, path locally originated
    172.20.11.4 (metric 0) from 0.0.0.0 (172.20.10.5)
      Origin incomplete, MED 0, localpref 100, weight 32768
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:1 Router MAC:ecc0.182a.7173

  Path-id 1 advertised to peers:
    172.20.10.4
```

While nxos-leaf-1101 generates a Type-5 route for the external network

```
leaf-1101# show bgp l2vpn evpn 10.58.30.64
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.6:5    (L3VNI 50000)
BGP routing table entry for [5]:[0]:[0]:[26]:[10.58.30.64]/224, version 2694
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn
```

```
  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65100 , path sourced external to AS
    172.20.11.5 (metric 0) from 0.0.0.0 (172.20.10.6)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:10000 Router MAC:4874.1001.a47b
          OSPF RT:0.0.0.0:0:0

  Path-id 1 advertised to peers:
    172.20.10.4
```

Next, we must ensure that the spine is receiving the right data from the NX-OS leaves.

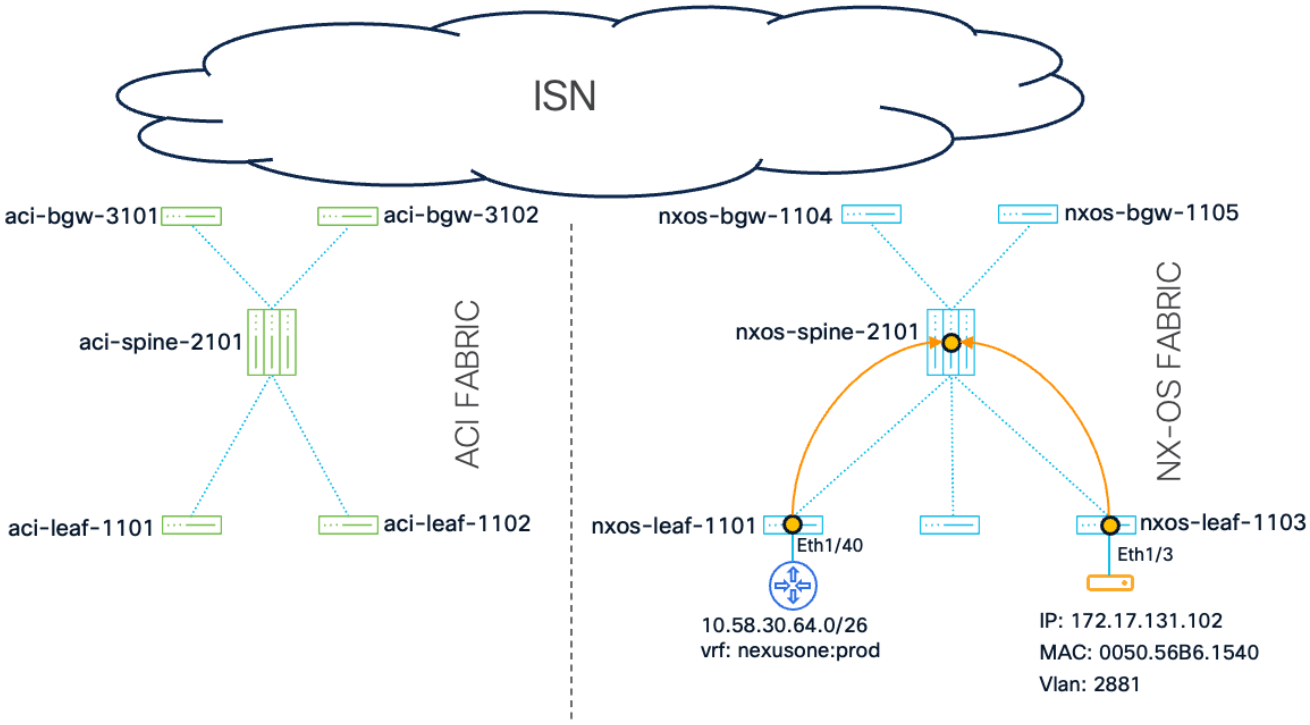

Figure 50 Control Plane verifications on NX-OS spine

The MAC-only EVPN Type-2 route is received by the spine. This route which carries the right SGT is then reflected to all other VTEPs within the fabric. Crucially, this propagation includes the border gateways (BGWs) at **172.20.10.1** and **172.20.10.3**, ensuring they have the necessary Layer 2 and security information.

```
nxos-spine-2101# show bgp l2vpn evpn 0050.56b6.1540
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:35648
```

```
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 115
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 2) from 172.20.10.5 (172.20.10.5)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:0:10002 MAC Mobility Sequence:00:1

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.2        172.20.10.3        172.20.10.6
```

The same is true for the MAC/IP Type-2 route.

```
nxos-spine-2101# show bgp l2vpn evpn 172.17.131.102
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:35648
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 112
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 2) from 172.20.10.5 (172.20.10.5)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:0:10002
          MAC Mobility Sequence:00:1 Router MAC:ecc0.182a.7173

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.2        172.20.10.3        172.20.10.6
```

The Type-5 associated with the DAG SVI is also received by the spine node; in this case notice that the SGT equals to 1 as expected.

```
nxos-spine-2101# show bgp l2vpn evpn 172.17.131.0
BGP routing table information for VRF default, address family L2VPN EVPN
```

```
Route Distinguisher: 172.20.10.5:4
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 183
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 2) from 172.20.10.5 (172.20.10.5)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:1 Router MAC:ecc0.182a.7173

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.2        172.20.10.3        172.20.10.6
```

Finally, the IP prefix of the external network is also received as EVPN Type-5 on the spine, with the associated pcTag value 10000.

```
nxos-spine-2101# show bgp l2vpn evpn 10.58.30.64
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.6:4
BGP routing table entry for [5]:[0]:[0]:[26]:[10.58.30.64]/224, version 121
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65100 , path sourced external to AS
    172.20.11.5 (metric 2) from 172.20.10.6 (172.20.10.6)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:10000 Router MAC:4874.1001.a47b
          OSPF RT:0.0.0.0:0:0

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.2        172.20.10.3        172.20.10.5
```

In this EVPN fabric, the spines function as route reflectors. Their role is to simply reflect the routes without altering them, meaning there is no modification to validate on the spines themselves. Consequently, the next logical step in our verification path is to examine the routes as they arrive on one of the NX-OS BGWs. For simplicity we will focus on one of them, nxos-bgw-1104:

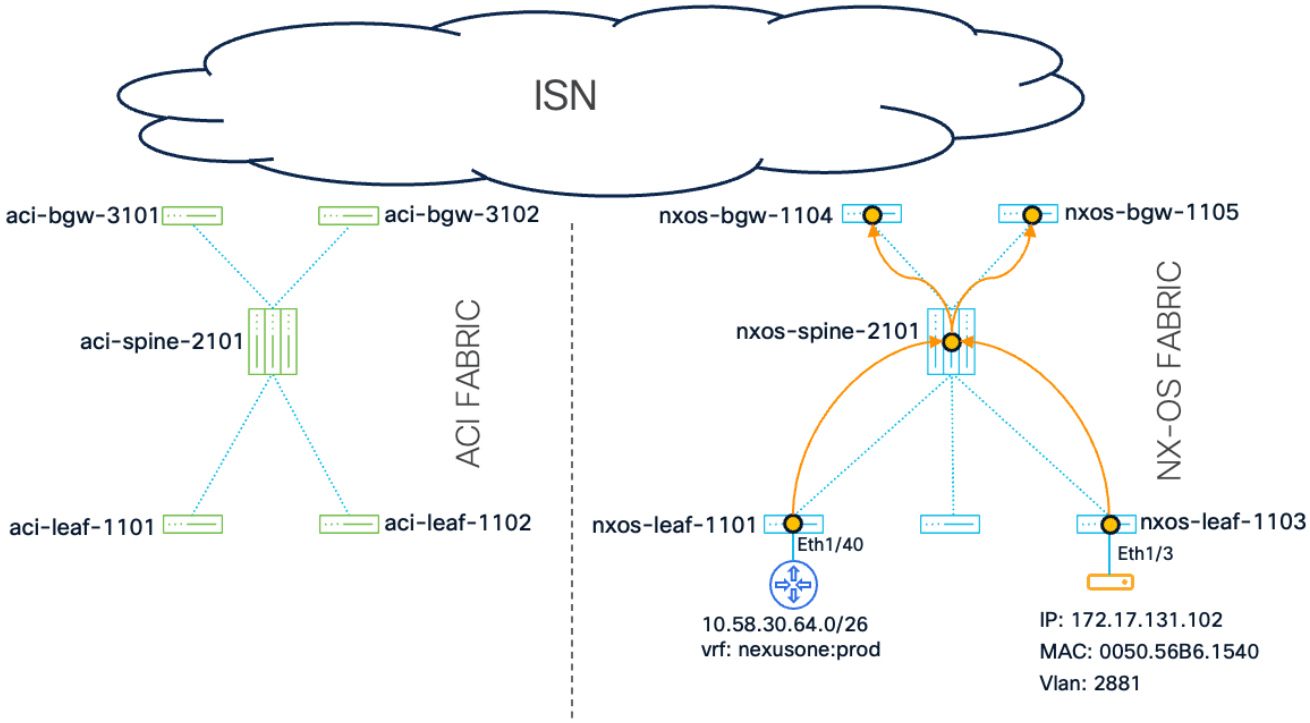

Figure 51 Control Plane verifications on NX-OS Border Gateway

When handling EVPN routes, like the MAC-Only Type-2 route that we will inspect next, the NX-OS BGW effectively processes two distinct versions of the route, both of which are visible in the snippet below. First, it receives the route from the internal fabric spines. This route is never sent to remote fabrics:

```
nxos-bgw-1104# show bgp l2vpn evpn 0050.56b6.1540
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:35648
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 212
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported to 1 destination(s)
             Imported paths list: L2-30000
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin IGP, MED not set, localpref 100, weight 0
```

```
      Received label 30000
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:0:10002 MAC Mobility Sequence:00:1
      Originator: 172.20.10.5 Cluster list: 172.20.10.4

  Path-id 1 not advertised to any peer
```

Second, it re-originates a new version of this route and advertises it toward the remote ACI BGWs (172.20.0.1 and 172.20.0.2).
Crucially, this re-originated route preserves the SGT information, ensuring the security classification is correctly propagated to the ACI fabric.

Note: We use the same command as before but, in both cases, we extract the partial and interesting output only

```
nxos-bgw-1104# show bgp l2vpn evpn 0050.56b6.1540
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.1:35648    (L2VNI 30000)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 213
Paths: (1 available, best #1)
Flags: (0x000212) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop, in rib
             Imported from
172.20.10.5:35648:[2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:0:10002 MAC Mobility Sequence:00:1
      Originator: 172.20.10.5 Cluster list: 172.20.10.4

  Path-id 1 (dual) advertised to peers:
    172.20.0.1          172.20.0.2          172.20.20.1         172.20.20.5
    172.20.30.3
```

The same pattern of receiving a route from the internal fabric and re-originating it toward the external fabric also applies to the other key EVPN routes.

The following output tracks the re-origination of the Type-2 MAC/IP route. Following the same logic as the MAC-only route, the BGW is expected to re-originate this route toward the ACI fabric.

```
nxos-bgw-1104# show bgp l2vpn evpn 172.17.131.102
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.5:35648
```

```
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 209
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported to 3 destination(s)
             Imported paths list: nexusone:prod L2-30000 L3-50000
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:0:10002
          MAC Mobility Sequence:00:1 Router MAC:ecc0.182a.7173
      Originator: 172.20.10.5 Cluster list: 172.20.10.4

  Path-id 1 not advertised to any peer

Route Distinguisher: 172.20.10.1:35648    (L2VNI 30000)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 219
Paths: (1 available, best #1)
Flags: (0x000212) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop, in rib
             Imported from
172.20.10.5:35648:[2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:0:10002
          MAC Mobility Sequence:00:1 Router MAC:ecc0.182a.7173
      Originator: 172.20.10.5 Cluster list: 172.20.10.4

  Path-id 1 (dual) advertised to peers:
    172.20.0.1          172.20.0.2          172.20.20.1         172.20.20.5
    172.20.30.3
```

Next, we examine the SVI Type-5 route. In addition to the locally originated route, the BGW also receives one from the remote ACI fabric, which is expected behavior in a stretched subnet configuration. This remote route is easily identified by its distinct origin AS number, in this case AS 65011.

A closer inspection of this ACI-originated route reveals a critical detail in its pcTag, which is set to 0:80:1. With the current implementation ACI BGWs cannot process VXLAN headers that contain GPO information and hence they efficiently use EVPN to notify this preference to the NX-OS BGWs. The middle value,:80:, in fact acts as a signal from the ACI BGWs.

```
nxos-bgw-1104# show bgp l2vpn evpn 172.17.131.0
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 16365:19005441
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 106
Paths: (2 available, best #2)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Path type: external, path is valid, not best reason: newer EBGP path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:80:1 Router MAC:c02c.17ff.b946

  Path-id 1 not advertised to any peer

Route Distinguisher: 172.20.10.5:4
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 273
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported to 1 destination(s)
             Imported paths list: nexusone:prod
  Gateway IP: 0.0.0.0
  AS-Path: NONE, path sourced internal to AS
    172.20.11.4 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:1 Router MAC:ecc0.182a.7173
      Originator: 172.20.10.5 Cluster list: 172.20.10.4

  Path-id 1 not advertised to any peer
```

```
Route Distinguisher: 172.20.10.1:4    (L3VNI 50000)
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 278
Paths: (1 available, best #1)
Flags: (0x000002) (high32 0x000400) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, reoriginated, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: NONE, path locally originated
    172.20.11.1 (metric 0) from 0.0.0.0 (172.20.10.1)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:1 Site-ID:0:65101
          Router MAC:4874.1038.2a4b
      Originator: 172.20.10.2 Cluster list: 172.20.10.4

  Path-id 1 (dual) advertised to peers:
    172.20.0.1          172.20.0.2          172.20.20.1         172.20.20.5
    172.20.30.3
```

And the External Type-5 route, which carries reachability information for external networks. Following the established pattern, the BGW re-originates this route toward the ACI fabric. During this process, the route is correctly tagged with the appropriate SGT. This crucial step ensures that the fabric's security policy is consistently applied not only to internal endpoints but also to traffic destined for external networks.

```
nxos-bgw-1104# show bgp l2vpn evpn 10.58.30.64
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.6:4
BGP routing table entry for [5]:[0]:[0]:[26]:[10.58.30.64]/224, version 227
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: internal, path is valid, is best path, no labeled nexthop
             Imported to 1 destination(s)
             Imported paths list: nexusone:prod
  Gateway IP: 0.0.0.0
  AS-Path: 65100 , path sourced external to AS
    172.20.11.5 (metric 3) from 172.20.10.4 (172.20.10.4)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:10000 Router MAC:4874.1001.a47b
```

```
          OSPF RT:0.0.0.0:0:0
      Originator: 172.20.10.6 Cluster list: 172.20.10.4

  Path-id 1 not advertised to any peer

Route Distinguisher: 172.20.10.1:4    (L3VNI 50000)
BGP routing table entry for [5]:[0]:[0]:[26]:[10.58.30.64]/224, version 251
Paths: (1 available, best #1)
Flags: (0x000002) (high32 0x000400) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, reoriginated, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65100 , path sourced external to AS
    172.20.11.1 (metric 0) from 0.0.0.0 (172.20.10.1)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:10000 Site-ID:0:65101
          Router MAC:4874.1038.2a4b OSPF RT:0.0.0.0:0:0
      Originator: 172.20.10.6 Cluster list: 172.20.10.4

  Path-id 1 (dual) advertised to peers:
    172.20.0.1          172.20.0.2          172.20.20.1         172.20.20.5
    172.20.30.3
```

Finally, we observe the appearance of the EVPN Type-3 route for the specific L2VNI 30000. This route is advertised to the ACI BGWs (172.20.0.1 and 172.20.0.2) to signal that BUM traffic for this L2VNI should be replicated to the NX-OS BGWs. The path IP **[172.20.11.1]/88** is the local BGW PIP. SGT is set to 1 but there is not policy enforcement for BUM traffic.

```
nxos-bgw-1104# show bgp l2vpn evpn vni-id 30000 route-type 3
BGP routing table entry for [3]:[0]:[32]:[172.20.11.1]/88, version 39
Paths: (1 available, best #1)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, no labeled nexthop, is extd
  AS-Path: NONE, path locally originated
    172.20.11.1 (metric 0) from 0.0.0.0 (172.20.10.1)
      Origin IGP, MED not set, localpref 100, weight 32768
      Origin flag 0x2
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:0:1
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
```

```
        Label: 30000, Tunnel Id: 172.20.11.1

  Path-id 1 advertised to peers:
    172.20.0.1          172.20.0.2          172.20.20.1         172.20.20.5
    172.20.30.3
```

Interestingly, the command output also reveals the corresponding Type-3 routes being received from both ACI BGWs. This is a clear indication that the same L2VNI is configured on the ACI side, and those devices are likewise signaling their intent to receive BUM traffic for that L2VNI. To differentiate between the locally originated (NX-OS) and the remotely received (ACI) routes, one can simply inspect the AS-Path.

```
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 172.20.10.1:35648    (L2VNI 30000)
BGP routing table entry for [3]:[0]:[32]:[172.20.0.1]/88, version 96
Paths: (1 available, best #1)
Flags: (0x000012) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported from 172.20.0.1:1:[3]:[0]:[32]:[172.20.0.1]/88
  AS-Path: 65011 , path sourced external to AS
    172.20.0.1 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin IGP, MED not set, localpref 100, weight 0
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:80:1
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 30000, Tunnel Id: 172.20.0.1

  Path-id 1 (dual) not advertised to any peer
BGP routing table entry for [3]:[0]:[32]:[172.20.0.2]/88, version 94
Paths: (1 available, best #1)
Flags: (0x000012) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported from 172.20.0.2:9:[3]:[0]:[32]:[172.20.0.2]/88
  AS-Path: 65011 , path sourced external to AS
    172.20.0.2 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin IGP, MED not set, localpref 100, weight 0
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:80:1
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
```

```
        Label: 30000, Tunnel Id: 172.20.0.2

  Path-id 1 (dual) not advertised to any peer
```

At this point, the EVPN routes will be advertised to the ACI BGWs:

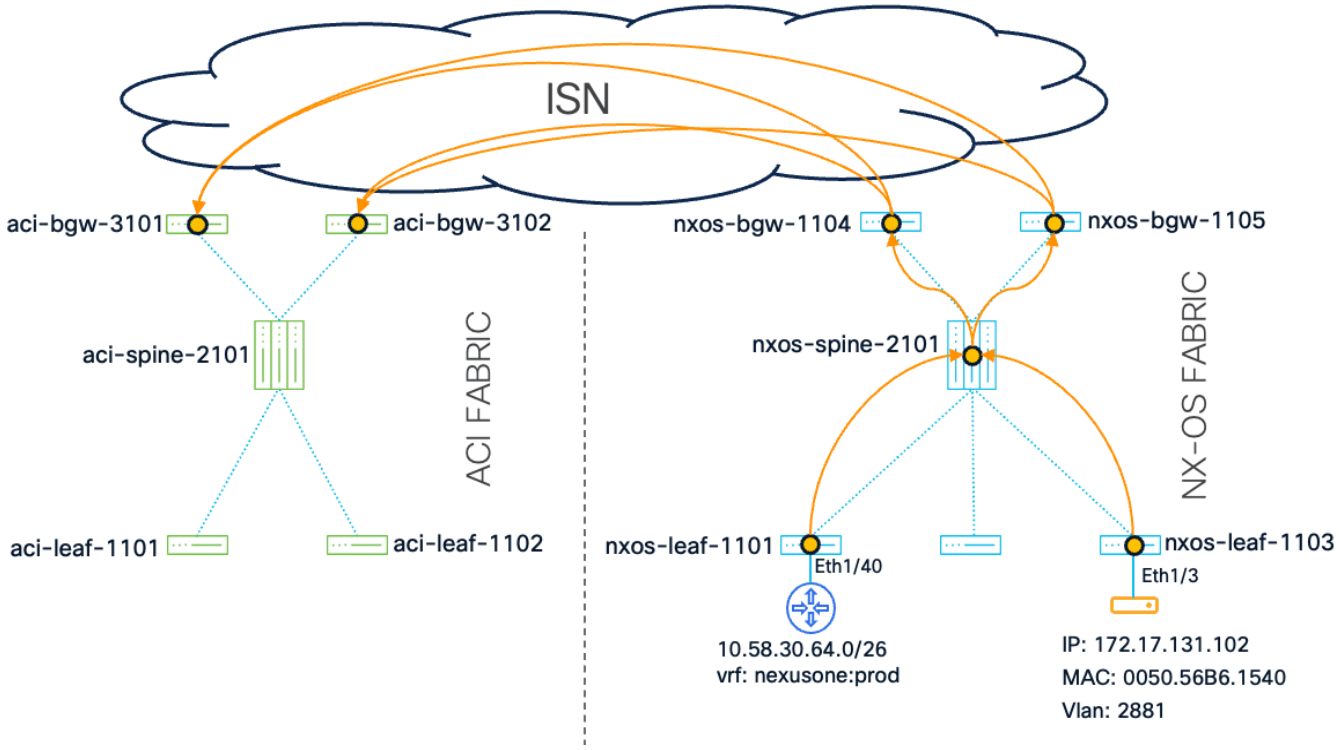

Figure 52 Control Plane verifications on ACI Border Gateway

To confirm that the EVPN routes we have been tracking were successfully received and processed by the ACI BGWs, we can now examine their EVPN tables.

Note: In addition to processing and re-originating specific routes, the ACI BGW populates its local forwarding tables with the entries required to bridge and route packets to remote endpoints and networks.

ACI BGWs similarly to the NX-OS BGWs re-originate the EVPN routes. We will now split the output of the same command into a couple of different blocks. In the first section we will analyze the routes received from the NX-OS BGWs. We can recognize them by the DUAL route distinguisher **65101:30000** which carries the AS:L2VNI_ID format. A common RD allows the devices to store a higher number of routes and to scale higher.

As expected, we are receiving the same route from both NX-OS BGWs. At this stage nothing has been manipulated or normalized except for the route-target (RT) autonomous system portion. This is a direct result of the rewrite-evpn-rt-asn configuration, which the APIC applies implicitly. It is crucial to note that without this rewrite function, the ACI BGW would not import the route at all. In this case the RT has been

converted from the original **65101:**30000 to **65011:**30000. Notice how the VNI and pcTag still show the NX-OS original values

```
aci-bgw-3101# show bgp l2vpn evpn 0050.56b6.1540 vrf overlay-1
Route Distinguisher: 65101:30000
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 279
dest ptr 0x7fc657c2f164
Paths: (2 available, best #1)
Flags: (0x0000000000000202 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc64c2e8f70): external 0x40000028 0x4002000 ref 1 adv path ref 1, path is
valid, is best path
             Imported to 1 destination(s)
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating
svi 0, tunnel resolved 0
      Received label 30000
      Extcommunity: RT:65011:30000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:10002
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1

  Path type (0x7fc64c2e9588): external 0x40000028 0x4002000 ref 0 adv path ref 0, path is
valid, not best reason: newer EBGP path
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.3 (172.20.10.3)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating
svi 0, tunnel resolved 0
      Received label 30000
      Extcommunity: RT:65011:30000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:10002
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1

  Path-id 1 not advertised to any peer
```

From the previous command output, we will now inspect a different section, and we will focus on the route that has been re-originated. We can recognize it by the RD **16365:31948749**; A critical indicator confirms the route has been processed correctly: the pcTag got normalized to the value known inside the ACI fabric namespace.

```
aci-bgw-3101# show bgp l2vpn evpn 0050.56b6.1540 vrf overlay-1
Route Distinguisher: 16365:31948749    (L2VNI 15171533)
```

```
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/216, version 332
dest ptr 0x7fc657c2dd8c
Paths: (1 available, best #1)
Flags: (0x0000000000000212 0000000000) on xmit-list, is in rib/evpn, is not in HW
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc64c75d374): external 0xc0000028 0x400 ref 0 adv path ref 1, path is valid,
is best path, in rib
             Imported from (0x7fc64c2e8f70)
65101:30000:[2]:[0]:[0]:[48]:[0050.56b6.1540]:[0]:[0.0.0.0]/112
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating
svi 0, tunnel resolved 0
      Received label 30000
      Extcommunity: RT:65011:30000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5474
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1

  Path-id 1 advertised to peers:
    172.21.10.129
```

Furthermore, it is important to remember that as these Network Layer Reachability Information (NLRI) details are re-originated by the BGWs and sent to the ACI spine (172.21.10.129), the VNIs (Label and RT last 16 bits) will be translated to match the ACI fabric's internal values during the export.

We will now verify the other routes, starting from the Type-2 MAC/IP, and highlight the important attributes.
For this specific route there is nothing different from the previous MAC-Only Type-2 route except that we also see the L3VNI id since we have an IP address in the EVPN route.

```
aci-bgw-3101# show bgp l2vpn evpn 172.17.131.102 vrf overlay-1
Route Distinguisher: 16365:31948749    (L2VNI 15171533)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 336 dest ptr 0x7fc657c2dc8e
Paths: (1 available, best #1)
Flags: (0x0000000000000212 0000000000) on xmit-list, is in rib/evpn, is not in HW
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc64c75d22c): external 0xc0000028 0x400 ref 0 adv path ref 1, path is valid,
is best path, remote nh not installed, in rib
             Imported from (0x7fc64c2e8ed4)
65101:30000:[2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/144
```

AS-Path: 65101 , path sourced external to AS

172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)

Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating svi 0, tunnel resolved 0

Received label 30000 50000

Extcommunity: **RT:65011:30000 RT:65011:50000** SOO:65011:4093640685

COST:pre-bestpath:170:1610612736 ENCAP:8 **PCTAG:00:0:0:5474**

OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1

Router MAC:0200.ac14.0005

Path-id 1 advertised to peers:

172.21.10.129

Route Distinguisher: 65101:30000

BGP routing table entry for [2]:[0]:[0]:**[48]:[0050.56b6.1540]:[32]:[172.17.131.102]**/272, version 277 dest ptr 0x7fc657c2f066

Paths: (2 available, best #1)

Flags: (0x0000000000000202 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is locked

Multipath: eBGP iBGP

Advertised path-id 1

Path type (0x7fc64c2e8ed4): external 0x40000028 0x4002000 ref 2 adv path ref 1, path is valid, is best path, remote nh not installed

Imported to 2 destination(s)

AS-Path: 65101 , path sourced external to AS

172.20.0.5 (metric 0) **from 172.20.10.1** (172.20.10.1)

Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating svi 0, tunnel resolved 0

Received label 30000 50000

Extcommunity: **RT:65011:30000 RT:65011:50000** SOO:65011:4093640685

COST:pre-bestpath:170:1610612736 ENCAP:8 **PCTAG:00:0:0:10002**

OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1

Router MAC:0200.ac14.0005

Path type (0x7fc64c2e94ec): external 0x40000028 0x4002000 ref 0 adv path ref 0, path is valid, not best reason: Router Id, remote nh not installed

AS-Path: 65101 , path sourced external to AS

172.20.0.5 (metric 0) **from 172.20.10.3** (172.20.10.3)

Origin IGP, MED 2000, localpref 100, weight 0 tag 4294966000, propagate 0, floating svi 0, tunnel resolved 0

Received label 30000 50000

Extcommunity: **RT:65011:30000 RT:65011:50000** SOO:65011:4093640685

COST:pre-bestpath:170:1610612736 ENCAP:8 **PCTAG:00:0:0:10002**

```
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1
          Router MAC:0200.ac14.0005

  Path-id 1 not advertised to any peer
```

We will look at the Type-3 route next. This route is not re-originated to the spines. ACI native underlay FTAG trees are used to replicate BUM traffic throughout the fabric.

```
aci-bgw-3101# show bgp l2vpn evpn route-type 3 vrf overlay-1 | egrep -A3 -B20 "Label: 30000"
Route Distinguisher: 172.20.10.1:35648
BGP routing table entry for [3]:[0]:[32]:[172.20.11.1]/88, version 228 dest ptr
0x7fc657c2e974
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc657e46f48): external 0x28 0x0 ref 1 adv path ref 1, path is valid, is best
path
              Imported to 1 destination(s)
  AS-Path: 65101 , path sourced external to AS
    172.20.11.1 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Extcommunity: RT:65011:30000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 30000, Tunnel Id: 172.20.11.1
Route Distinguisher: 172.20.10.3:35648
BGP routing table entry for [3]:[0]:[32]:[172.20.11.3]/88, version 216 dest ptr
0x7fc657c2f55c
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc657e475fc): external 0x28 0x0 ref 1 adv path ref 1, path is valid, is best
path
              Imported to 1 destination(s)
  AS-Path: 65101 , path sourced external to AS
```

```
    172.20.11.3 (metric 0) from 172.20.10.3 (172.20.10.3)
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Extcommunity: RT:65011:30000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 30000, Tunnel Id: 172.20.11.3

  Path-id 1 not advertised to any peer
```

A noteworthy behavior can be observed with the SVI Type-5 route. The ACI BGW receives this route from the NX-OS BGWs and simultaneously advertises its own local route back to them, which is expected since the subnet is stretched. However, a key detail is that the ACI BGW does not propagate the received NX-OS route downstream to its own spines.

This is not an error but is the intended design of ACI. The architecture treats routes learned from an external peer, including an EVPN peer like an NX-OS BGW, as external routes. Therefore, instead of reflecting the EVPN route internally, the ACI BGW redistributes the Type-5 route from the EVPN address family into its VPNv4/VPNv6 table for propagation.

```
aci-bgw-3101# show bgp l2vpn evpn 172.17.131.0 vrf overlay-1
Route Distinguisher: 65101:50000
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]:[0.0.0.0]/224, version 387 dest
ptr 0x7fc657c2c0c6
Paths: (2 available, best #2)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Path type (0x7fc64c2e75d8): external 0x40000028 0x4000000 ref 0 adv path ref 0, path is
valid, not best reason: newer EBGP path, remote nh not installed
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.3 (172.20.10.3)
      Origin incomplete, MED 1, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:50000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005

  Advertised path-id 1
  Path type (0x7fc64c2e7674): external 0x40000028 0x4000000 ref 1 adv path ref 1, path is
valid, is best path, remote nh not installed
```

```
            Imported to 1 destination(s)
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin incomplete, MED 1, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:50000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005

  Path-id 1 not advertised to any peer

Route Distinguisher: 16365:19005441    (L3VNI 2228225)
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]:[0.0.0.0]/224, version 287 dest
ptr 0x7fc657c3161a
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7fc64c2ec0cc): local 0x4000008c 0x4400000 ref 0 adv path ref 1, path is
valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.20.0.1)
      Origin incomplete, MED 0, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Received label 2228225
      Extcommunity: RT:65011:2228225 VNID:2228225

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.3        172.20.20.1        172.20.20.5
    172.20.30.3
```

Although this SVI route exists on the ACI BGW, the device installs a static route pointing toward the spine proxy but does not advertise the subnet back to the spines via VPNv4. This is expected behavior, as there is no need to advertise the BD subnet to the spines since those devices already have reachability information for all connected endpoints in the COOP database.

```
aci-bgw-3101# show bgp vpnv4 uni 172.17.131.0/24 vrf overlay-1
BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 3101:2228225    (VRF nexusone:prod)
BGP routing table entry for 172.17.131.0/24, version 95 dest ptr 0x7fc64bf7658c
Paths: (2 available, best #1)
Flags: (0x00000000080c0002 0x00000200) on xmit-list, is not in urib, exported
```

```
  vpn: version 284, (0x0000000000100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type (0x7fc657e4870c): redist 0x408 0x404001 ref 0 adv path ref 2, path is valid, is
best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.20.0.1)
      Origin incomplete, MED 0, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Extcommunity: RT:65011:2228225 VNID:2228225

  VPN AF advertised path-id 2
  Path type (0x7fc64c75bef4): external 0xc0000028 0x0 ref 0 adv path ref 1, path is valid,
not best reason: Weight, remote nh not installed
            Imported from (0x7fc64c2e7674)
65101:50000:[5]:[0]:[0]:[24]:[172.17.131.0]:[0.0.0.0]/120
  AS-Path: 65101 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin incomplete, MED 1, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:2228225 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d VNID:2228225

  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 not advertised to any peer
  Path-id 2 not advertised to any peer
```

In contrast, if the subnet is locally deployed within the NX-OS fabric rather than extended, the BGW classifies it as an external route. In this case, the BGW would propagate the reachability information into the fabric by advertising the route to the spines via VPNv4.

For verification, please refer to the subsequent check for the remote external subnet 10.58.30.64/26. The EVPN table is only reporting the Type-5 route received from the NX-OS BGWs which is identified by the RD **65101:50000** (**AS:L3VNI_ID**). There is no EVPN Type-5 re-originated and sent to the spines.

```
aci-bgw-3101# show bgp l2vpn evpn 10.58.30.64 vrf overlay-1
Route Distinguisher: 65101:50000
```

```
BGP routing table entry for [5]:[0]:[0]:[26]:[10.58.30.64]:[0.0.0.0]/224, version 367 dest
ptr 0x7fc657c2f360
Paths: (2 available, best #2)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Path type (0x7fc64c2e8208): external 0x40000028 0x4000000 ref 0 adv path ref 0, path is
valid, not best reason: newer EBGP path, remote nh not installed
  AS-Path: 65101 65100 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.3 (172.20.10.3)
      Origin incomplete, MED 45, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:50000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:10000
          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005
          OSPF RT:0.0.0.0:0:0

  Advertised path-id 1
  Path type (0x7fc64c2e96c0): external 0x40000028 0x4000000 ref 1 adv path ref 1, path is
valid, is best path, remote nh not installed
              Imported to 1 destination(s)
  AS-Path: 65101 65100 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin incomplete, MED 45, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:50000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:10000
          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005
          OSPF RT:0.0.0.0:0:0

  Path-id 1 not advertised to any peer
```

To advertise this, we use the VPNv4 Address Family.

```
aci-bgw-3101# show bgp vpnv4 unicast 10.58.30.64/26 vrf overlay-1
BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 3301:2228225    (VRF nexusone:prod)
BGP routing table entry for 10.58.30.64/26, version 108 dest ptr 0x7fc64bf75e8c
Paths: (1 available, best #1)
Flags: (0x00000000000c001a 0000000000) on xmit-list, is in urib, is best urib route, is in
HW, exported
```

```
  vpn: version 325, (0x0000000000100002) on xmit-list
Multipath: eBGP iBGP


  Advertised path-id 1, VPN AF advertised path-id 1
  Path type (0x7fc64c75c890): external 0xc0000028 0x400 ref 0 adv path ref 2, path is valid,
is best path, remote nh not installed, in rib
            Imported from (0x7fc64c2e96c0)
65101:50000:[5]:[0]:[0]:[26]:[10.58.30.64]:[0.0.0.0]/120
  AS-Path: 65101 65100 , path sourced external to AS
    172.20.0.5 (metric 0) from 172.20.10.1 (172.20.10.1)
      Origin incomplete, MED 45, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:2228225 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5475
          OSPF RT:03100000:0000fe4d VNID:2228225 OSPF RT:0.0.0.0:0:0

  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    172.21.10.129
```

Since the BGW is now advertising these routes to the spines, the next logical step in our verification path is to examine the ACI spines themselves:

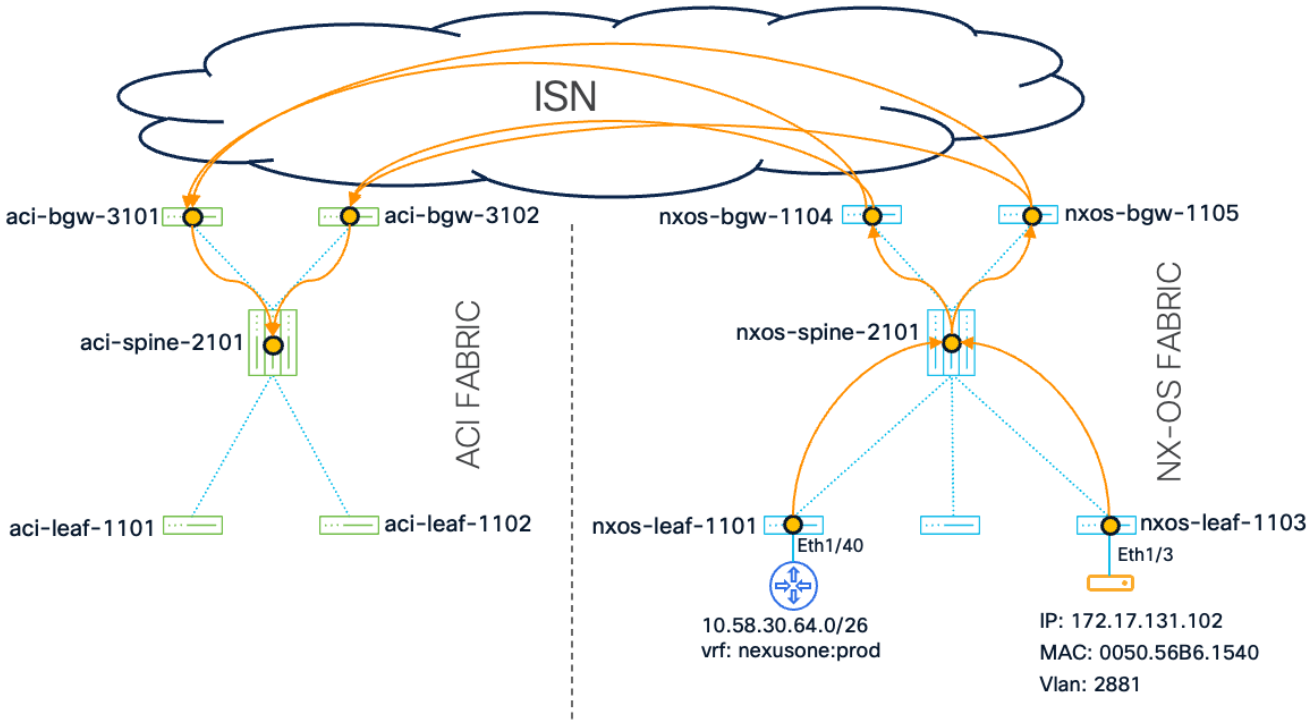

Figure 53 Control Plane verifications on ACI Spines

In the Nexus One control plane scenario the ACI spines serve two primary and distinct functions in the overall flow.

First, facilitate remote endpoint reachability. They receive EVPN Type-2 routes from the ACI BGWs, redistribute this information into the COOP (Council of Oracle Protocol) database, and subsequently install these entries into their hardware forwarding tables. Thanks to this, spines can provide their native Layer-2 and Layer-3 spine proxy functionality for inter fabric conversations too.

Second, they act as VPNv4/VPNv6 route reflectors and propagate to all the ACI leaves the external subnets learned from local L3Outs and the EVPN Type-5 routes received and converted by the ACI BGWs.

The command output below displays both the MAC-Only and MAC/IP Type-2 routes for the remote endpoint. The key process shown here is the redistribution of this information from the EVPN table into the COOP database. During this redistribution, two essential attributes are maintained:

- The tunnel next-hop, which correctly points to the VTEP (172.21.11.164, the fabric **vxlan-evpn-anycast-v4-internal-tep**).
- The pcTag.

It is important to note how the pcTag is represented within COOP. The value appears as a hex-encoded PcTag, such as 0x1001562. This format is because COOP sets an internal flag by performing a logical OR operation with 0x01000000. Therefore, to find the original SGT, simply disregard the leading 1 and convert the remaining hex value (0x1562) to decimal, which yields 5474.

```
aci-spine-2101# show bgp l2vpn evpn 172.17.131.102 vrf overlay-1
Route Distinguisher: 16365:31948749
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/272,
version 282 dest ptr 0x7f64b6a0dfe8
Paths: (2 available, best #1)
Flags: (0x0000000000000202 0x00000800) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7f64ac447ab8): internal 0x40000018 0x4002000 ref 1 adv path ref 1, path is
valid, is best path, remote nh not installed
              Imported to 1 destination(s)
  AS-Path: 65101 , path sourced external to AS
    172.21.11.164 (metric 2) from 172.21.10.131 (172.20.0.1)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:65011:15171533 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5474
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1
          Router MAC:0200.ac14.0005

  Path type (0x7f64ac447c8c): internal 0x40000018 0x4002000 ref 0 adv path ref 0, path is
valid, not best reason: Router Id, remote nh not installed
  AS-Path: 65101 , path sourced external to AS
    172.21.11.164 (metric 2) from 172.21.10.133 (172.20.0.2)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:65011:15171533 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5474
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1
          Router MAC:0200.ac14.0005

  Path-id 1 not advertised to any peer

Local Route Distinguisher: 1:16777199    (L2VNI 1)
Paths: (2 available, best #2)
Flags: (0x000000000000031a 0x00000009) on xmit-list, is in rib/evpn, is not in HW, is in
l2rib mpod shard, is in l2rib
Multipath: eBGP iBGP
```

```
  Path type (0x7f64ac44bbec): local 0x4000008c 0x4000000 ref 0 adv path ref 0, path is
valid, not best reason: MAC Mobility Sequence
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.21.1.115)
      Origin IGP, MED not set, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:5:16 PCTAG:01:0:0:5474

  Advertised path-id 1

  Path type (0x7f64abf319dc): internal 0xc0000018 0x220000 ref 0 adv path ref 1, path is
valid, is best path, is in l2rib, remote nh not installed
            Imported from (0x7f64ac447ab8)
16365:31948749:[2]:[0]:[0]:[48]:[0050.56b6.1540]:[32]:[172.17.131.102]/144
  AS-Path: 65101 , path sourced external to AS
    172.21.11.164 (metric 2) from 172.21.10.131 (172.20.0.1)
      Origin IGP, MED 2000, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:65011:15171533 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5474
          OSPF RT:03100000:0000fe4d MAC Mobility Sequence:00:1
          Router MAC:0200.ac14.0005

  Path-id 1 not advertised to any peer



aci-spine-2101# show coop internal info repo ep key 15171533 0050.56b6.1540 | egrep
"EP|Tunnel nh|PcTag"
EP bd vnid : 15171533
EP mac :  00:50:56:B6:15:40
PcTag : 0x1001562
Tunnel nh : 172.21.11.164
ETEP Tunnel : 0.0.0.0
Current published TEP : 172.21.11.164
BackupTunnel nh : 0.0.0.0
Tunnel EP entry: 0x7ff46001f338
Backup Tunnel EP entry: (nil)
Real IPv4 EP : 172.17.131.102
Synthetic Flags IPv4 EP : 0x25
PcTag: 0x1001562
```

The handling of external subnets follows a different logic. The remote external subnet (10.58.30.64/26), which was advertised to the spine as a VPNv4 prefix, is not installed into the spine's own forwarding tables. Instead, the spine's role here is to simply reflect this route down to all ACI leaf switches.

```
aci-spine-2101# show bgp vpnv4 uni 172.17.137.0/24 vrf overlay-1
Route Distinguisher: 3102:2228225
BGP routing table entry for 10.58.30.64/26, version 770 dest ptr 0x7f64b6b40b84
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in urib, is not in HW, is locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7f64ac444e3c): internal 0x40000018 0x40 ref 1 adv path ref 1, path is valid,
is best path
             Imported to 1 destination(s)
  AS-Path: 65101 65100 , path sourced external to AS
    172.21.10.133 (metric 2) from 172.21.10.133 (172.20.0.2)
      Origin incomplete, MED 45, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 0
      Received path-id 1
      Extcommunity: RT:65011:2228225 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5475
          OSPF RT:03100000:0000fe4d VNID:2228225 OSPF RT:0.0.0.0:0:0

  Path-id 1 advertised to peers:
    172.21.10.128      172.21.10.130
Route Distinguisher: 3101:2228225
BGP routing table entry for 10.58.30.64/26, version 750 dest ptr 0x7f64b6b43424
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in urib, is not in HW, is locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7f64ac446a44): internal 0x40000018 0x40 ref 1 adv path ref 1, path is valid,
is best path
             Imported to 1 destination(s)
  AS-Path: 65101 65100 , path sourced external to AS
    172.21.10.131 (metric 2) from 172.21.10.131 (172.20.0.1)
      Origin incomplete, MED 45, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 0
      Received path-id 1
      Extcommunity: RT:65011:2228225 SOO:65011:4093640685
```

```
        COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:5475
        OSPF RT:03100000:0000fe4d VNID:2228225 OSPF RT:0.0.0.0:0:0

  Path-id 1 advertised to peers:
    172.21.10.128      172.21.10.130
```

Because this reflection occurs over an internal BGP (iBGP) session, the next-hop attribute of the VPNv4 route remains unchanged throughout the process. As a result, the ACI leaves will install the route in their forwarding tables with the next-hop correctly pointing to the origin of the route: the ACI BGW's Primary TEP IP. This final step on the leaf is precisely what we will verify in the next section.

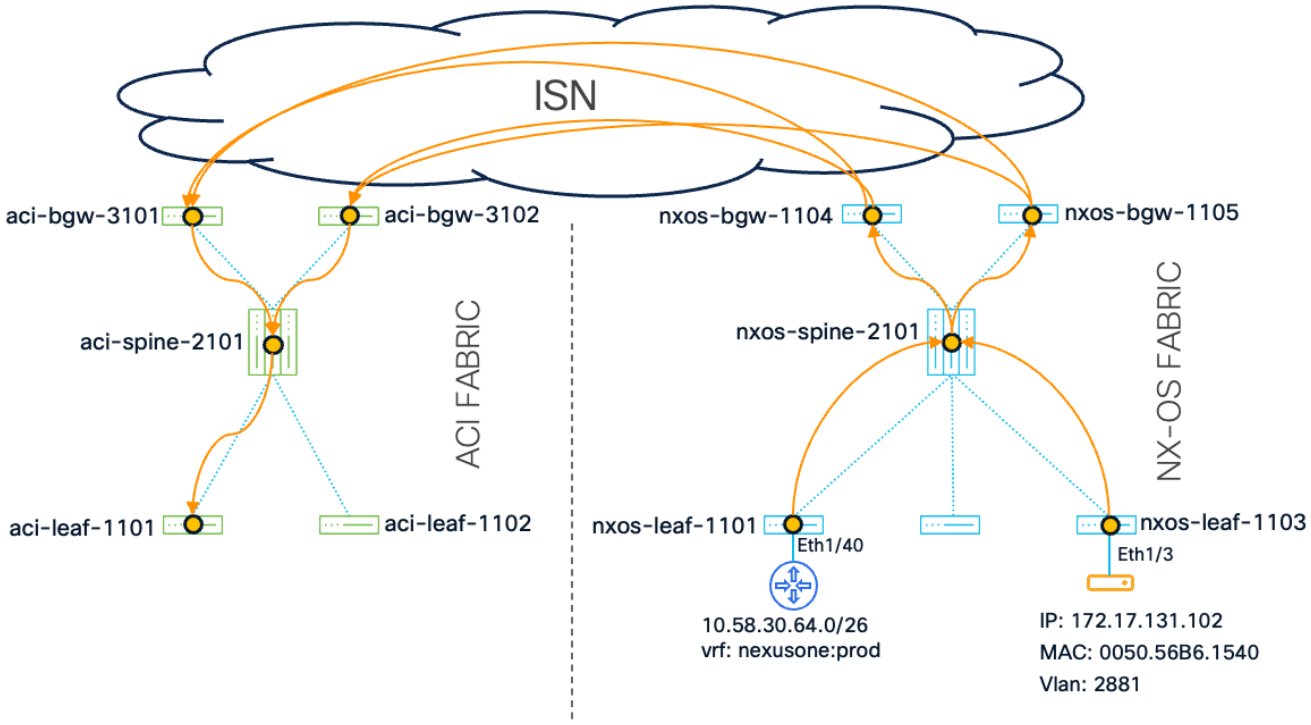

Figure 54 Control Plane verifications on ACI Leaf

Before examining the command outputs on the ACI leaf, it is essential to understand the key principles governing how it learns endpoint and external route information from a connected NX-OS fabric:

- Remote MAC Address Learning: ACI leaves do not learn remote MAC addresses from the control plane. The spine's COOP database is never advertised to the leaves. Instead, leaves rely on conversational data-plane learning. This means that aci-leaf-1101 will only learn the MAC address 0050.56b6.1540 after it decapsulates a bridged frame from that original source and forwards it to a local endpoint. This data-plane process also includes learning the translated pcTag associated with that MAC address. In our output we can see the remote endpoint in aci-leaf-1101 as there is live conversational traffic.
- Remote IP Address Learning: By design (as of ACI 6.1(5)), ACI leaves do not learn the specific IP addresses of remote fabrics endpoints in the same way they learn local fabric remote endpoints (not

connected to the same leaf). Remote endpoint learning only happens via ARP-based data plane learning, not IP-based data plane learning. Consequently, if the ACI leaf doesn't have an entry for the remote endpoint it will result in two different forwarding behaviors depending on the subnet configuration:

  - For a stretched subnet the leaf has a pervasive static route for the entire subnet that points to the spine-proxy VTEP address. Therefore, it forwards the routed traffic to the spine proxy. The spine, which does have the full endpoint information in COOP, then forwards the packet correctly toward the ACI BGW.
  - For a non-stretched subnet, the leaf learns a specific route for the endpoint via the control plane. This route, as we covered in the previous steps, originates as an EVPN Type-5 from NX-OS and is propagated from the ACI BGW to the spine and leaves as a VPNv4 route. The leaf installs this route with the ACI BGW Primary TEP IP (ptep) as the next-hop. In this case, policy enforcement does not happen on the leaf; the leaf uses a default PcTag of 1, and the actual pcTag-based policy is then enforced on the ACI BGW.

- Remote External Subnet Learning: Remote external subnets are always learned by the ACI leaves. The mechanism is the same as for a non-stretched subnet: the route is propagated via VPNv4 from the spine, and the leaf installs it with the ACI BGW Primary TEP IP (ptep) as its next-hop.

Let's now check the status of the remote endpoint and external subnet learning on the ACI leaf

```
aci-leaf-1101# show system internal epm endpoint mac 0050.56b6.1540
MAC : 0050.56b6.1540 ::: Num IPs : 0
Vlan id : 6 ::: Vlan vnid : 15171533 ::: VRF name : nexusone:prod
BD vnid : 15171533 ::: VRF vnid : 2228225
Phy If : 0 ::: Tunnel If : 0x18010008
Interface : Tunnel8
Flags : 0x80004810 ::: sclass : 5474 ::: Ref count : 4
EP Create Timestamp : 04/01/2026 12:57:27.251520
EP Update Timestamp : 04/08/2026 20:37:55.685396
EP Flags : locally-aged|MAC|sclass|timer|

aci-leaf-1101# show system internal epm endpoint ip 172.17.131.102
## NO ENTRY AS EXPECTED

aci-leaf-1101# show ip route 172.17.131.102 vrf nexusone:prod
172.17.131.0/24, ubest/mbest: 1/0, attached, direct, pervasive
    *via 172.21.11.162%overlay-1, [1/0], 08w06d, static
         recursive next hop: 172.21.11.162/32%overlay-1

aci-leaf-1101# show isis dteps vrf overlay-1 | grep 172.21.11.162
172.21.11.162      SPINE   N/A             PHYSICAL,PROXY-ACAST-V4

aci-leaf-1101# show ip route 10.58.30.64/26 vrf nexusone:prod
```

```
10.58.30.64/26, ubest/mbest: 2/0
    *via 172.21.10.131%overlay-1, [200/45], 1d11h, bgp-65011, internal, tag 65101, pc-tag: 1
         recursive next hop: 172.21.10.131/32%overlay-1
    *via 172.21.10.133%overlay-1, [200/45], 1d11h, bgp-65011, internal, tag 65101, pc-tag: 1
         recursive next hop: 172.21.10.133/32%overlay-1
```

With these detailed verifications, we have mapped the complete lifecycle of endpoint and external network reachability from NX-OS to ACI, confirming that every critical control-plane element is functioning as intended. The seamless propagation of reachability, security, and policy information across both fabrics demonstrates not only technical correctness but also the robustness and scalability of this architecture.

The control plane journey for endpoints and external networks learned in the ACI fabric

Having established a thorough understanding of how reachability and policy information is propagated from NX-OS to ACI, we now shift our focus to the reverse scenario: the control plane journey for endpoints and external networks as they are learned within the ACI fabric and propagated all the way to NX-OS environments.

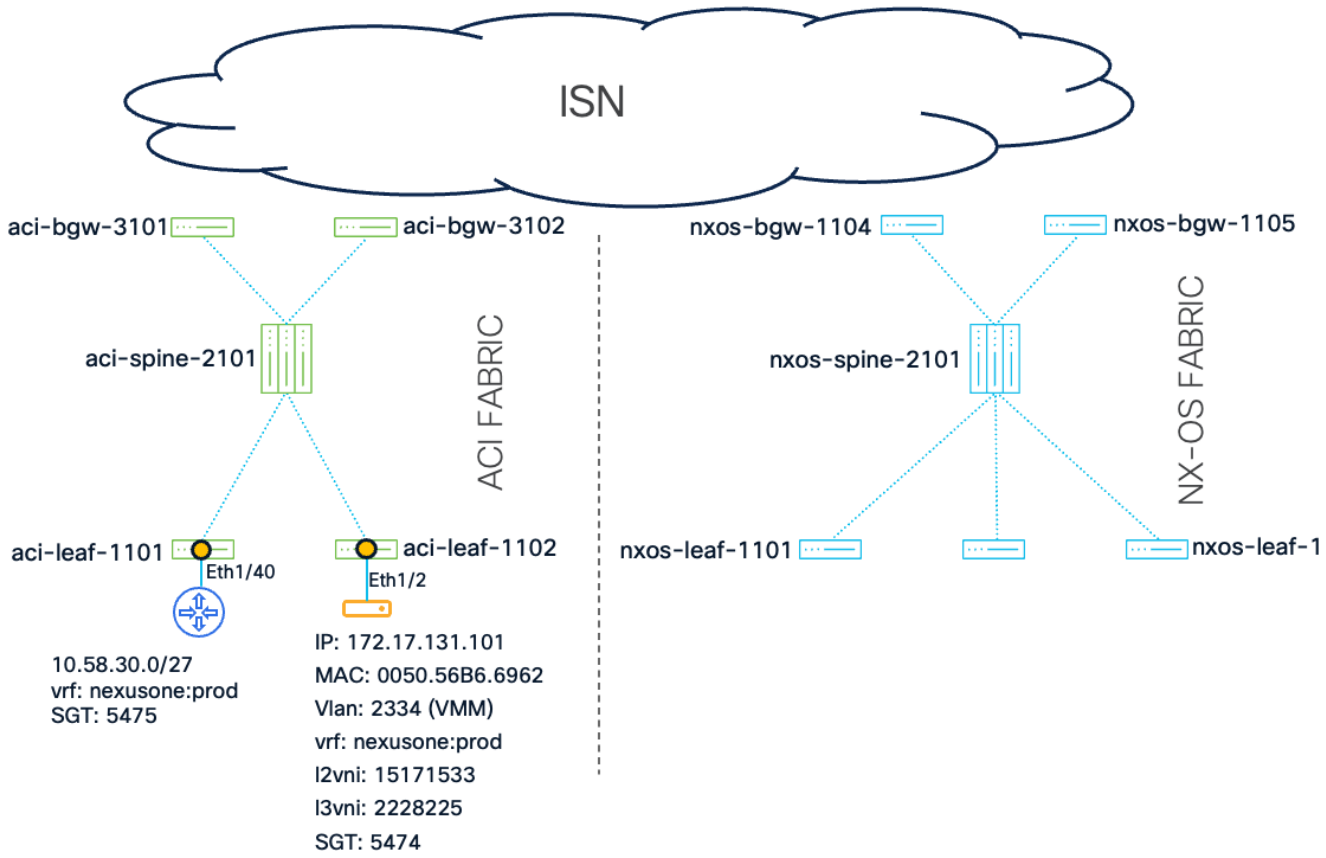

Figure 55 ACI attached resources

The following endpoint has been connected and discovered by a leaf node in the ACI fabric:

IP Address	MAC Address	VRF	L3VNI	VLAN	L2VNI	SGT
172.17.131.101	0050.56b6.6962	nexusone:prod	2228225	2334	15171533	5474

As well as the external route:

IP Address	Next-Hop	VRF	L3VNI	SGT
10.58.30.0/27	172.19.255.69/30	nexusone:prod	2228225	5475

VRF and BD are the same we have used in the previous example, hence both have been stretched from each fabric; We will now learn how ACI leaves initially discover the endpoint and the external route and propagate that all the way to NX-OS fabric. Over the journey the ACI spine and BGWs will generate the following EVPN routes:

- MAC-only Type-2
- MAC/IP Type-2
- BD SVI Type-5
- External Route Type-5
- Inclusive Ethernet Multicast Tag (IMET) Type-3 for BUM Ingress Replication.

Let's start our verifications from the aci-leaf-1102 and aci-leaf-1101, the devices that discover our end-host and the external network respectively.

The two commands that follow provide a comprehensive validation that the leaf switch has successfully discovered and programmed the local endpoint. The initial command confirms the basic identity and location (IP, MAC, and port), while the internal EPM details reveal the critical fabric-level context. The presence of the correct VNIDs and, most importantly, the pcTag, confirms that the endpoint is not only reachable but has also been correctly classified within its security policy group.

```
aci-leaf-1102# show endpoint ip 172.17.131.101
+-----------------------------------+---------------+-----------------+--------------+-------------+
      VLAN/                           Encap           MAC Address       MAC Info/       Interface
      Domain                          VLAN            IP Address        IP Info
+-----------------------------------+---------------+-----------------+--------------+-------------+
165                                       vlan-2334     0050.56b6.6962 L                      eth1/2
nexusone:prod                             vlan-2334     172.17.131.101 L                      eth1/2

aci-leaf-1102# show system internal epm endpoint mac 0050.56b6.6962

MAC : 0050.56b6.6962 ::: Num IPs : 1
IP# 0 : 172.17.131.101 ::: IP# 0 flags :  ::: l3-sw-hit: No
Vlan id : 165 ::: Vlan vnid : 8926 ::: VRF name : nexusone:prod
BD vnid : 15171533 ::: VRF vnid : 2228225
```

```
Phy If : 0x1a001000 ::: Tunnel If : 0
Interface : Ethernet1/2
Flags : 0x80004c04 ::: sclass : 5474 ::: Ref count : 5
EP Create Timestamp : 04/01/2026 12:55:36.550970
EP Update Timestamp : 04/09/2026 06:30:44.307996
EP Flags : local|IP|MAC|sclass|timer|

::::
```

Next, the show ip route command confirms that aci-leaf-1101 has learned a path to the external prefix 10.58.30.0/27 via BGP. While this establishes basic reachability, the 'show zoning-prefixes scope 2228225' command, which prints only the prefixes associated with the 2228225 VNI/VRF, reveals the critical ACI-specific detail: this external prefix has been mapped to ESG with pcTag 5475.

The final "show bgp vpnv4 unicast" command shows the external route VPNv4 advertisement to the spine.

```
aci-leaf-1101# show ip route 10.58.30.0/27 vrf nexusone:prod
IP Route Table for VRF "nexusone:prod"

10.58.30.0/27, ubest/mbest: 1/0
    *via 172.19.255.69%nexusone:prod, [20/44], 00:11:14, bgp-65011, external, tag 65100
         recursive next hop: 172.19.255.69/32%nexusone:prod

aci-leaf-1101# show zoning-prefixes scope 2228225
+---------+---------------+----------------+-------+-----------+
| Vrf-Vni |    Vrf-Name   |    Address     | Class | OperState |
+---------+---------------+----------------+-------+-----------+
| 2228225 | nexusone:prod |      ::/0      |   15  |  enabled  |
| 2228225 | nexusone:prod |   0.0.0.0/0    |   15  |  enabled  |
| 2228225 | nexusone:prod | 10.58.30.64/26 |  5475 |  enabled  |
| 2228225 | nexusone:prod | 10.58.30.0/27  |  5475 |  enabled  |
+---------+---------------+----------------+-------+-----------+

aci-leaf-1101# show bgp vpnv4 unicast 10.58.30.0/27 vrf overlay-1
BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 1101:2228225    (VRF nexusone:prod)
BGP routing table entry for 10.58.30.0/27, version 269 dest ptr 0x7f6a2e95908c
Paths: (1 available, best #1)
Flags: (0x00000000080c001a 0000000000) on xmit-list, is in urib, is best urib route, is in
HW, exported
  vpn: version 1495, (0x0000000000100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
```

```
  Path type (0x7f6a6406ca94): external 0x28 0x400 ref 0 adv path ref 2, path is valid, is
best path, in rib
  AS-Path: 65100 , path sourced external to AS
    172.19.255.69 (metric 0) from 172.19.255.69 (172.31.255.3)
      Origin incomplete, MED 44, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Extcommunity: RT:65011:2228225 PCTAG:01:0:0:5475 VNID:2228225
          OSPF RT:0.0.0.0:0:0

  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 advertised to peers:
    172.21.10.129
```

The external route prefix is getting propagated to the spines, and this is where we will focus our next verifications.

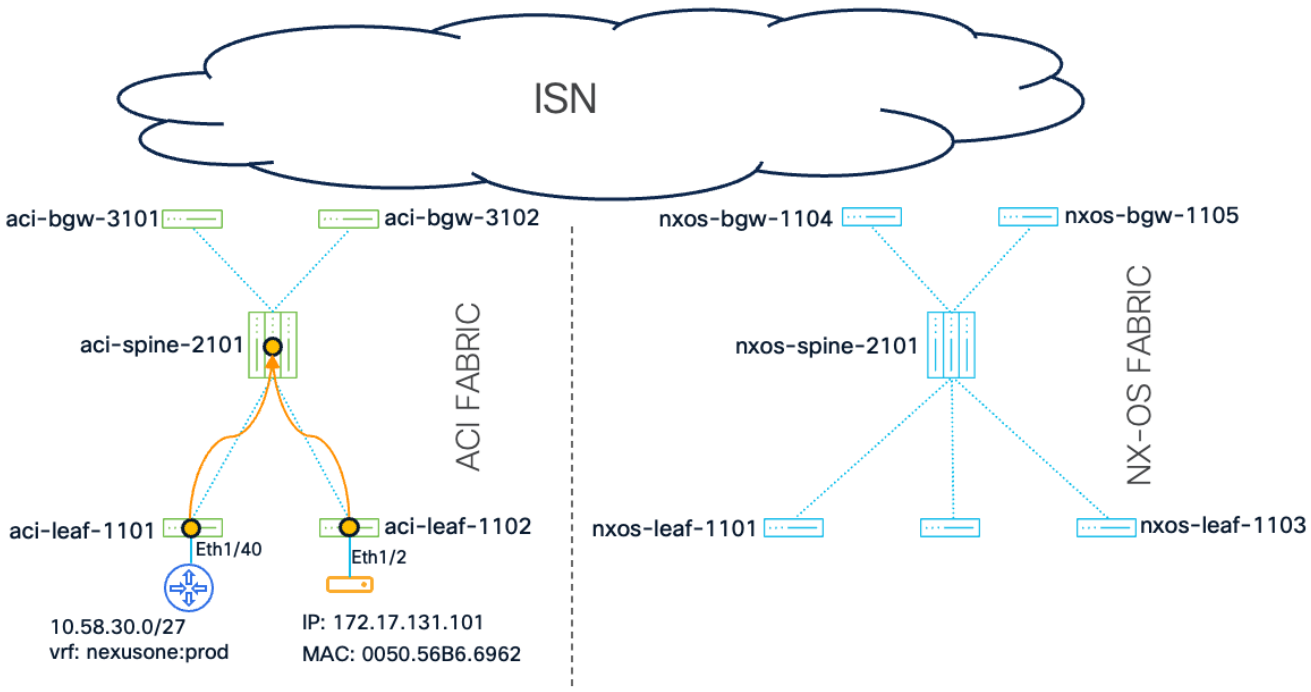

Figure 56 Control Plane verifications on ACI Spines

The verification begins with the COOP. The show coop commands confirms that the spine has successfully learned the endpoint's MAC and IP addresses from the leaf switch and has mapped them to the correct leaf VTEP address. Critically, the COOP database also stores the endpoint's security classification (pcTag), which is essential for the correct propagation via EVPN. The command “show coop internal info repo ep

key 15171533 00:50:56:B6:69:62” searches the MAC address in the 15171533 VNI which represents the Bridge Domain.

The spine then performs its key role in preparing this information for external advertisement by redistributing the host route from COOP into the BGP L2VPN EVPN address family. The show bgp output validates that the spine has generated EVPN Type-2 routes (for both MAC-Only and MAC/IP) and, most importantly, has embedded the security pcTag as a BGP extended community. This action ensures that when the route is advertised to the BGWs, its security context travels with it, allowing for consistent policy application across the entire architecture. The absence of a Type-5 route for the BD subnet at this stage is expected.

```
aci-spine-2101# show coop internal info ip-db key 2228225 172.17.131.101
IP address : 172.17.131.101
Vrf : 2228225
Flags : 0
EP bd vnid : 15171533
EP mac :  00:50:56:B6:69:62
Publisher Id : 172.21.10.130
Record timestamp : 04 09 2026 05:56:48 307685813
Publish timestamp : 04 09 2026 05:56:48 309040889
Seq No: 0
Remote publish timestamp: 01 01 1970 00:00:00 0
URIB Tunnel Info
Num tunnels : 1
Tunnel address : 172.21.10.130

aci-spine-2101# show coop internal info repo ep key 15171533 00:50:56:B6:69:62 | egrep
"EP|Tunnel nh|PcTag"

EP bd vnid : 15171533
EP mac :  00:50:56:B6:69:62
PcTag : 0x1001562
Tunnel nh : 172.21.10.130

aci-spine-2101# show bgp l2vpn evpn 00:50:56:B6:69:62 vrf overlay-1
Route Distinguisher: 16365:15171533    (L2VNI 15171533)
BGP routing table entry for [2]:[0]:[15171533]:[48]:[0050.56b6.6962]:[0]:[0.0.0.0]/216,
version 247 dest ptr 0x7f64b6a0fcae
BGW RD: 16365:15171533    (L2VNI 15171533)
Local Route Distinguisher: 1:16777199    (L2VNI 1)
Paths: (1 available, best #1)
Flags: (0x000000000000010a 0000000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP
```

```
  Advertised path-id 1
  Path type (0x7f64ac445fe8): local 0x4000008c 0x4000000 ref 0 adv path ref 1, path is
valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.21.1.115)
      Origin IGP, MED not set, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Received label 15171533
      Extcommunity: RT:5:16 PCTAG:01:0:0:5474

  Path-id 1 advertised to peers:
    172.21.10.131      172.21.10.133
BGP routing table entry for
[2]:[0]:[15171533]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272, version 250 dest ptr
0x7f64b6a0fbb0
BGW RD: 16365:15171533    (L2VNI 15171533)
Local Route Distinguisher: 1:16777199    (L2VNI 1)
Paths: (1 available, best #1)
Flags: (0x000000000000010a 0000000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7f64ac4462f4): local 0x4000008c 0x4000000 ref 0 adv path ref 1, path is
valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.21.1.115)
      Origin IGP, MED not set, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:5:16 PCTAG:01:0:0:5474

  Path-id 1 advertised to peers:
    172.21.10.131      172.21.10.133
```

We can observe a similar behavior for the 10.58.30.0/27 subnet. This subnet is received by the Spine via a VPNv4 NLRI that contains the pcTag 5475 and reflected to the remaining VTEPs in the fabric, including the BGWs (172.21.10.131 and 172.21.10.133).

```
aci-spine-2101# show bgp vpnv4 unicast 10.58.30.0/27 vrf overlay-1
BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 1101:2228225
BGP routing table entry for 10.58.30.0/27, version 890 dest ptr 0x7f64b6b40f6c
```

```
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in urib, is not in HW, is locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x7f64ac445148): internal 0x40000018 0x40 ref 1 adv path ref 1, path is valid,
is best path
             Imported to 1 destination(s)
  AS-Path: 65100 , path sourced external to AS
    172.21.10.128 (metric 2) from 172.21.10.128 (172.21.10.128)
      Origin incomplete, MED 44, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 0
      Received path-id 1
      Extcommunity: RT:65011:2228225 PCTAG:01:0:0:5475 VNID:2228225
          OSPF RT:0.0.0.0:0:0

  Path-id 1 advertised to peers:
    172.21.10.130      172.21.10.131      172.21.10.133
```

Next, we will verify how this gets propagated to the BGWs:

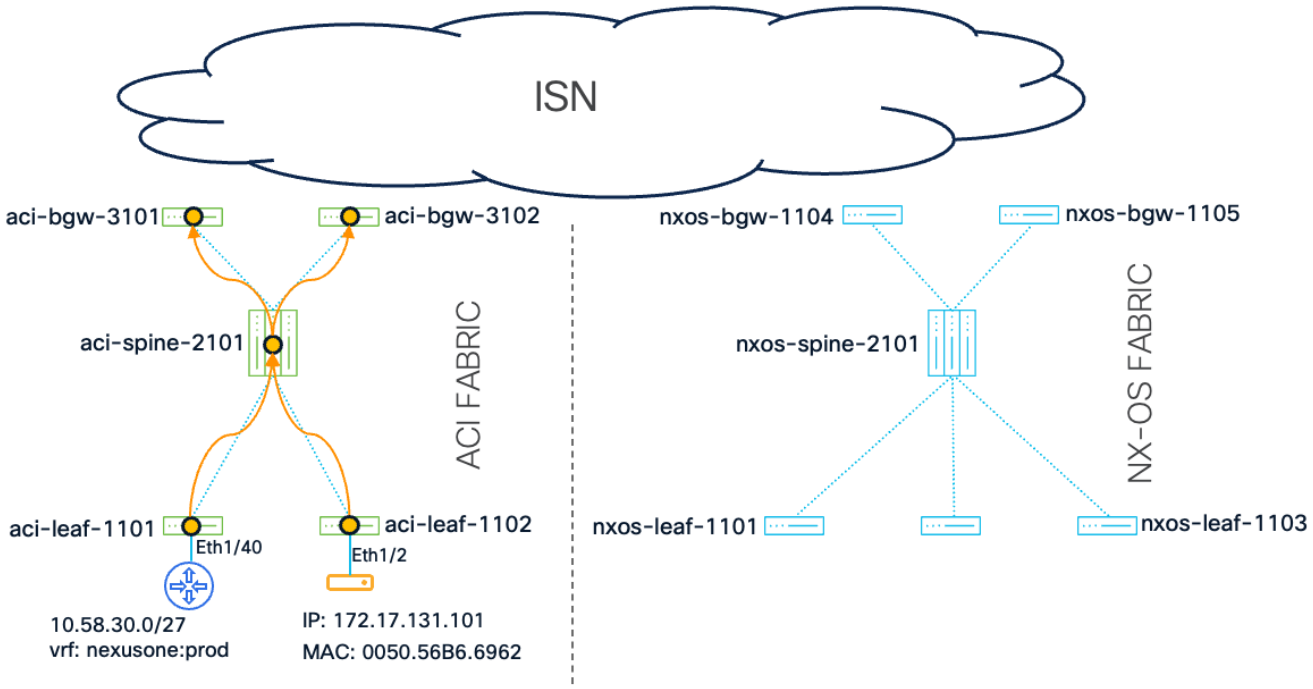

Figure 57 Control Plane verifications on ACI BGW

The NLRI prefix re-origination which has been covered in the previous sections and is at the base of the BGW functionality can be seen again now. The prefixes for the ACI attached endpoint are received from the ACI spines via EVPN advertisements and re-originated in a new EVPN route before being sent to the remote NX-OS BGWs (172.20.10.1 and 172.20.10.3). At this stage the route next-hop is 172.21.11.162 which identifies the spine anycast VTEP address.

The next output will focus on the MAC/IP route, the same will happen to the MAC-Only.

Note: Note that namespace normalization has not happened yet; those modifications are triggered when the prefixes get sent to the remote NX-OS BGWs.

```
aci-bgw-3102# show bgp l2vpn evpn 172.17.131.101 vrf overlay-1
Route Distinguisher: 16365:15171533
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272,
version 251 dest ptr 0xa2883012
Paths: (1 available, best #1)
Flags: (0x0000000000000202 0x00000800) on xmit-list, is not in rib/evpn, is not in HW, is
locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x97190bf8): internal 0x40000018 0x4002000 ref 2 adv path ref 1, path is valid,
is best path, remote nh not installed
              Imported to 2 destination(s)
  AS-Path: NONE, path sourced internal to AS
    172.21.11.162 (metric 2) from 172.21.10.129 (172.21.1.115)
```

```
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:65011:2228225 RT:65011:15171533 SOO:1:1 ENCAP:8
          PCTAG:01:0:0:5474 Router MAC:0200.0000.0000

  Path-id 1 not advertised to any peer

Route Distinguisher: 16365:31948749    (L2VNI 15171533)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272,
version 246 dest ptr 0xa288372e
Paths: (1 available, best #1)
Flags: (0x0000000000000212 0000000000) on xmit-list, is in rib/evpn, is not in HW
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x973e6594): internal 0xc0000018 0x400 ref 0 adv path ref 1, path is valid, is
best path, remote nh not installed, in rib
             Imported from (0x97190bf8)
16365:15171533:[2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/144
  AS-Path: NONE, path sourced internal to AS
    172.21.11.162 (metric 2) from 172.21.10.129 (172.21.1.115)
      Origin IGP, MED not set, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 15171533 2228225
      Extcommunity: RT:65011:2228225 RT:65011:15171533 SOO:1:1 ENCAP:8
          PCTAG:01:0:0:5474 Router MAC:0200.0000.0000

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.3        172.20.20.1        172.20.20.5
    172.20.30.3
```

The next command shows that our BGW is originating an EVPN Type-5 route, which represents the entire Bridge Domain subnet, and advertising it toward the remote BGWs. The command output also implicitly confirms the bidirectional nature of a stretched subnet; while our BGW advertises the prefix, it simultaneously receives an equivalent advertisement from the remote NX-OS BGWs as we saw before.

It is important to note that at this stage of the control plane process, the security policy information (pcTag) is not yet associated with the re-generated EVPN prefix. This security tag is dynamically attached later in the advertisement pipeline as the route is being prepared for transmission to its peers.

```
aci-bgw-3102# show bgp l2vpn evpn 172.17.131.0 vrf overlay-1
Route Distinguisher: 16365:19005441    (L3VNI 2228225)
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]:[0.0.0.0]/224, version 196 dest
ptr 0xa2886b14
```

Paths: (1 available, best #1)

Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn

Multipath: eBGP iBGP

  Advertised path-id 1

  Path type (0x97193088): local 0x4000008c 0x4400000 ref 0 adv path ref 1, path is valid, is best path

  AS-Path: NONE, path locally originated

    0.0.0.0 (metric 0) **from 0.0.0.0 (172.20.0.2)**

      Origin incomplete, MED 0, localpref 100, weight 32768 tag 0, propagate 0, floating svi 0, tunnel resolved 0

      Received label 2228225

      Extcommunity: **RT:65011:2228225 VNID:2228225**

  Path-id 1 advertised to peers:

    172.20.10.1 **172.20.10.3** 172.20.20.1 172.20.20.5

    172.20.30.3

Route Distinguisher: 65101:50000

BGP routing table entry for **[5]**:[0]:[0]:**[24]:[172.17.131.0]**:[0.0.0.0]/224, version 296 dest ptr 0xa28824b2

Paths: (2 available, best #2)

Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn, is not in HW, is locked

Multipath: eBGP iBGP

  Path type (0x9718fe90): external 0x40000028 0x4000000 ref 0 adv path ref 0, path is valid, not best reason: newer EBGP path, remote nh not installed

  AS-Path: 65101 , path sourced external to AS

    172.20.0.5 (metric 0) **from 172.20.10.3** (172.20.10.3)

      Origin incomplete, MED 1, localpref 100, weight 0 tag 4294966000, propagate 0, floating svi 0, tunnel resolved 0

      Received label 50000

      Extcommunity: **RT:65011:50000** SOO:65011:4093640685

          COST:pre-bestpath:170:1610612736 ENCAP:8 **PCTAG:00:0:0:1**

          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005

  Advertised path-id 1

  Path type (0x9718fef8): external 0x40000028 0x4000000 ref 1 adv path ref 1, path is valid, is best path, remote nh not installed

              Imported to 1 destination(s)

  AS-Path: 65101 , path sourced external to AS

    172.20.0.5 (metric 0) **from 172.20.10.1** (172.20.10.1)

```
      Origin incomplete, MED 1, localpref 100, weight 0 tag 4294966000, propagate 0,
floating svi 0, tunnel resolved 0
      Received label 50000
      Extcommunity: RT:65011:50000 SOO:65011:4093640685
          COST:pre-bestpath:170:1610612736 ENCAP:8 PCTAG:00:0:0:1
          OSPF RT:03100000:0000fe4d Router MAC:0200.ac14.0005

  Path-id 1 not advertised to any peer
```

At this stage we can also verify the Type-3 route generation which will be advertised to the NX-OS BGWs and signal them to send the BUM traffic for this specific L2VNI. Even in this case the VNI value must be normalized, the one indicated in the output is the ACI BD VNI.

```
aci-bgw-3101# aci-bgw-3102# show bgp l2vpn evpn route-type 3 vrf overlay-1 | egrep -A3 -B15
"Label: 15171533"

Route Distinguisher: 16365:31948749    (L2VNI 15171533)
BGP routing table entry for [3]:[0]:[32]:[172.20.0.2]/88, version 247 dest ptr 0xa2886ea2
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0xa2b4f8f8): local 0x8c 0x0 ref 0 adv path ref 1, path is valid, is best path
  AS-Path: NONE, path locally originated
    0.0.0.0 (metric 0) from 0.0.0.0 (172.20.0.2)
      Origin IGP, MED not set, localpref 100, weight 32768 tag 0, propagate 0, floating svi
0, tunnel resolved 0
      Extcommunity: RT:65011:15171533
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 15171533, Tunnel Id: 172.20.0.2

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.3        172.20.20.1        172.20.20.5
```

And finally, our external network is also learned via the VPNv4 address family and re-originated in a standard EVPN Type-5 route advertised to the remote NX-OS BGWs.

```
aci-bgw-3102# show bgp vpnv4 unicast 10.58.30.0/27 vrf overlay-1
BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 1101:2228225
BGP routing table entry for 10.58.30.0/27, version 296 dest ptr 0xa294c2b8
```

```
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in urib, is not in HW, is locked
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x97190850): internal 0x40000018 0x40 ref 1 adv path ref 1, path is valid, is
best path
              Imported to 1 destination(s)
  AS-Path: 65100 , path sourced external to AS
    172.21.10.128 (metric 3) from 172.21.10.129 (172.21.1.115)
      Origin incomplete, MED 44, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 0
      Received path-id 1
      Extcommunity: RT:65011:2228225 PCTAG:01:0:0:5475 VNID:2228225
          OSPF RT:0.0.0.0:0:0
      Originator: 172.21.10.128 Cluster list: 172.21.1.115

  Path-id 1 not advertised to any peer

BGP routing table information for VRF overlay-1, address family VPNv4 Unicast
Route Distinguisher: 3102:2228225    (VRF nexusone:prod)
BGP routing table entry for 10.58.30.0/27, version 178 dest ptr 0x96ced658
Paths: (1 available, best #1)
Flags: (0x00000000000c001a 0x00000200) on xmit-list, is in urib, is best urib route, is in
HW, exported
  vpn: version 324, (0x0000000000100002) on xmit-list
Multipath: eBGP iBGP

  Advertised path-id 1, VPN AF advertised path-id 1
  Path type (0x973e615c): internal 0xc0000018 0x440 ref 0 adv path ref 2, path is valid, is
best path, in rib
              Imported from (0x97190850) 1101:2228225:10.58.30.0/27
  AS-Path: 65100 , path sourced external to AS
    172.21.10.128 (metric 3) from 172.21.10.129 (172.21.1.115)
      Origin incomplete, MED 44, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 0
      Received path-id 1
      Extcommunity: RT:65011:2228225 PCTAG:01:0:0:5475 VNID:2228225
          OSPF RT:0.0.0.0:0:0
      Originator: 172.21.10.128 Cluster list: 172.21.1.115
```

```
  VRF advertise information:
  Path-id 1 not advertised to any peer

  VPN AF advertise information:
  Path-id 1 not advertised to any peer

aci-bgw-3102# show bgp l2vpn evpn 10.58.30.0 vrf overlay-1
Route Distinguisher: 16365:19005441    (L3VNI 2228225)
BGP routing table entry for [5]:[0]:[0]:[27]:[10.58.30.0]:[0.0.0.0]/224, version 369 dest
ptr 0xa2882ea6
Paths: (1 available, best #1)
Flags: (0x0000000000000002 0000000000) on xmit-list, is not in rib/evpn
Multipath: eBGP iBGP

  Advertised path-id 1
  Path type (0x97191070): local 0x4000008c 0x4000000 ref 0 adv path ref 1, path is valid, is
best path
  AS-Path: 65100 , path sourced external to AS
    0.0.0.0 (metric 0) AS-Path: 65100 (172.20.0.2)
      Origin incomplete, MED 44, localpref 100, weight 0 tag 0, propagate 0, floating svi 0,
tunnel resolved 0
      Received label 2228225
      Extcommunity: RT:65011:2228225 PCTAG:01:0:0:5475 VNID:2228225
          OSPF RT:0.0.0.0:0:0
      Originator: 172.21.10.128 Cluster list: 172.21.1.115

  Path-id 1 advertised to peers:
    172.20.10.1        172.20.10.3        172.20.20.1        172.20.20.5
    172.20.30.3
```

The ACI control plane verifications are now done, let's move to the NX-OS fabric starting from the NX-OS BGW:

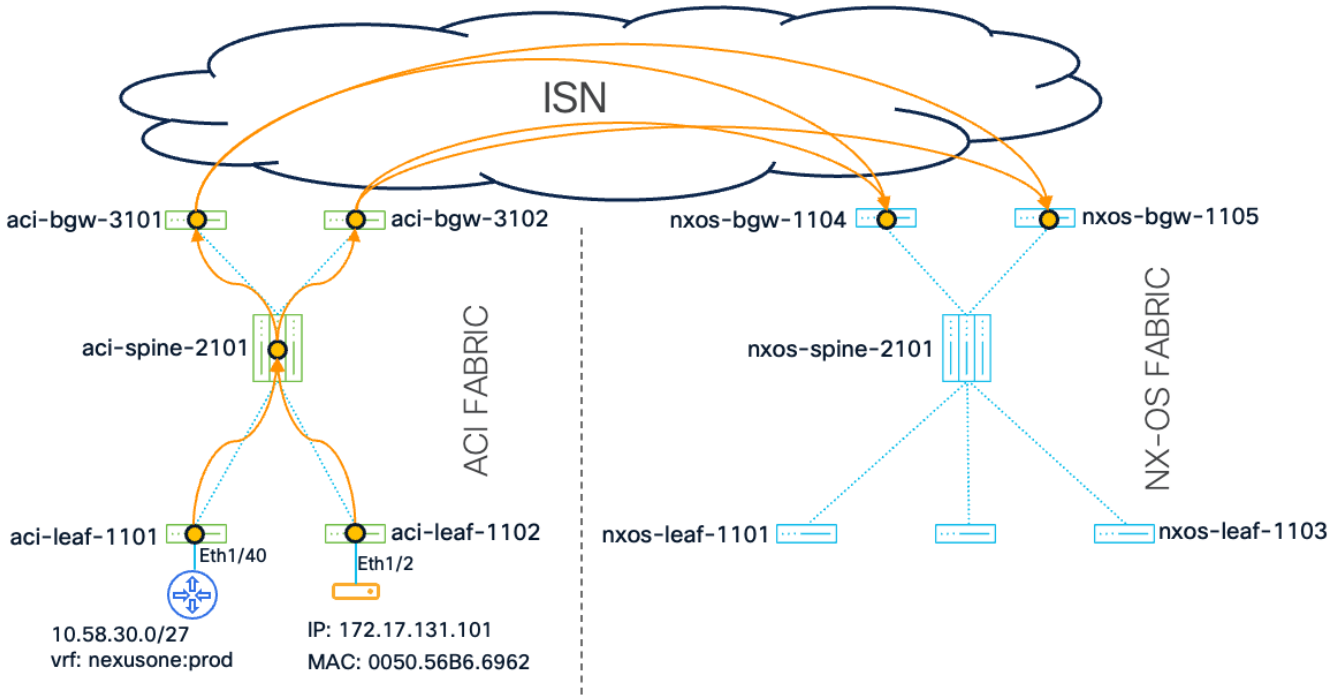

Figure 58 Control Plane verifications on NX-OS BGW

Beginning with endpoint route validation, the output confirms that both MAC-only and MAC/IP EVPN Type-2 routes are received from the ACI BGWs, imported in the local BGP table and subsequently re-originated toward the local spine (172.20.10.4). The most critical observation is the successful namespace normalization performed by the ACI BGW; it correctly translated the ACI-specific identifiers into values compatible with the NX-OS fabric. Specifically, the output shows the L2 VNI is translated to 30000, the L3 VNI to 50000, and the security policy is mapped to SGT 10002.

We focus once more on the MAC-IP route as the MAC-Only shares the same fate.

```
nxos-bgw-1104# show bgp l2vpn evpn 172.17.131.101
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 16365:31948749
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272,
version 195
Paths: (2 available, best #2)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Path type: external, path is valid, not best reason: newer EBGP path, no labeled nexthop
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:80:10002
```

```
          Router MAC:c02c.17ff.b946

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported to 3 destination(s)
             Imported paths list: nexusone:prod L3-50000 L2-30000
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:80:10002
          Router MAC:7cad.4f2c.1199

  Path-id 1 not advertised to any peer

Route Distinguisher: 172.20.10.1:35648    (L2VNI 30000)
BGP routing table entry for [2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272,
version 207
Paths: (1 available, best #1)
Flags: (0x000212) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop, in rib
             Imported from
16365:31948749:[2]:[0]:[0]:[48]:[0050.56b6.6962]:[32]:[172.17.131.101]/272
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin IGP, MED not set, localpref 100, weight 0
      Received label 30000 50000
      Extcommunity: RT:65101:30000 RT:65101:50000 ENCAP:8 PCTAG:0:80:10002
          Router MAC:7cad.4f2c.1199

  Path-id 1 (dual) advertised to peers:
    172.20.10.4
```

The Type-3 IMET route from the ACI BGWs is also confirmed, signaling that BUM traffic for the L2VNI should be replicated to the ACI BGW's external Primary TEP IP (ptep). Notably, this route is not re-originated internally within the NX-OS fabric. This is expected behavior, as the NX-OS fabric in this design relies on underlay multicast to propagate L2 BUM frames, making an internal EVPN advertisement for this purpose unnecessary.

```
nxos-bgw-1104# show bgp l2vpn evpn vni-id 30000 route-type 3
BGP routing table information for VRF default, address family L2VPN EVPN
```

```
Route Distinguisher: 172.20.10.1:35648    (L2VNI 30000)
BGP routing table entry for [3]:[0]:[32]:[172.20.0.1]/88, version 96
Paths: (1 available, best #1)
Flags: (0x000012) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported from 172.20.0.1:1:[3]:[0]:[32]:[172.20.0.1]/88
  AS-Path: 65011 , path sourced external to AS
    172.20.0.1 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin IGP, MED not set, localpref 100, weight 0
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:80:1
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 30000, Tunnel Id: 172.20.0.1

  Path-id 1 (dual) not advertised to any peer
BGP routing table entry for [3]:[0]:[32]:[172.20.0.2]/88, version 94
Paths: (1 available, best #1)
Flags: (0x000012) (high32 0x000400) on xmit-list, is in l2rib/evpn, is not in HW

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported from 172.20.0.2:9:[3]:[0]:[32]:[172.20.0.2]/88
  AS-Path: 65011 , path sourced external to AS
    172.20.0.2 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin IGP, MED not set, localpref 100, weight 0
      Extcommunity: RT:65101:30000 ENCAP:8 PCTAG:0:80:1
      PMSI Tunnel Attribute:
        flags: 0x00, Tunnel type: Ingress Replication
        Label: 30000, Tunnel Id: 172.20.0.2

  Path-id 1 (dual) not advertised to any peer
```

The route for the stretched Bridge Domain subnet is also received from the ACI BGW, with its SGT correctly mapped to 1, as expected. However, this route is not propagated further toward the spine. This is a direct result of standard BGP best-path selection, as the same subnet is already learned locally within the NX-OS fabric with a more preferred metric.

```
nxos-bgw-1104# show bgp l2vpn evpn 172.17.131.0
BGP routing table information for VRF default, address family L2VPN EVPN
```

```
Route Distinguisher: 16365:19005441
BGP routing table entry for [5]:[0]:[0]:[24]:[172.17.131.0]/224, version 106
Paths: (2 available, best #2)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Path type: external, path is valid, not best reason: newer EBGP path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:80:1 Router MAC:c02c.17ff.b946

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported to 1 destination(s)
             Imported paths list: nexusone:prod
  Gateway IP: 0.0.0.0
  AS-Path: 65011 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin incomplete, MED 0, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:80:1 Router MAC:7cad.4f2c.1199

  Path-id 1 not advertised to any peer
```

The external route learned via the ACI L3Out follows the expected control plane behavior. The EVPN Type-5 route is received from the remote ACI BGW and is subsequently re-originated toward the local spines with the correct security mapping, SGT 10000.

```
nxos-bgw-1104# show bgp l2vpn evpn 10.58.30.0
BGP routing table information for VRF default, address family L2VPN EVPN
Route Distinguisher: 16365:19005441
BGP routing table entry for [5]:[0]:[0]:[27]:[10.58.30.0]/224, version 411
Paths: (2 available, best #2)
Flags: (0x000002) (high32 00000000) on xmit-list, is not in l2rib/evpn, is not in HW

  Path type: external, path is valid, not best reason: newer EBGP path, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65011 65100 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.2 (172.20.0.2)
      Origin incomplete, MED 44, localpref 100, weight 0
```

```
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:80:10000 Router MAC:7cad.4f2c.1199
          OSPF RT:0.0.0.0:0:0

  Advertised path-id 1
  Path type: external, path is valid, is best path, no labeled nexthop
             Imported to 1 destination(s)
             Imported paths list: nexusone:prod
  Gateway IP: 0.0.0.0
  AS-Path: 65011 65100 , path sourced external to AS
    172.20.0.3 (metric 0) from 172.20.0.1 (172.20.0.1)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:80:10000 Router MAC:c02c.17ff.b946
          OSPF RT:0.0.0.0:0:0

  Path-id 1 not advertised to any peer

Route Distinguisher: 172.20.10.1:4    (L3VNI 50000)
BGP routing table entry for [5]:[0]:[0]:[27]:[10.58.30.0]/224, version 391
Paths: (1 available, best #1)
Flags: (0x000002) (high32 0x000400) on xmit-list, is not in l2rib/evpn

  Advertised path-id 1
  Path type: local, path is valid, is best path, reoriginated, no labeled nexthop
  Gateway IP: 0.0.0.0
  AS-Path: 65011 65100 , path sourced external to AS
    172.20.11.1 (metric 0) from 0.0.0.0 (172.20.10.1)
      Origin incomplete, MED 44, localpref 100, weight 0
      Received label 50000
      Extcommunity: RT:65101:50000 ENCAP:8 PCTAG:0:0:10000 Site-ID:0:0
          Router MAC:4874.1038.2a4b OSPF RT:0.0.0.0:0:0

  Path-id 1 (dual) advertised to peers:
    172.20.10.4
```

Verification on the spine devices can be safely omitted in this scenario. In sharp contrast to ACI spines, which play an active role by maintaining the COOP database and performing redistribution, NX-OS spines do not participate in complex control-plane functions. Their sole responsibility is to ensure that all EVPN routes received from the VTEPs are reliably reflected and distributed throughout the fabric.

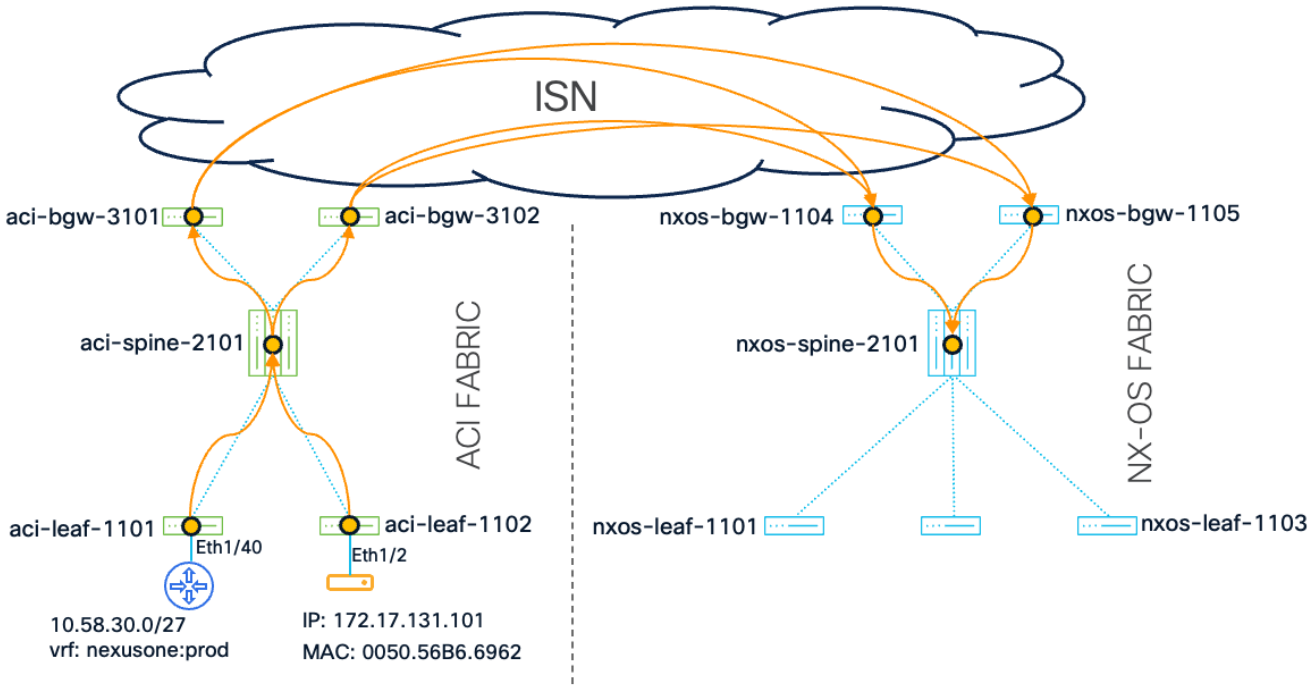

Figure 59 Control Plane verifications on NX-OS Spine

Finally, the verification process concludes on the NX-OS leaf switch, where we will confirm that all prefixes have been successfully received and installed.

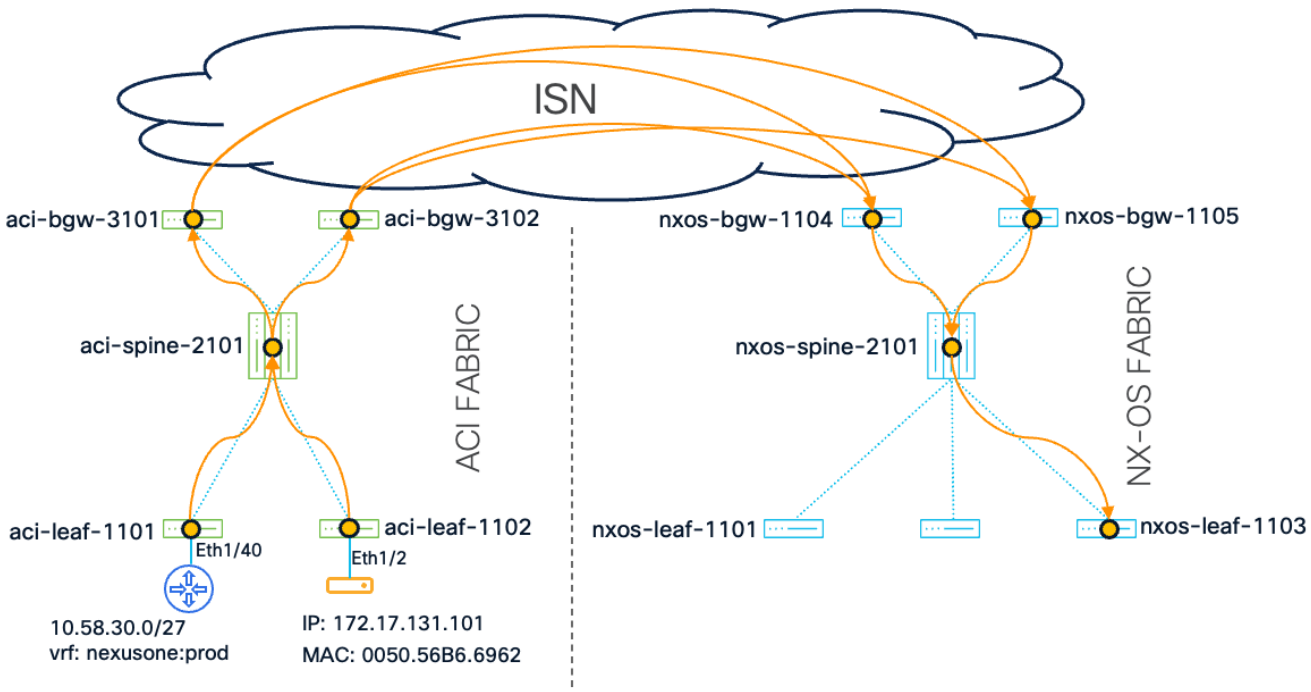

Figure 60 Control Plane verifications on NX-OS Leaf

This final check is paramount, as an NX-OS fabric fully depends on control plane advertisements to build and program its routing and forwarding tables. Without a valid BGP EVPN route received, no forwarding path will exist for the traffic.

You will probably notice that the next-hop for bridged and routed traffic is always **172.20.0.5** which in this case is represents the NX-OS BGWs Multi-Site VIP.

```
nxos-leaf-1103# show mac address-table esg address 0050.56b6.6962
   VNI      MAC Address      Type      age     Secure NTFY Ports     ESG Tag
---------+-----------------+--------+---------+------+----+------------------
C 2881     0050.56b6.6962   dynamic  NA         F      F    nve1(172.20.0.5) 10002

nxos-leaf-1103# show ip route  172.17.131.101 detail vrf nexusone:prod
172.17.131.101/32, ubest/mbest: 1/0
    *via 172.20.0.5%default, [200/0], 1w0d, bgp-65101, internal, tag 65011, remote-site,
segid: 50000 tunnelid: 0xac140005 encap: VXLAN
        BGP-EVPN: VNI=50000 (EVPN)
        recursive next hop: 172.20.0.5/32%default
        xri tag type: Security Group; value: 10002

nxos-leaf-1103# show ip route  10.58.30.0/27 detail vrf nexusone:prod
10.58.30.0/27, ubest/mbest: 1/0
    *via 172.20.0.5%default, [200/44], 01:44:32, bgp-65101, internal, tag 65011, remote-
site, segid: 50000 tunnelid: 0xac140005 encap: VXLAN
```

```
BGP-EVPN: VNI=50000 (EVPN)
recursive next hop: 172.20.0.5/32%default
xri tag type: Security Group; value: 10000
```

The verification confirms that all relevant prefixes have been successfully installed in both the Layer 2 and Layer 3 forwarding tables, and crucially, each prefix is associated with its correct SGT. This outcome validates the end-to-end integrity of the security policy.

In this scenario, an endpoint and an external route were intentionally classified into two distinct Endpoint Security Groups (ESGs) within the ACI fabric. As verified on the destination NX-OS leaf, these separate security classifications were perfectly preserved, with each prefix mapped to its corresponding, unique ESG. The critical action enabling this was performed by the ACI Border Gateway during its namespace normalization process. It is the BGW's responsibility to translate the ACI fabric's locally significant SGTs into the normalized values understood by the NX-OS domain, confirming its role as the essential policy translation point between the two domains.

The data plane journey for endpoints between ACI and NX-OS – Step by Step

Based on the previously covered scenario, we will now observe a packet walk for bridged traffic originating from an endpoint connected to the ACI fabric and destined for an endpoint connected to the remote NX-OS fabric. In this flow, the traffic remains within the same stretched Bridge Domain (L2VNI), and forwarding decisions are made based on MAC address learning.

We assume that no traffic has been sent between the endpoints prior to this flow. Therefore, both endpoints must perform ARP resolution to discover each other's MAC addresses across the multi-fabric domain.

Packet 1: ARP Resolution and BUM Traffic Flooding

- The ACI-connected endpoint 172.17.131.101 sends a broadcast ARP request to discover the MAC address of the NX-OS endpoint 172.17.131.102. The source MAC address will be 0050.56B6.6962.
- The ACI leaf switch (leaf 1102) receives this ARP broadcast frame.
- The leaf learns the source endpoint and classifies it into the appropriate Endpoint Security Group (ESG) based on APIC-defined classification rules.
- Since the destination MAC is unknown locally, the leaf treats the frame as unknown unicast and floods it to all local interfaces within the same bridge domain.
- Simultaneously, the leaf encapsulates the frame in an iVXLAN header with these key fields:
  - Outer Source IP: 172.21.10.130 (ACI leaf TEP address)
  - Outer Destination IP: 225.0.74.166 (multicast group address for the FTAG tree used for BUM replication)
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag)
  - iVXLAN Policy Applied Bit: 0

- The ACI fabric replicates this broadcast frame across the local fabric using the FTAG-based underlay multicast distribution tree. The assumption at this stage is that also aci-leaf-1101 has the Bridge Domain locally configured

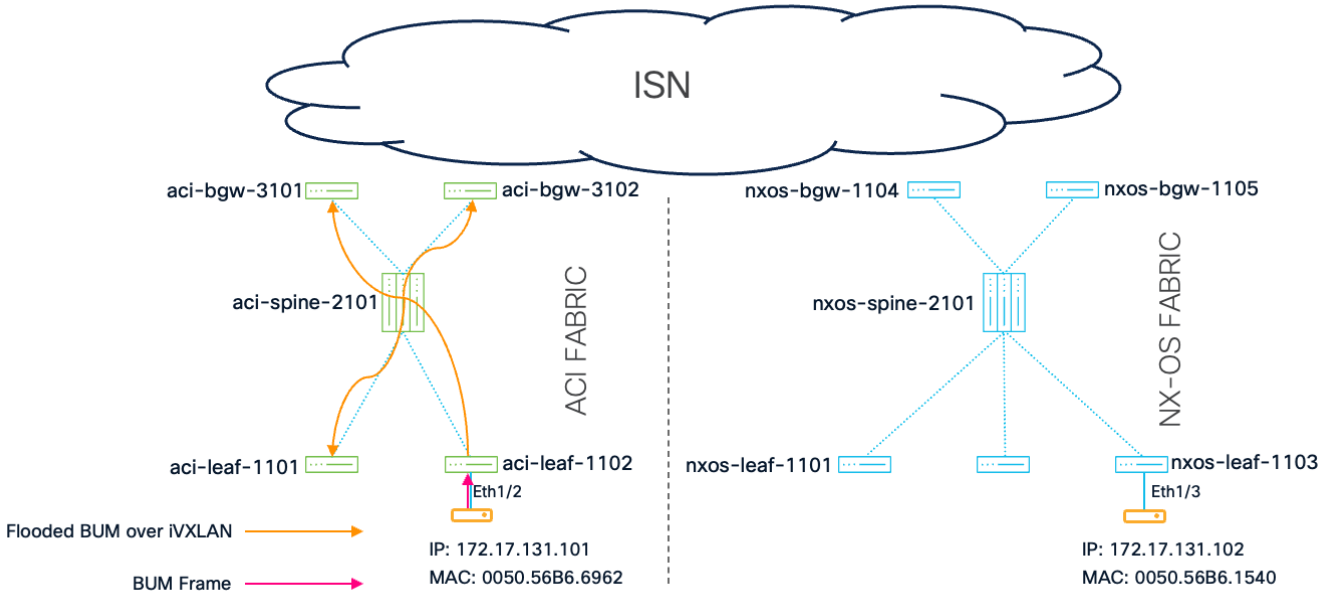

Figure 61 ARP request flooded inside ACI fabric

- ACI Border Gateways receive the encapsulated broadcast frame.
- The BGWs decapsulate the iVXLAN packet and performs a destination MAC lookup.
- Since the NX-OS endpoint MAC is not yet learned in EVPN, only the Designated Forwarder BGW continues to flood the original frame to the remote site.
- The DF BGW creates two separate VXLAN packets for the remote NX-OS fabric, performing ingress replication to the NX-OS BGWs.
- Namespace normalization is applied:
  - The ACI L2VNI 15171533 is mapped to the external normalized L2VNI 30000.
- The VXLAN headers for these packets have:
  - Outer Source IP: 172.20.0.1 (The rtep for aci-bgw-3101, the Designated Forwarder for this L2VNI)
  - Outer Destination IPs:

    Packet 1: 172.20.11.1 (PIP of nxos-bgw-1104)

    Packet 2: 172.20.11.3 (PIP of nxos-bgw-1105)
  - VXLAN VNI: 30000

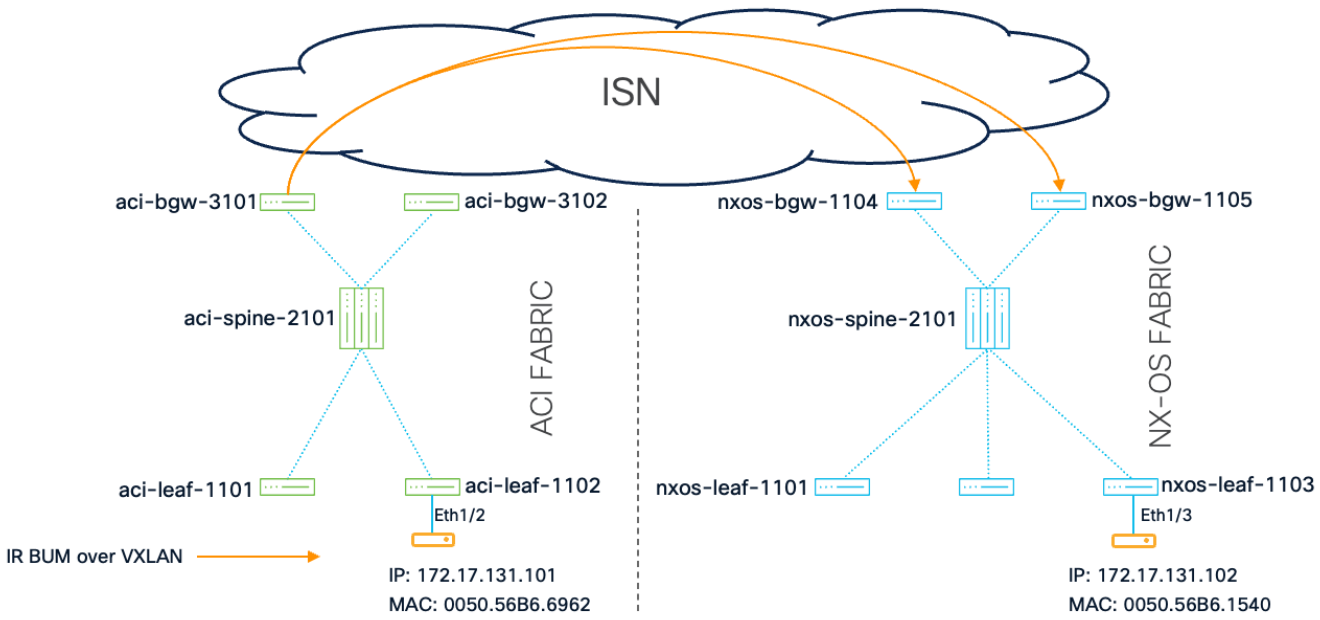

Figure 62 ARP request flooded in the Inter-Site network via Ingress Replication

- Both NX-OS BGWs receive the BUM packet, decapsulate it and look at the content.
- Only the Designated Forwarder BGW forwards the broadcast packet further into the local NX-OS fabric.
- Since this is a broadcast packet, no policy enforcement is applied at this stage.
- The packet follows the native underlay multicast distribution tree to reach the NX-OS-connected endpoint.
  - Outer Source IP: 172.20.11.1 (nxos-bgw-1104, the Designated Forwarder for this L2VNI)
  - Outer Destination IP: 239.1.1.0 (multicast group address for the L2VNI)
  - VXLAN VNI: 30000
- The assumption is that both nxos-leaf-1102 and nxos-leaf-1103 are attached to the VNI 30000

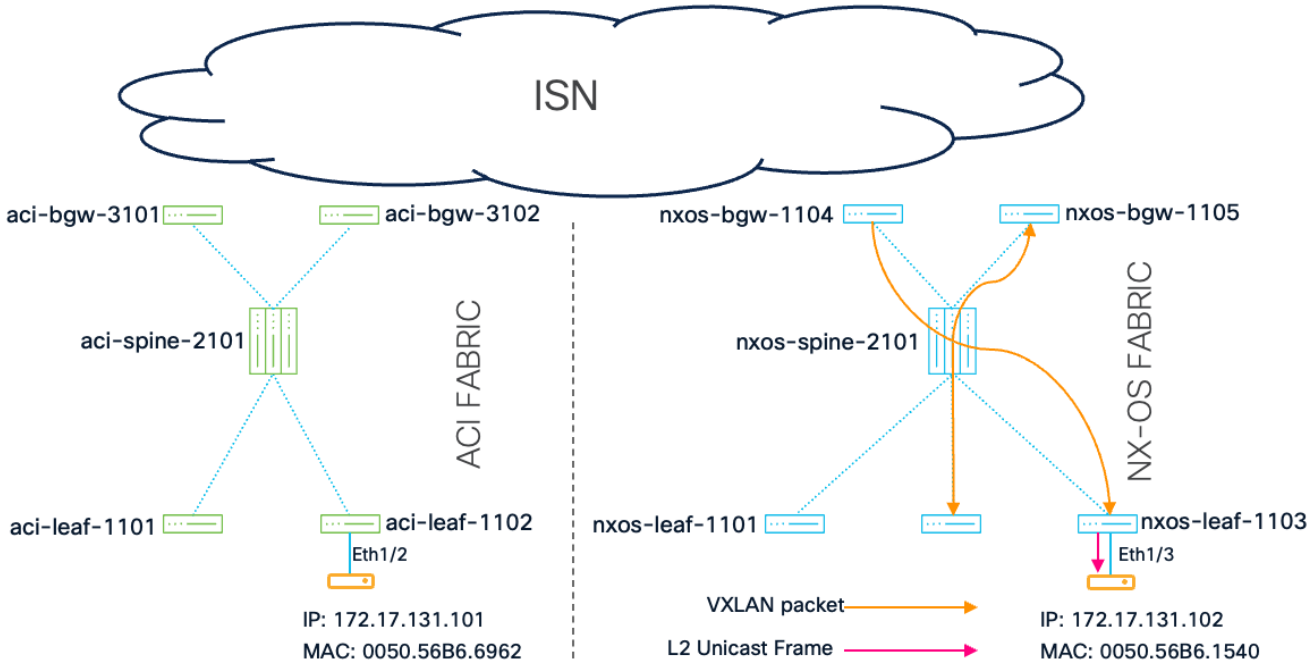

Figure 63 ARP request flooded in the NX-OS fabric

Packet 2: The Unicast ARP Reply

- The NX-OS-attached endpoint generates a unicast ARP reply via a unicast L2 frame. The source will be 0050.56B6.1540 and the destination 0050.56B6.6962. The frame is received by nxos-leaf-1103.
- nxos-leaf-1103 finds the destination MAC address 0050.56B6.6962 in the mac address table; the information was received via EVPN a few milliseconds before when the MAC address was discovered for the first time in the ACI fabric and advertised; the next-hop is the NX-OS site Multi-Site VIP.
- At the same time nxos-leaf-1103 learns about the existence of the MAC address 0050.56B6.1540 and advertises that via a Type-2 MAC-Only EVPN route. The route will also be propagated to the ACI BGWs.
- As the traffic is unicast, the frame is subject to the policy enforcement. But as this is an ARP reply, the frame will hit an implicit rule that will permit it.
- nxos-leaf-1103 encapsulates the packet in a VXLAN and sends it toward the NX-OS fabric Multi-Site VIP:
  - Outer Source IP: 172.20.11.4 (nxos-leaf-1103)
  - Outer Destination IP: 172.20.0.5 (NX-OS Multi-Site VIP)
  - VXLAN VNI: 30000
  - VXLAN SGT: 10002
  - VXLAN Policy Applied Bit: 0

- The packet will be received by either one of the BGWs (say nxos-bgw-1104), based on the underlay ECMP hashing algorithm.

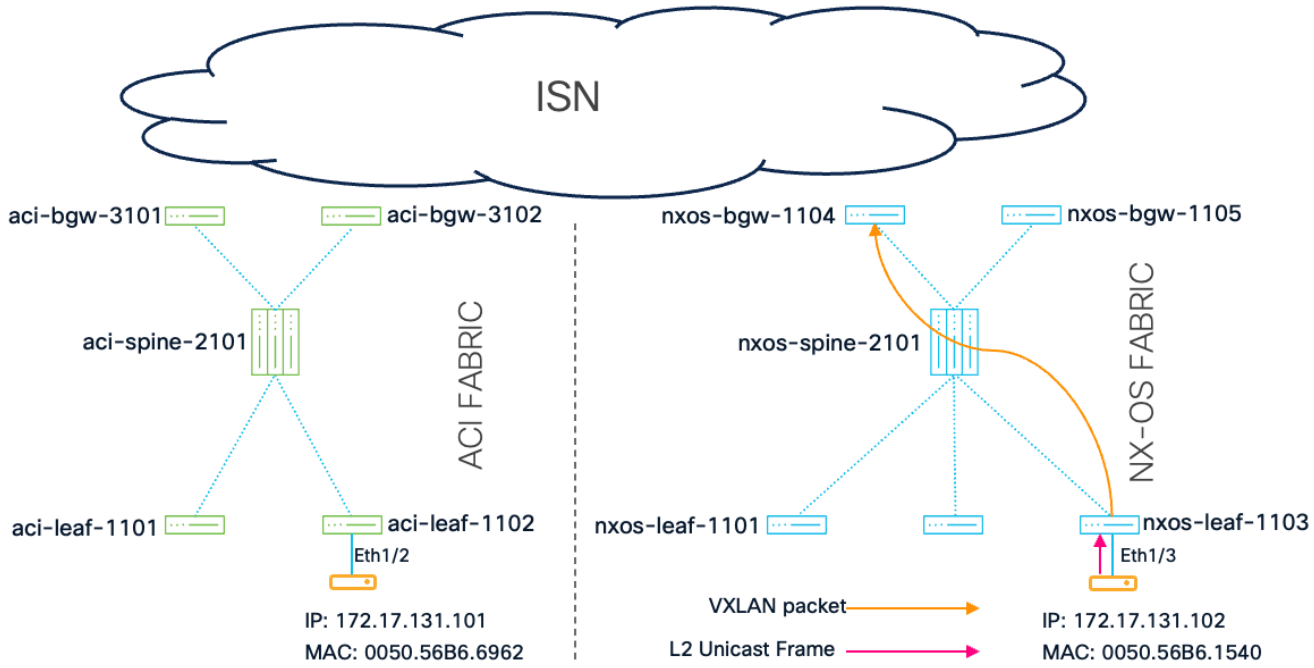

Figure 64 ARP reply forwarded in the NX-OS fabric

- The NX-OS BGW, during the decapsulation, detects that the original packet needs to be bridged in the L2VNI 30000 and looks up the destination MAC address in the local MAC address table associated with that VNI.
- The destination MAC address points to the remote ACI vxlan-evpn-anycast-v4-external-tep 172.20.0.3.
- nxos-bgw-1104 encapsulates the original unicast ARP reply in a new VXLAN packet and sends it in that direction:
  - Outer Source IP: 172.20.11.1 (nxos-bgw-1104)
  - Outer Destination IP: 172.20.0.3 (ACI vxlan-evpn-anycast-v4-external-tep)
  - VXLAN VNI: 30000

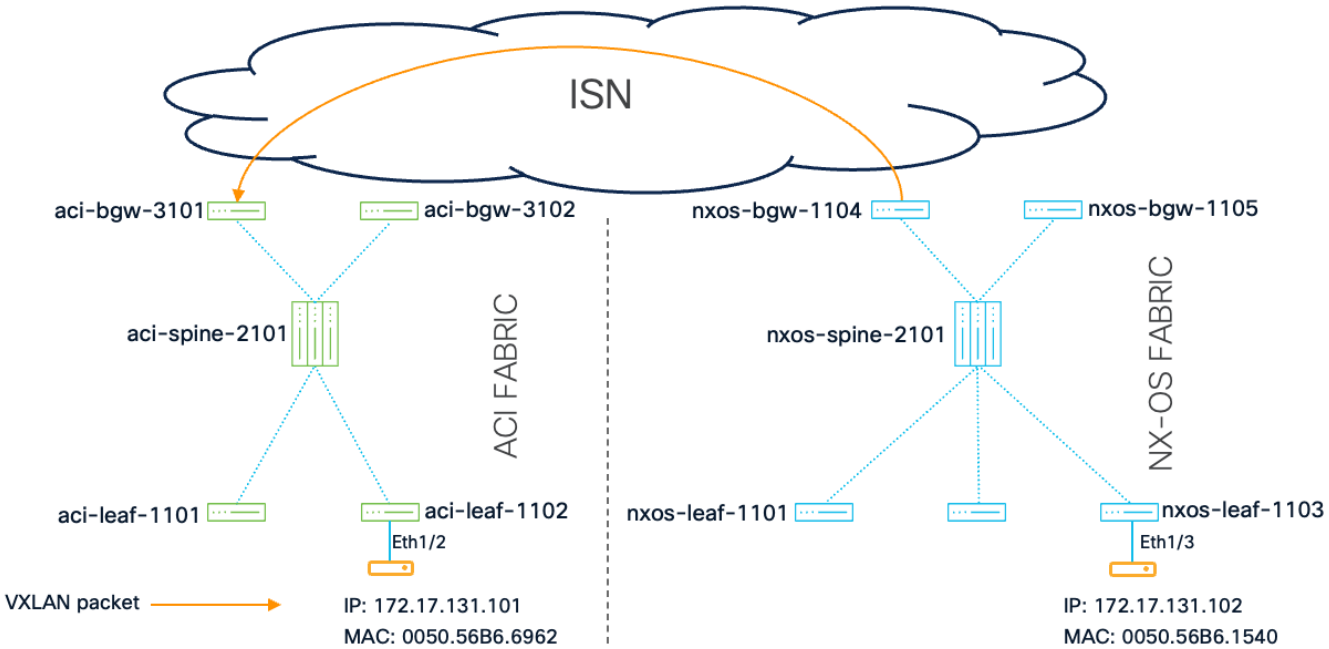

Figure 65 ARP reply forwarded in the Inter-Site network

- Upon receiving the ARP reply, the ACI BGW (in this case aci-bgw-3101) performs the standard tunnel stitching operations and translates the overlay VNI.
- As the local ACI MAC addresses are installed with pcTag 1 on the BGWs, the devices do not apply any policy, encapsulate the frame in an iVXLAN packet, and forward it to the L2 MAC Spine-Proxy.
  - Outer Source IP: 172.21.11.164 (ACI vxlan-evpn-anycast-v4-internal-tep)
  - Outer Destination IP: 172.21.11.161 (L2 MAC Spine-Proxy)
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag
  - iVXLAN Policy Applied: 0
- The spine performs a fast lookup thanks to the COOP entries downloaded in the forwarding table and proxies the iVXLAN packet to the aci-leaf-1102
  - Outer Source IP: 172.21.11.164 (ACI vxlan-evpn-anycast-v4-internal-tep)
  - Outer Destination IP: 172.21.11.130 (aci-leaf-1102)
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag
  - iVXLAN Policy Applied: 0
- aci-leaf-1102 receives the packet, decapsulates the original frame and after a local table lookup it forwards it toward the destination endpoint. The policy implicitly allows the packet as this is a unicast ARP reply.

- At the same time, the leaf performs data-plane learning, mapping the remote MAC address (0050.56b6.1540) to the ACI BGW's vxlan-evpn-anycast-v4-internal-tep 172.21.11.164 and as well the related pcTag.

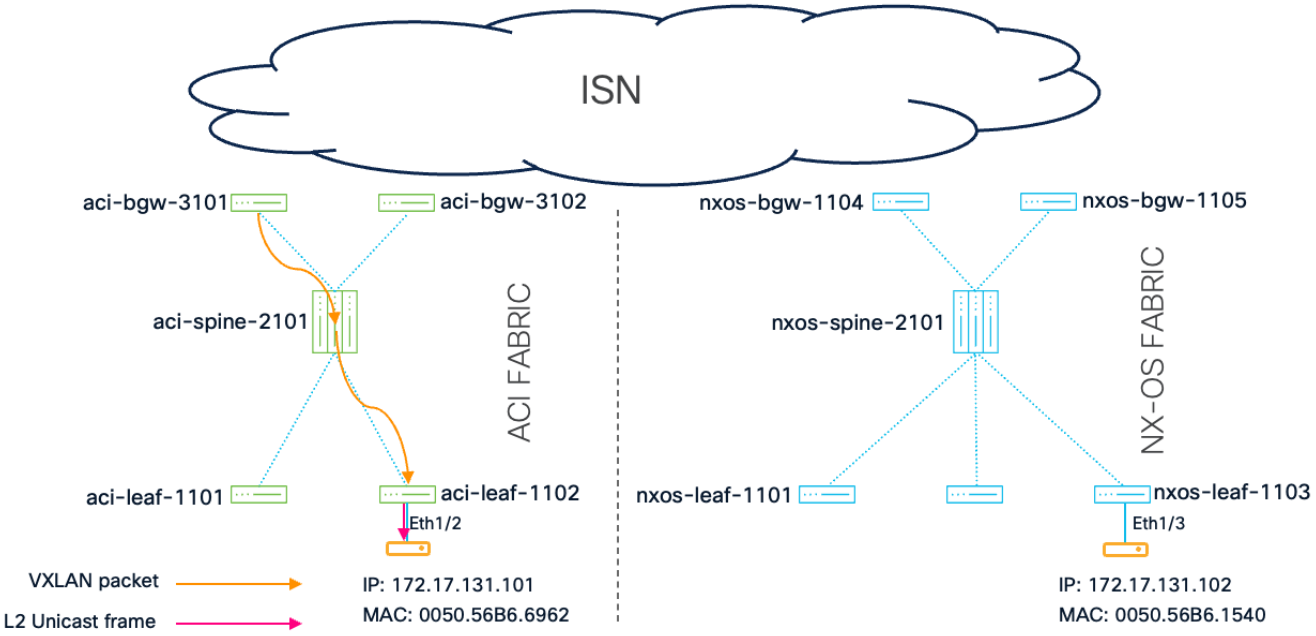

Figure 66 ARP reply forwarded in the ACI Fabric

Packet 3: ICMP Echo

- With the MAC address learned, the ACI-connected endpoint, 172.17.131.101, sends an ICMP packet to 172.17.131.102.
- The aci-leaf-1102 receives the frame. It performs a lookup, finds the remote MAC in its table. The entry contains both the next-hop and a pcTag as they were learned via data plane with the previous unicast ARP reply frame.
- The leaf applies the policy; as the frame remains within the same security-group the traffic is allowed, encapsulated in an iVXLAN and sent to the next-hop
  - Outer Source IP: 172.21.10.130 (aci-leaf-1102).
  - Outer Destination IP: 172.21.11.164 (ACI vxlan-evpn-anycast-v4-internal-tep).
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag
  - iVXLAN Policy Applied: 1
- The BGWs decapsulate the iVXLAN packet and performs a destination MAC lookup. As Policy Applied bit is set to 1, the policy enforcement is skipped.
- The BGW now knows about the remote NX-OS endpoint as that was advertised in the meantime via EVPN, hence it encapsulates the original frame, translates the VNID to the normalized value 30000 and send it to NX-OS Multi-Site VIP address:

- Outer Source IP: 172.20.0.1 (aci-bgw-3101)
- Outer Destination IPs: 172.20.0.5 (NX-OS Multi-Site VIP)
- VXLAN VNI: 30000

- The NX-OS BGW receives the standard VXLAN packet. Because the GPO header is not preserved across the ISN, the NX-OS BGW re-classifies the frame and applies the security enforcement.
- It then forwards the frame to the nxos-leaf-1103 via the VXLAN tunnel:
  - Outer Source IP: 172.20.11.1 (nxos-bgw-1104)
  - Outer Destination IPs: 172.20.11.4 (nxos-leaf-1103)
  - VXLAN VNI: 30000
  - VXLAN SGT: 10002
  - VXLAN Policy Applied Bit: 1
- Nxos-leaf-1103 decapsulates the packet and delivers the original L2 frame to the 172.17.131.102 endpoint.

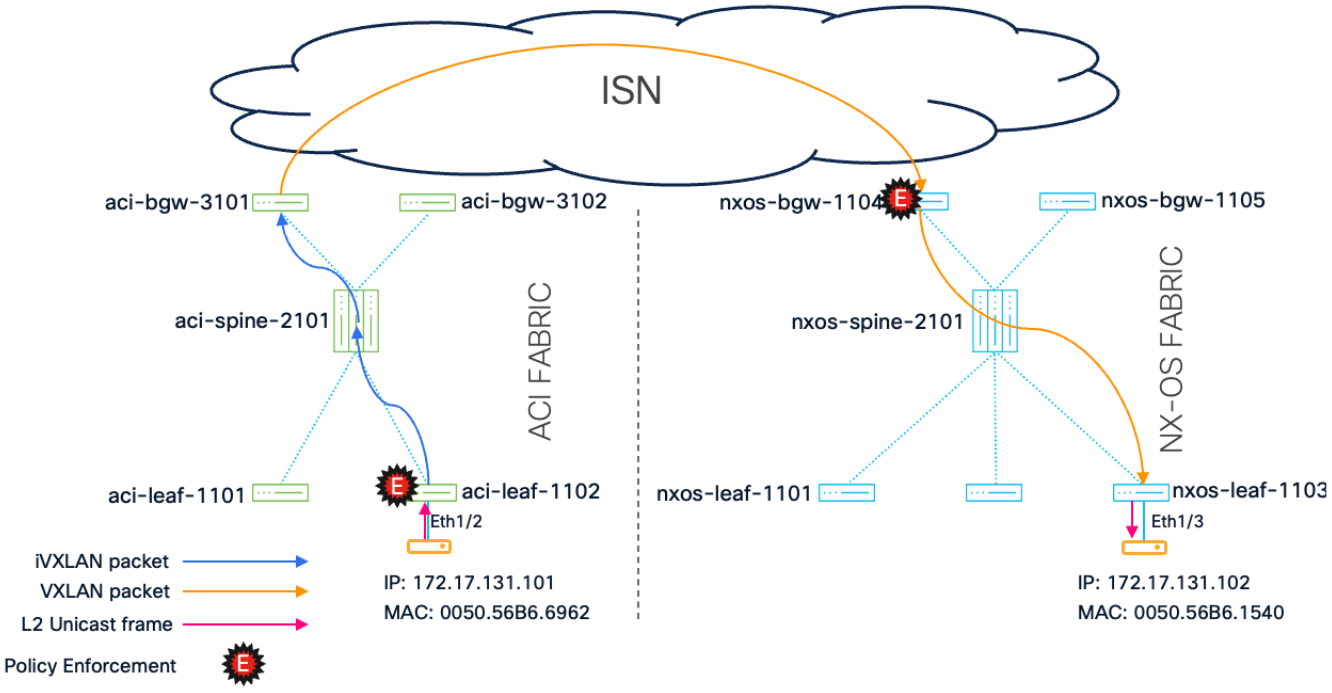

Figure 67 ICMP echo packet from ACI to NX-OS

Packet 4: ICMP Echo Reply (Return Flow)

- The NX-OS-connected endpoint (172.17.131.102) sends the ICMP Echo Reply back to the ACI-connected endpoint (172.17.131.101).
- The nxos-leaf-1103 receives the frame. It performs a lookup, finds the destination MAC address (0050.56B6.6962) in the local table. The entry contains the NX-OS Multi-Site VIP as the next-hop.

- The leaf applies the policy based on the source and destination mac address lookup; the traffic is allowed, encapsulated in a standard VXLAN, and sent to the NX-OS BGW Multi-Site VIP:
  - Outer Source IP: 172.20.11.4 (nxos-leaf-1103)
  - Outer Destination IP: 172.20.0.5 (NX-OS Multi-Site VIP)
  - VXLAN VNI: 30000
  - VXLAN SGT: 10002
  - VXLAN Policy Applied Bit: 1
- One of the NX-OS BGWs receives the VXLAN packet, decapsulates the content, and performs a destination MAC lookup. Since the destination MAC is associated with the remote ACI fabric, it encapsulates the original frame in a new VXLAN packet directed to the ACI vxlan-evpn-anycast-v4-external-tep:
  - Outer Source IP: 172.20.11.1 (nxos-bgw-1104)
  - Outer Destination IP: 172.20.0.3 (ACI vxlan-evpn-anycast-v4-external-tep)
  - VXLAN VNI: 30000
- The ACI BGW that receives the packet performs the standard tunnel stitching operations. It translates the VNID from the normalized value 30000 back to the ACI local L2VNI 15171533 and maps the source of the traffic to the ACI local ESG pcTag 5474.
- No policy is applied as local ACI endpoint MAC addresses are installed with pcTag 1. Due to this the Policy Applied bit will be set to 0 and the egress leaf will apply the policy.
- The BGW encapsulates the original frame in an iVXLAN packet and forwards it to the L2 Spine Proxy:
  - Outer Source IP: 172.21.11.164 (ACI vxlan-evpn-anycast-v4-internal-tep)
  - Outer Destination IP: 172.21.11.161 (L2 MAC Spine-Proxy)
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag)
  - iVXLAN Policy Applied: 0
- The spine performs a fast lookup thanks to the COOP entries downloaded in the forwarding table and proxies the iVXLAN packet to the aci-leaf-1102
  - Outer Source IP: 172.21.11.164 (ACI vxlan-evpn-anycast-v4-internal-tep)
  - Outer Destination IP: 172.21.11.130 (aci-leaf-1102)
  - iVXLAN VNI: 15171533 (ACI local L2VNI)
  - iVXLAN pcTag: 5474 (ACI local ESG pcTag
  - iVXLAN Policy Applied: 0
- The aci-leaf-1102 receives the packet, decapsulates it, and looks up the destination. As the Policy Applied bit is set to 0, the leaf applies the policy.

- Finally, as the policy allows the intra-ESG traffic aci-leaf-1102 delivers the original ICMP echo reply to the ACI endpoint 172.17.131.101.

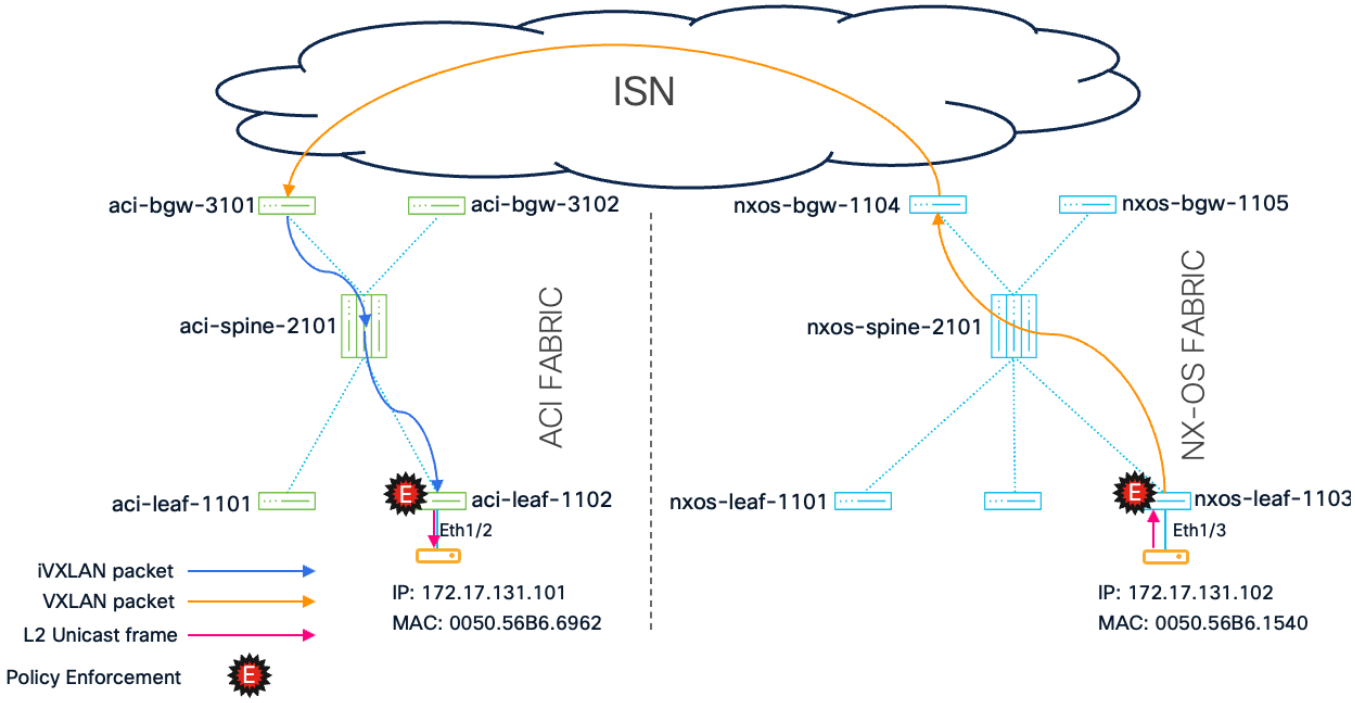

Figure 68 ICMP reply packet from NX-OS to ACI

Note: An additional white paper containing a list of more exhaustive packet walk scenarios is expected to be published separately.

Service Redirection

The redirect action in contracts (ACI) and SGACLs (NX-OS) enables flexible, policy-based service chaining. Traffic flowing between two security groups can be transparently steered through one or more service functions, such as firewalls for deep packet inspection, load balancers for traffic distribution, or Intrusion Prevention Systems (IPS). This powerful feature decouples service insertion from the physical network topology, allowing security architects to insert services based on business intent (e.g., "Web to App traffic must be inspected") rather than IP addressing or VLAN location.

In Nexus One, while the intent is defined globally via Security Groups, the implementation remains native to each fabric's capabilities:

- ACI: utilizes Service Graphs to define the service chain.
- NX-OS: utilizes Enhanced Policy-Based Redirection (ePBR).

A critical architectural detail in the Nexus One context is the handling of policy state across the Inter-Site Network (ISN), as the policy information in the VXLAN GPO header is not preserved when traffic traverses the ISN.

Consequently, the destination fabric receives the traffic as "unenforced" and will re-evaluate the packet against its local policies. This results in a sequential service chaining model:

- Traffic is classified and redirected through the service chain in the source fabric (e.g., ACI Service Graph).
- Traffic traverses the ISN.
- Traffic is re-classified and, if a corresponding redirect policy exists at the destination, is redirected through the service chain in the destination fabric (e.g., NX-OS ePBR).

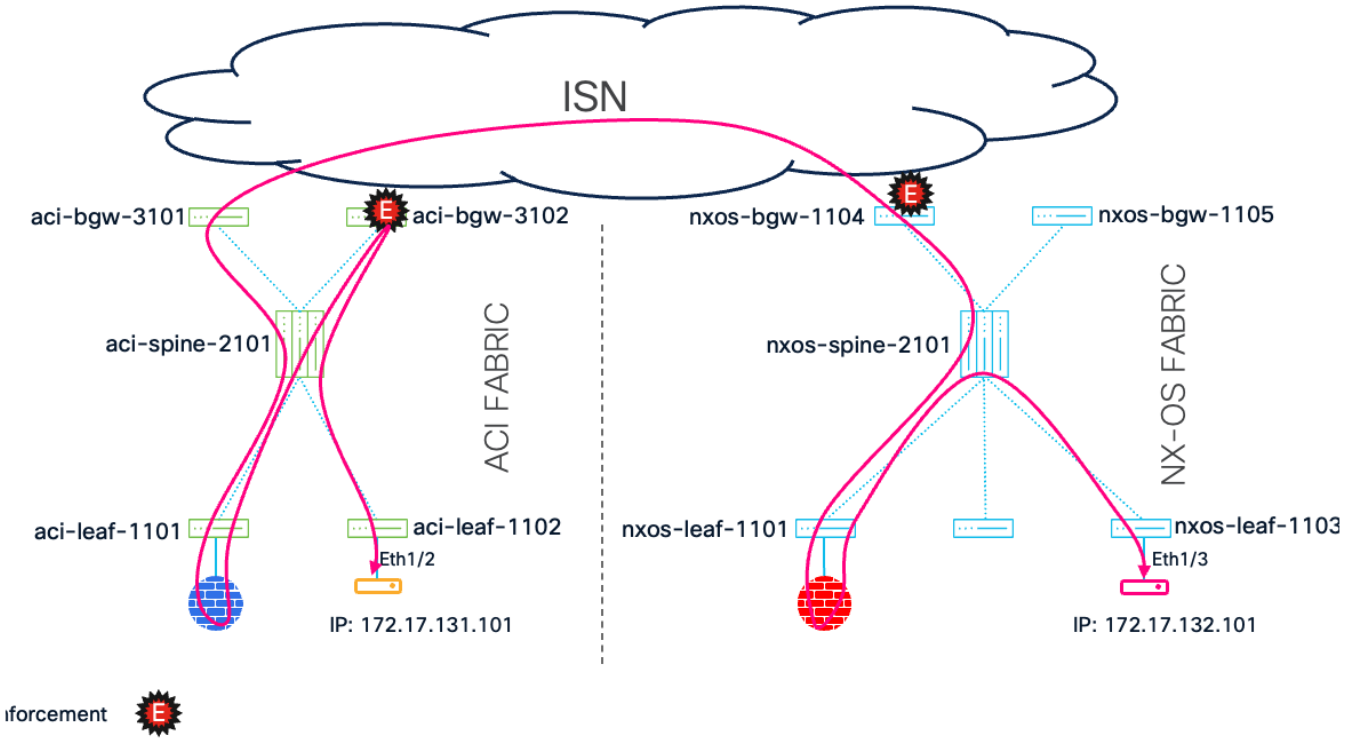

Figure 69 Double PBR Effect

Operators must plan for this "double inspection" behavior when deploying symmetric policies across sites, ensuring that service chains are designed to accommodate traffic that has already been inspected by the remote site.

Due to the above, it is important to note that the high availability of each service chain must be maintained within each fabric. In the current implementation, it is not possible to redirect intra-site traffic to a remote-site service chain when the local one fails.

ACI Multi-Site With BGWs

Over the years, Cisco ACI has continuously evolved to address the requirements of modern data centers, particularly when it comes to interconnecting geographically dispersed sites and extending the fabric beyond the confines of a single spine and leaf topology. Before diving into the specifics of how the ACI Border Gateway introduces a new approach to building Multi-Site architectures, it is worth taking a moment to revisit the journey that has brought ACI to this point. Understanding the strengths and trade-offs of the existing ACI multi-fabric solutions is essential to appreciate the architectural value that the BGW-based approach brings to the table.

The Evolution of ACI Multi-Fabric Architectures

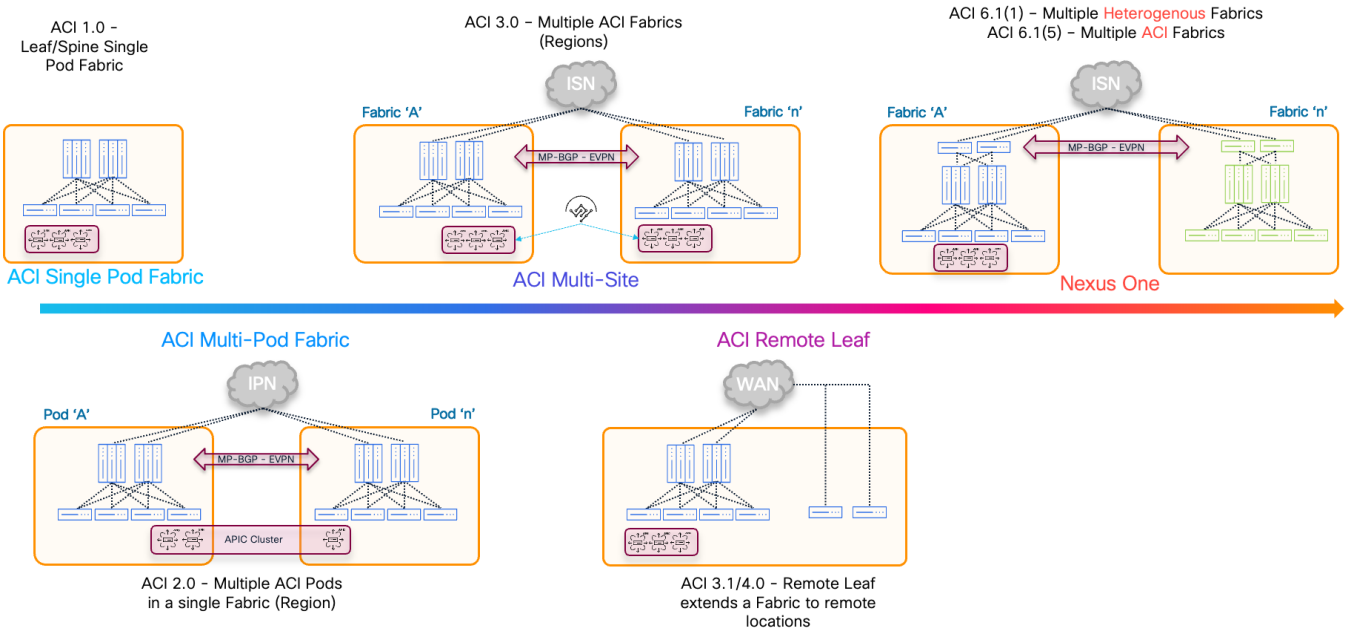

Figure 70 ACI DCI Evolution

The first meaningful step in extending ACI beyond a single fabric was the introduction of ACI Multi-Pod. Multi-Pod was designed to address scenarios where multiple data center locations, typically within a metropolitan area or campus, needed to be operated as a single logical ACI fabric. In this model, all pods share a common APIC cluster, hence a single management domain. The pods are interconnected through an Inter-Pod Network (IPN), and while the data plane is distributed, the fabric behaves as one. Multi-Pod is particularly well-suited for environments where low latency is a given, and where the operational simplicity of managing a single fabric outweighs the need for strict fault isolation between sites. However, because the management plane is shared across pods, a misconfiguration or software issue in one pod can potentially have an impact on the others, limiting the blast radius containment that some customers require.

To address this limitation, Cisco introduced ACI Multi-Site, a solution that takes a fundamentally different approach. Instead of stretching a single fabric, Multi-Site interconnects multiple, fully independent ACI fabrics, each with its own APIC cluster, control plane, and management plane. This separation provides the strong fault isolation that large-scale and geographically dispersed deployments demand. Centralized orchestration is provided today by the Orchestration service on Nexus Dashboard, which evolved from the earlier Multi-Site Orchestrator (MSO) and later Nexus Dashboard Orchestrator (NDO). This orchestration layer acts as a single pane of glass for defining tenants, VRFs, bridge domains, EPGs, ESGs, contracts, and

more and then pushing the appropriate configurations to each fabric. The orchestration layer also takes care of building the inter-site connectivity, programming the spines to talk to each other and establishing the necessary overlay tunnels between sites. Multi-Site has become the de-facto standard for interconnecting production ACI fabrics, but it does come with the requirement of deploying and operating the additional orchestration layer.

In parallel, Cisco introduced the Remote Leaf architecture to address a different class of use cases: extending the ACI fabric to smaller satellite locations, such as branch offices, edge sites, or co-location facilities, without deploying a full spine-leaf fabric at each remote site. Remote leaves are leaf switches physically located at a remote site but logically part of the main ACI fabric, connected back through an IP network. This model is particularly valuable for customers who want to extend consistent ACI policy and automation to smaller footprints without the overhead of a full fabric deployment.

Each of these architectures has its place, and together they have provided ACI customers with a rich set of options for building distributed data center environments.

The BGW-based Multi-Site approach described in this appendix does not replace any of them; rather, it adds another tool to the toolbox, one that leverages the standards-based VXLAN EVPN foundation of Nexus One to enable a more flexible and open way of interconnecting ACI fabrics.

A New Approach: Multi-Site via the ACI Border Gateway

With the introduction of the ACI Border Gateway in ACI release 6.1(1), and the continued maturity of the feature in subsequent releases, a new possibility emerges: building a Multi-Site architecture between two or more ACI fabrics using the BGW as the interconnection point, rather than relying exclusively on the traditional spine-to-spine connectivity orchestrated by the Orchestration service on ND. Because the BGW speaks standard MP-BGP EVPN and performs standards-based VXLAN encapsulation, it can naturally peer with BGWs deployed in other ACI fabrics, creating an EVPN-based overlay that extends Layer 2 and Layer 3 connectivity, as well as policy, across sites.

From a high-level perspective, much of what has already been discussed in the main sections of this document applies directly to this scenario. The BGW performs the same tunnel stitching, namespace normalization, and policy translation functions, regardless of whether the remote fabric is an NX-OS VXLAN EVPN fabric or another ACI fabric. The control plane mechanics, the use of anycast VIPs for resiliency, the hierarchical overlay model, and the role of the BGW as a control point for BUM traffic all remain consistent. What changes is simply the nature of the peer on the other side of the Inter-Site Network: instead of an NX-OS BGW, we now have another ACI BGW, with its own APIC cluster behind it.

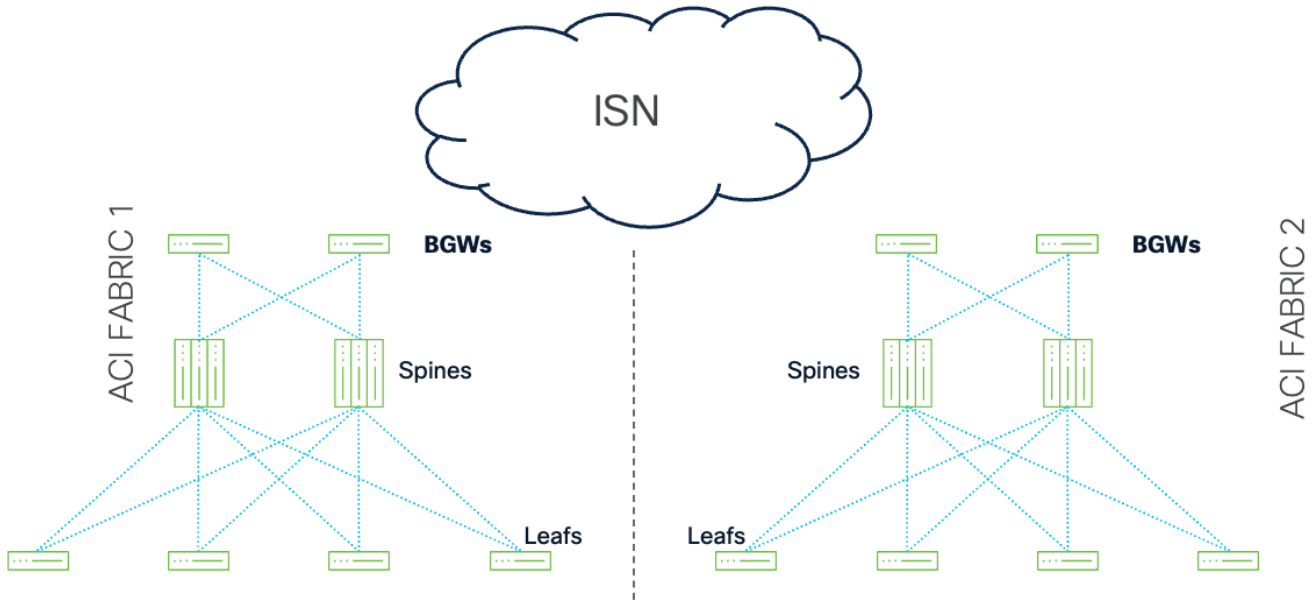

Figure 71 ACI BGW-based Multi-Site

One of the most important aspects of this new approach, and one that deserves particular attention, is the operational model. At the time of this writing, customers who wish to deploy a BGW-based ACI-to-ACI Multi-Site architecture are required to configure each APIC cluster independently. There is no single orchestration entity that, today, programs all fabrics from a unified intent. This means that each fabric must be provisioned separately, either through direct manual configuration on each APIC or through automation workflows built by the customer (for example, using Terraform, Ansible, or custom scripts against the APIC REST API).

This is a significant architectural distinction compared to the traditional Multi-Site model. With NDO, the orchestrator is a hard requirement for building and operating the Multi-Site domain. With the BGW-based approach, no orchestrator is required. Customers are free to choose their own operational model: some may prefer the flexibility and control of managing each fabric independently, while others may want a centralized orchestration experience. For the latter, Cisco will introduce, starting with Nexus Dashboard release 4.4, the concept of ACI-ACI Fabric Groups. With this upcoming functionality, Nexus Dashboard will take care of building all the required underlay and overlay connectivity policies on the participating ACI fabrics and on the ISN fabric, providing that it's under its administrative control. ND will also deploy the required overlay constructs, such as VRFs, networks, security groups, contracts, and filters, across all the fabrics that are part of the group. Administrators will be able to define their intent once and have Nexus Dashboard translate and deploy the appropriate configuration to each APIC cluster. Importantly, the use of Nexus Dashboard in this context is optional: customers who prefer to operate each fabric independently can continue to do so, while those who value centralized orchestration will have a native path to it.

Extending VRFs, Networks, and Security Groups Across Fabrics

A natural question that arises when designing a BGW-based ACI-to-ACI Multi-Site deployment is how VRFs, bridge domains, and ESGs are extended across fabrics, especially considering that each APIC cluster independently allocates its own VNIs and pcTags to these constructs.

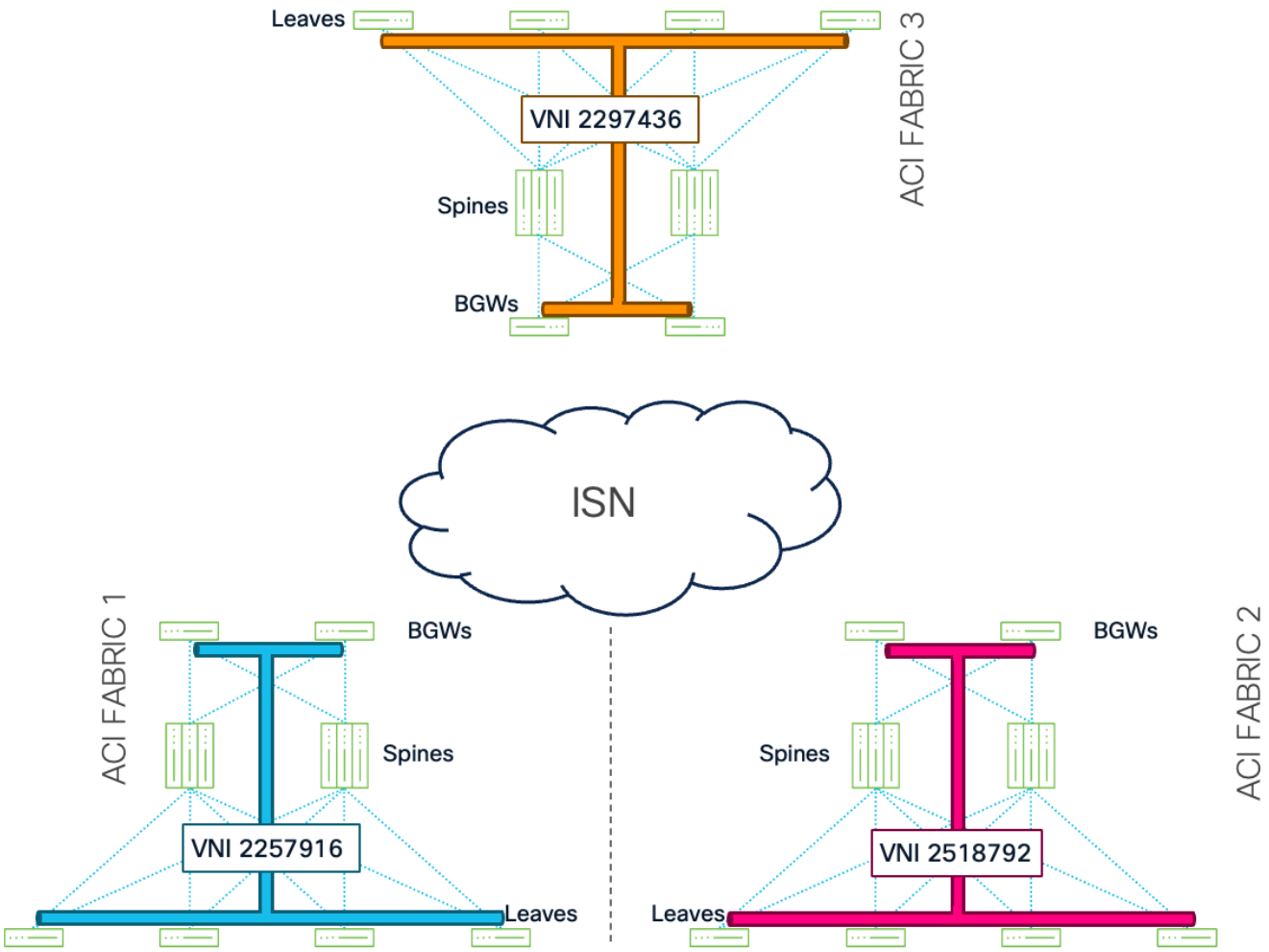

Figure 72 Asymmetric VNIs

The answer lies in the namespace normalization capability of the BGW, which was discussed in detail earlier in this document. In the context of ACI-to-ACI Multi-Site, this capability becomes particularly relevant, because namespace mismatches are effectively guaranteed: there is no mechanism that forces two independent APIC clusters to assign the same VNI to the same logical resource.

So, what should an administrator do when the same VRF or Bridge Domain, deployed independently in three different sites, needs to be extended across this architecture?

The recommended design approach is straightforward. For each resource that needs to be extended across sites, such as a VRF or a bridge domain, the administrator chooses a global, common identifier that will be used as the normalized value across all participating fabrics. This global identifier does not need to match any of the locally assigned VNIs; it simply needs to be unique and consistent across the Multi-Site domain. The BGW in each fabric is then configured with the appropriate normalization entries, mapping the local VNI or pcTag to the chosen global value in both control and data plane.

To make this more concrete, consider a VRF that needs to be extended across three ACI fabrics. ACI Fabric 1 may have assigned VNI 2257916 to this VRF, ACI Fabric 2 may have assigned VNI 2518792, and ACI Fabric 3 may have assigned VNI 2297436. Rather than attempting to force all three fabrics to use the same local VNI, which is impossible, the administrator picks a global value, for example 464646, and

configures each BGW to translate between its local VNI and this global value. When a route is advertised from Fabric 1 toward the other fabrics, the local VNI 2257916 is translated to 464646 before leaving the BGW. When the same route is received by Fabric 2, the BGW in Fabric 2 translates the global 464646 back into its local 2518792. The same logic applies in the data plane: VXLAN packets traversing the Inter-Site Network carry the global VNI, and each BGW performs the appropriate translation as packets enter and leave its local fabric.

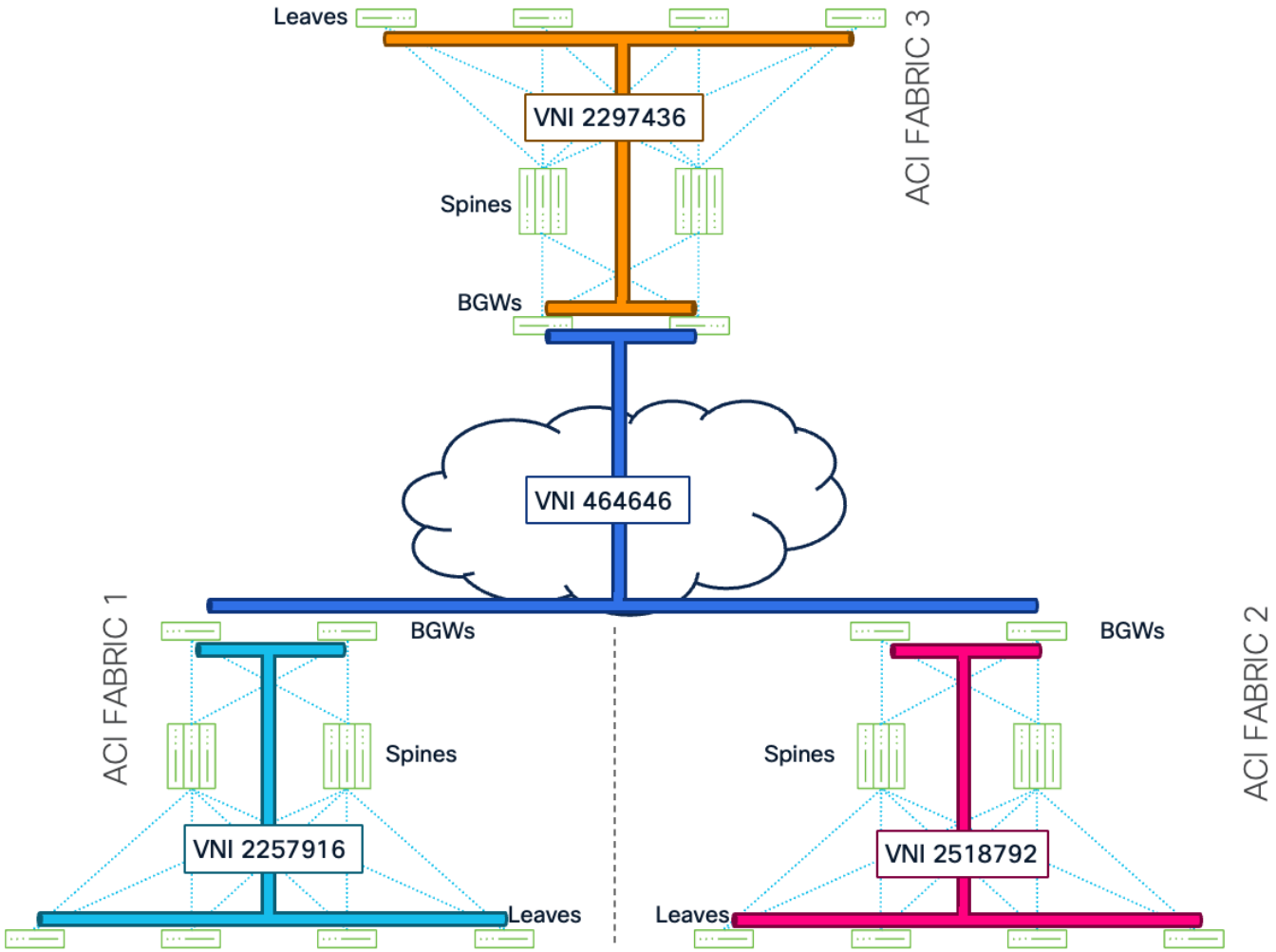

Figure 73 Two-way translation

The same concept applies to ESGs. Each APIC cluster assigns its own local pcTag to an ESG, but for the Multi-Site domain to operate consistently, administrators select a global pcTag that represents the same logical security group across all fabrics. The BGW then ensures that when EVPN routes carrying security classification information are exchanged between sites, the local pcTag is normalized to the global value and vice-versa. This is what makes it possible to define a contract between, for example, a "Web Tier" ESG and a "Database Tier" ESG once, and have it applied consistently across all fabrics, although the underlying pcTags are different in each site.

This global-VNI and global-pcTag approach is simple to understand, easy to document, and scales naturally as additional fabrics are added to the Multi-Site domain. When Nexus Dashboard 4.4 becomes available with ACI-ACI Fabric Groups, this translation logic will be fully automated, with Nexus Dashboard proposing and managing the global identifiers transparently on behalf of the administrator.

Requirements

Deploying a BGW-based ACI-to-ACI Multi-Site architecture builds on top of the same foundational requirements already discussed for the Nexus One architecture. From a software perspective, ACI release 6.1(5)M provides a stable baseline where this design can be confidently implemented in production environments. As the first maintenance release of the 6.1 train, it brings the level of maturity and stability expected for deployments of this scope, and it includes all the BGW functionality, including the asymmetric namespace normalization, that is relevant for ACI-to-ACI Multi-Site scenarios.

The platform requirements for the BGW role remain the same as described in the main sections of this document. The BGW function is supported on Cisco Nexus 9300-FX, -FX3, -GX, -GX2, and -HX series switches, provided they are equipped with a minimum of 32 GB of RAM. FX2 devices do not meet this memory requirement and therefore cannot function as BGWs. BGWs must be deployed as dedicated leaf nodes connected to the ACI spines, and a minimum of two BGWs per site (or per pod, in Multi-Pod deployments) is required for redundancy. In Multi-Pod fabrics, BGWs must be deployed in each pod that requires connectivity to the remote ACI fabrics.

Licensing-wise, as with any other Nexus One deployment, all nodes in the ACI fabric must be licensed with the Advantage tier. From a policy perspective, ESGs are mandatory across all fabrics participating in the Multi-Site domain. Customers still relying on EPGs as their primary security classification construct will need to migrate to ESGs before adopting this architecture. As mentioned earlier in the document, Cisco has released a script to assist customers in uplifting their existing EPG-based configurations into ESGs, simplifying this transition.

Caveats and Limitations

As with any architectural option, the BGW-based ACI-to-ACI Multi-Site approach comes with a set of caveats and limitations that administrators must carefully evaluate before committing to this design.

The first and perhaps most important limitation is that Tenant Routed Multicast (TRM) is not currently supported in this architecture. Customers with applications that depend on inter-site multicast routing will need to stick with the traditional NDO orchestrated Multi-Site, where TRM is supported.

The second area that deserves careful attention is scalability. The BGW-based Multi-Site architecture currently supports lower scale numbers compared to traditional ND orchestrated Multi-Site, and these figures become even more constrained when Multi-Pod is also part of the design. Because scale limits evolve release after release, administrators are strongly encouraged to consult the version-specific verified scale guidelines in the **ACI Border Gateways (BGW)** section of the Cisco ACI Verified Scalability Guide to understand the applicable limits for their deployment.

A third consideration, already mentioned but worth reiterating, is that ESGs are mandatory in any extended, enforced VRF. There is no supported path for using EPGs as the primary security classifier in this architecture. Customers must plan their migration to ESGs as part of the adoption path for BGW-based Multi-Site.

Finally, administrators need to be aware of how policy enforcement, and therefore service redirection behave in this architecture. As discussed in the main sections of the document, service chains remain local to each fabric. A packet that traverses the Inter-Site Network arrives at the destination fabric without

preserved policy state in the VXLAN header and will be re-evaluated and re-redirected through a local service chain at the destination site. This results in the "double inspection" behavior already described. Practically, this means that each fabric must maintain its own independent service appliance cluster: there is no mechanism today to redirect traffic from one site to a service chain deployed in another site, and in the event of a local service chain failure, traffic cannot automatically fall back to a service chain in a remote fabric. Designs that rely on centralized service insertion must plan accordingly, ensuring that each site has its own high-availability service cluster sized for the traffic it needs to inspect. This last consideration is true in the context of classic NDO-based ACI Multi-Site too.

Despite these caveats, the BGW-based ACI-to-ACI Multi-Site architecture represents a meaningful addition to the ACI multi-fabric portfolio. It provides customers with an open, standards-based, and orchestrator-optional path to interconnecting ACI fabrics, while preserving the policy consistency and operational simplicity that have always been at the heart of the ACI value proposition. And because it shares the same foundation as the rest of the Nexus One solution, this architecture is not a silo: customers can seamlessly extend their ACI-to-ACI Multi-Site deployments to interconnect with NX-OS VXLAN EVPN fabrics today, and with the broader Nexus One ecosystem as it continues to evolve, including Hyperfabric, campus, and public cloud domains. For customers who prefer to keep operating each fabric independently, or who want to progressively adopt Nexus Dashboard as their unified control point, this architecture offers the flexibility to do so on their own terms.

Deployment Considerations and Best Practices

A successful implementation of the Nexus One architecture necessitates meticulous planning. This section synthesizes the technical requirements detailed throughout this document into actionable deployment strategies, focusing on the critical dependencies between ACI and NX-OS architectures.

Hardware, Software, and Licensing Prerequisites

Meeting the foundational requirements is non-negotiable for a robust Nexus One deployment.

- Platform: The BGW role is exclusively supported on Cisco Nexus 9300 series switches with Cloud Scale ASICs (FX, FX3, GX, GX2, and HX). A minimum of 32 GB of RAM is mandatory. This is not for throughput, but to accommodate the expanded Control Plane scale required to store remote EVPN routes and SGT mappings.
- The entire ACI fabric requires the Advantage tier license.
- For the ACI fabric, release 6.1(5)M or later is the recommended baseline. As the first maintenance release for the 6.1 train, it offers the necessary stability and maturity required for production deployments of the Border Gateway function.
- For the NX-OS fabric, release 10.6(1)F or later is strongly recommended when Security Groups are also implemented in the NX-OS fabric as this release introduces MAC-based segmentation, an important feature for Nexus One solution that enables policy enforcement on Layer 2 traffic following a consistent behavior with ACI. If Security Groups are not needed and planned in the NX-OS fabric, then any currently supported version would work.

Topology Design and BGW Placement

- As of current ACI releases, BGWs must be dedicated leaf nodes and cannot function as compute leaves or host tenant L3Out connections.
- The ISN must support additional 50 bytes for the VXLAN overhead. Control plane sessions (MP-BGP) are highly sensitive to MTU. If the MTU is inconsistent, the BGP session might establish

(Open/Keepalive messages are small), but fail later when large BGP Update messages containing extensive route data are sent. By default, ACI BGP speakers use 9000 bytes for the control-plane MTU, whereas NX-OS looks at the IP MTU of the interface used to reach the BGP neighbor and uses that. For additional information, review the “Overlay Control Plane” section in this document.

Control Plane and Connectivity Best Practices

- You must configure a full mesh of MP-BGP EVPN sessions between all BGWs across all sites. Each BGW must explicitly peer with every other remote BGW. Support for the Route-Server functionality was introduced in ACI 6.2(2)F and will also be available in the 6.1 train starting from ACI 6.1(6)M.
- Even for purely routed traffic, you must configure the L2VNI on the BGWs for any Bridge Domain part of a VRF that is extended.
- For ACI subnets to be reachable from NX-OS, the Bridge Domain subnet configuration must have the "Advertised Externally" scope enabled.

Security Policy Design

- ACI Micro-Segmentation (uSeg) EPGs are not supported for the stretched BDs in a Nexus One environment.

Operational Considerations and Constraints

- Currently, the VXLAN GPO header is not sent across the Inter-Site Network (ISN) between ACI and NX-OS domains, therefore, traffic is effectively treated as "unenforced" upon arrival at the destination fabric. Consequently, the receiving Border Gateway will always re-lookup the traffic and sometimes re-enforce the policy. Administrators must ensure that security policies are symmetrically configured or compatible on both sides to prevent unintended drops, a task streamlined by Nexus Dashboard, starting from version 4.2.

Conclusion

The modern data center is now a dynamic, distributed collection of interconnected resources. In this landscape, managing and securing these environments as a unified whole is essential. The Cisco Nexus One architecture emerges as the strategic architectural blueprint for this new reality, providing a cohesive framework that transcends the traditional boundaries of individual fabric types.

Leveraging Open NEtworking (ONE) standards it ensures maximum flexibility and interoperability, allowing organizations to integrate existing infrastructure and scale with best-of-breed solutions.

By meticulously unifying connectivity, segmentation, and policy enforcement across Cisco's diverse fabric portfolio encompassing ACI, NX-OS VXLAN EVPN, and with a clear roadmap for Hyperfabric and additional products, Nexus One solution delivers a transformative solution that enables customers to benefit from a seamless, secure, and automated operational model.

The journey to a fully unified, automated, and secure multi-fabric data center is complex and continuous. However, with Nexus One, Cisco provides a clear, strategic, and technically robust path forward, empowering organizations to confidently build and manage the data center of tomorrow.

Acronyms

Acronym	Full form	Description
ACI	Application Centric Infrastructure	Cisco policy-driven data center fabric architecture.
ACL	Access Control List	Rule used to permit or deny traffic based on matching criteria.
APIC	Application Policy Infrastructure Controller	Controller used to manage and configure Cisco ACI fabrics.
API	Application Programming Interface	Interface used by software systems or automation tools to interact with a platform.
ARP	Address Resolution Protocol	Protocol used to resolve an IP address to a MAC address within a Layer 2 domain.
AS	Autonomous System	Routing domain used by BGP.
AS-Path	Autonomous System Path	BGP attribute that lists the AS sequence a route has traversed.
BD	Bridge Domain	ACI Layer 2 forwarding construct
BGP	Border Gateway Protocol	Routing protocol used to exchange reachability information between network devices.
BGW	Border Gateway	VXLAN EVPN gateway used to interconnect fabrics and exchange reachability information.

Acronym	Full form	Description
BUM	Broadcast, Unknown Unicast, and Multicast	Traffic type that may require flooding or replication across a Layer 2 domain.
CNI	Container Network Interface	Interface framework commonly used to provide networking for containers.
COOP	Council of Oracle Protocol	ACI control-plane protocol used to store and distribute endpoint reachability information.
DAG	Distributed Anycast Gateway	Default gateway model where the same gateway IP/MAC can exist on multiple fabric nodes.
DCI	Data Center Interconnect	Technology or design used to interconnect data centers.
DF	Designated Forwarder	Elected forwarding node used to prevent duplicate forwarding or loops for certain traffic types.
DNS	Domain Name System	System used to resolve names to IP addresses.
DR	Disaster Recovery	Design or process used to restore services after a site or system failure.
eBGP	External Border Gateway Protocol	BGP session established between different autonomous systems.
ePBR	Enhanced Policy-Based Redirection	NX-OS feature used to redirect traffic to service devices based on policy.
EP	Endpoint	Host, VM, bare-metal server, container, or other resource connected to the fabric.
EPG	Endpoint Group	ACI policy construct used to group endpoints.
ESG	Endpoint Security Group	ACI security construct used to classify endpoints or prefixes for policy enforcement.
EVPN	Ethernet VPN	BGP-based control-plane technology used also with VXLAN overlays.
FTAG	Fabric tag / flooding tree tag	ACI fabric mechanism associated with BUM replication trees.
GARP	Gratuitous ARP	ARP message sent by an endpoint to announce or update its own IP-to-MAC mapping.
GPO	Group Policy Option	VXLAN and EVPN extension used to carry policy information such as Security Group Tags.

Acronym	Full form	Description
GUI	Graphical User Interface	Visual interface used to operate or configure a system.
iBGP	Internal Border Gateway Protocol	BGP session established within the same autonomous system.
IaC	Infrastructure as Code	Operational model where infrastructure is provisioned and managed through code or automation.
IGP	Interior Gateway Protocol	Routing protocol used within an autonomous system.
IMET	Inclusive Multicast Ethernet Tag	EVPN Type-3 route used to signal BUM replication membership.
IP	Internet Protocol	Network-layer protocol used for addressing and routing packets.
IPS	Intrusion Prevention System	Security system used to detect and block malicious traffic.
IS-IS	Intermediate System to Intermediate System	Link-state routing protocol used in the underlay.
ISN	Inter-Site Network	Network that connects fabrics or sites together.
L2	Layer 2	Data-link layer; commonly associated with switching and MAC forwarding.
L3	Layer 3	Network layer; commonly associated with routing and IP forwarding.
L2VNI	Layer 2 VXLAN Network Identifier	VXLAN identifier used for a Layer 2 segment or bridge domain.
L3VNI	Layer 3 VXLAN Network Identifier	VXLAN identifier used for a routed VRF overlay.
L3Out	Layer 3 Outside	ACI and Nexus Dashboard construct used to connect the fabric to external routed networks.
MAC	Media Access Control	Layer 2 address used for Ethernet forwarding.
MED	Multi-Exit Discriminator	BGP attribute used to influence path selection.
MP-BGP	Multiprotocol Border Gateway Protocol	BGP extension used to carry multiple address families, including EVPN and VPNv4/VPNv6.
MSS	Maximum Segment Size	TCP parameter that defines the maximum TCP payload size.

Acronym	Full form	Description
MTU	Maximum Transmission Unit	Maximum packet size supported on a link or path.
MTTI	Mean Time to Innocence	Operational metric referring to the time required to prove a component is not the cause of an issue.
MTTR	Mean Time to Resolution	Operational metric referring to the time required to resolve an issue.
ND	Nexus Dashboard	Cisco platform used for centralized operations, visibility, and orchestration.
NDO	Nexus Dashboard Orchestrator	Nexus Dashboard service used to orchestrate multi-site policy and connectivity.
NLRI	Network Layer Reachability Information	BGP route information carried in an update.
OSPF	Open Shortest Path First	Link-state routing protocol sometimes shown in routing outputs.
OTV	Overlay Transport Virtualization	Data center interconnect technology used to extend Layer 2 networks.
PA bit	Policy Applied bit	VXLAN GPO header bit indicating whether policy has already been enforced.
PBR	Policy-Based Redirection	Policy mechanism used to redirect traffic through a service device.
pcTag	Policy Control Tag	ACI security/policy identifier associated with an EPG or ESG
PIP	Primary IP	NX-OS BGW loopback address used as source, and sometimes destination, for VXLAN packets.
PMSI	Provider Multicast Service Interface	BGP/EVPN tunnel attribute used in multicast or replication signaling.
PTEP / ptep	Physical Tunnel Endpoint	ACI loopback used for internal fabric reachability and, depending on flow, as a VXLAN tunnel endpoint.
PVLAN	Private VLAN	VLAN feature used for traffic isolation within the same broadcast domain.
RD	Route Distinguisher	BGP VPN/EVPN value used to make routes unique.
REST	Representational State Transfer	API architectural style commonly used for web-based APIs.
RESTCONF	RESTCONF Protocol	REST-based protocol used to configure and operate network devices.

Acronym	Full form	Description
RID	Router ID	NX-OS BGW loopback used to establish BGP control-plane adjacencies.
RT	Route Target	BGP extended community used to control route import and export.
RTEP / rtep	Remote Tunnel Endpoint	ACI BGW loopback used for inter-site reachability and signaling.
SG	Security Group	Logical group of endpoints or prefixes sharing the same security posture.
SGACL	Security Group Access Control List	NX-OS security contract that defines permit, deny, or redirect actions between Security Groups.
SGT	Security Group Tag	Standards 16-bit identifier associated with a Security Group.
SIEM	Security Information and Event Management	Security platform used to collect, correlate, and analyze security events.
SOAR	Security Orchestration, Automation, and Response	Security platform used to automate and coordinate incident response workflows.
STP	Spanning Tree Protocol	Layer 2 loop-prevention protocol.
SVI	Switched Virtual Interface	Logical Layer 3 interface associated with a VLAN or bridge domain.
TCO	Total Cost of Ownership	Overall cost associated with owning and operating a system.
TCP	Transmission Control Protocol	Transport-layer protocol used for reliable communication.
TEP	Tunnel Endpoint	Device or logical interface that encapsulates and decapsulates VXLAN traffic.
uSeg	Micro-segmentation	ACI shorthand for micro-segmentation capabilities.
VIP	Virtual IP	Shared IP address used by multiple devices for resiliency or anycast behavior.
VLAN	Virtual Local Area Network	Layer 2 segmentation construct.
VM	Virtual Machine	Software-based compute instance.
VMM	Virtual Machine Manager	Platform or integration used to manage virtualized workloads.

Acronym	Full form	Description
VNI	VXLAN Network Identifier	VXLAN segment identifier carried in the VXLAN header.
VNID	Virtual Network Identifier	Identifier used for virtual network segments; often used in the document alongside VNI.
VPNv4	VPN IPv4 address family	MP-BGP address family used to carry IPv4 VPN routes.
VPNv6	VPN IPv6 address family	MP-BGP address family used to carry IPv6 VPN routes.
VRF	Virtual Routing and Forwarding	Logical routing table that provides traffic isolation.
VRF-Lite	Virtual Routing and Forwarding Lite	VRF-based routing isolation without VPN services.
VTEP	VXLAN Tunnel Endpoint	Device that encapsulates and decapsulates VXLAN traffic.
VXLAN	Virtual Extensible LAN	Overlay encapsulation technology used to extend Layer 2 and Layer 3 segments over an IP network.

References

- EVPN Multi-Site Draft -draft-sharma-bess-multi-site-evpn-05
  - https://www.ietf.org/archive/id/draft-sharma-bess-multi-site-evpn-05.txt
- IETF Group Policy Option (GPO) Draft
  - https://www.ietf.org/archive/id/draft-lrss-bess-evpn-group-policy-02.txt
- Cisco White Paper: Securing Data Centers with Microsegmentation using VXLAN GPO
  - https://www.cisco.com/c/en/us/td/docs/dcn/whitepapers/securing-datacenters-with-microsegmentation-and-vxlan-gpo.html
- Cisco Configuration Guide: NX-OS VXLAN Configuration Guide
  - https://www.cisco.com/c/en/us/td/docs/dcn/nx-os/nexus9000/106x/configuration/vxlan/cisco-nexus-9000-series-nx-os-vxlan-configuration-guide-release-106x/microsegmentation_for_vxlan_fabrics_using_gpo.html
- Cisco Configuration Guide: ACI Border Gateways
  - https://www.cisco.com/c/en/us/td/docs/dcn/aci/apic/6x/l3-configuration/cisco-apic-layer-3-networking-configuration-guide-61x/vxlan-layer3-config-61x.html