White Paper

Cisco Application Centric Infrastructure and Cisco Application Virtual Switch

Introduction

The Cisco® Application Virtual Switch (AVS) is a hypervisor-resident virtual network switch that is specifically designed for the Cisco Application Centric Infrastructure (ACI) architecture. Based on the highly successful Cisco Nexus® 1000V Switch, Cisco AVS provides feature support for the Cisco ACI application policy model, full switching capabilities, and more advanced telemetry features.

Compared to other hypervisor-based virtual switches, Cisco AVS provides cross-consistency in features, management, and control through the Cisco Application Policy Infrastructure Controller (APIC), rather than through hypervisor-specific management stations. As a main component of the overall Cisco ACI framework, Cisco AVS enables intelligent policy enforcement and optimal traffic steering for virtual applications.

Main features include:

- Purpose-built, virtual network edge for Cisco ACI fabric architecture
- Cisco ACI full policy enforcement on Cisco AVS with the Cisco OpFlex protocol
- Integration with the Cisco ACI management and orchestration platform to automate virtual network provisioning and application services deployments
- Integrated visibility of both physical and virtual workloads and network paths
- Open APIs to extend the software-based control and orchestration of the virtual network fabric
- High performance and throughput

Cisco AVS offers:

- Single point of management and control for both physical and virtual workloads and infrastructure
- Optimal traffic steering to application services
- Transparent workload mobility
- Support for all leading hypervisors with a consistent operating model across implementations for simplified operations in heterogeneous data centers

OpFlex

OpFlex is an extensible policy protocol designed to exchange abstract policy between a network controller and a set of smart devices capable of rendering policy. OpFlex relies on a separate information model understood by agents in both the controller and the devices. This information model, which exists outside the OpFlex protocol itself, must be based on abstract policy, giving each device the freedom and flexibility to render policy within the semantic constraints of the abstraction. For this reason, OpFlex can support any device, including hypervisor switches, physical switches, and Layer 4 through 7 network services.

OpFlex can be used for a number of purposes in Cisco ACI. One common use case is Cisco AVS. With OpFlex, the Cisco ACI fabric can be extended all the way to the virtualized layer, allowing full policy enforcement and visibility directly into the hypervisor. Each edge device running Cisco AVS is handled as a virtual leaf (vLeaf). This approach allows traffic between two virtual machines on the same host to be switched locally with full policy enforcement. From an OpFlex perspective, each vLeaf peers with the physical leaf to which it attached to request and exchange policy information. Additionally, OpFlex and Cisco APIC interact with various hypervisor management systems to configure the virtual switches as needed.

Fundamentals of Cisco ACI

Cisco ACI abstracts policy and connectivity from the underlying fundamental network constructs of VLANs, IP addresses, access lists, and quality-of-service (QoS) polices. Application network connectivity should be described in more abstract terms, such as endpoint groups, providers and consumers, and service-level agreements (SLAs), so it is relevant for the end user of the fabric.

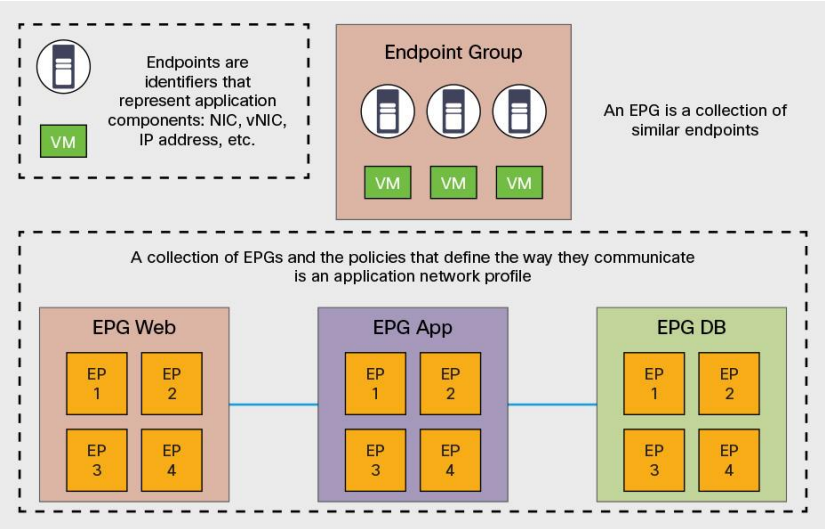
Cisco ACI provides a secure multitenant solution, allowing the network infrastructure administration and data flows to be segregated. Tenants can be customers, business units, groups, etc. Tenants can be further divided into Layer 3 constructs known as Virtual Routing and Forwarding (VRF) instances. Contexts provide a way to further separate the organizational and forwarding requirements of a tenant. Within each VRF instance, a bridge domain is created. A bridge domain is a Layer 2 namespace in which you define the various subnets. You assign all the subnets and default gateways within the bridge domain. By using separate forwarding instances, you can duplicate IP addressing in separate contexts for multitenancy.

Within the context, the model provides a series of objects that define the application. These objects are defined as endpoints and endpoint groups (EPGs; Figure 1).

EPG

Endpoints are devices (physical or virtual) that connect to the fabric and use it to interface with the network infrastructure. These endpoints can be computing, storage, and network services and security devices that attach to the fabric. At first customer shipment (FCS), Cisco ACI will support endpoints classified as network interface cards (NICs) or vNICs and their associated VLAN or Virtual Extensible LAN (VXLAN). In the future, endpoint support will be extended to IP addresses, MAC addresses, Domain Name System (DNS) names, virtual machine attributes, IEEE 802.1x identity, and other common attributes.

Figure 1. Endpoint Groups



An EPG is a collection of endpoints with the same types of attributes and identical network behavior (for example, connectivity, security, and QoS requirements).

Here are some examples of EPGs:

- EPG defined by traditional network VLANs: All endpoints connected to a given VLAN placed in an EPG
- EPG defined by VXLAN: Same as for VLANs except using VXLAN
- EPGs defined as security zones
- EPGs defined as application tiers (web, application, and database)
- EPG mapped to a VMware ESXi port group

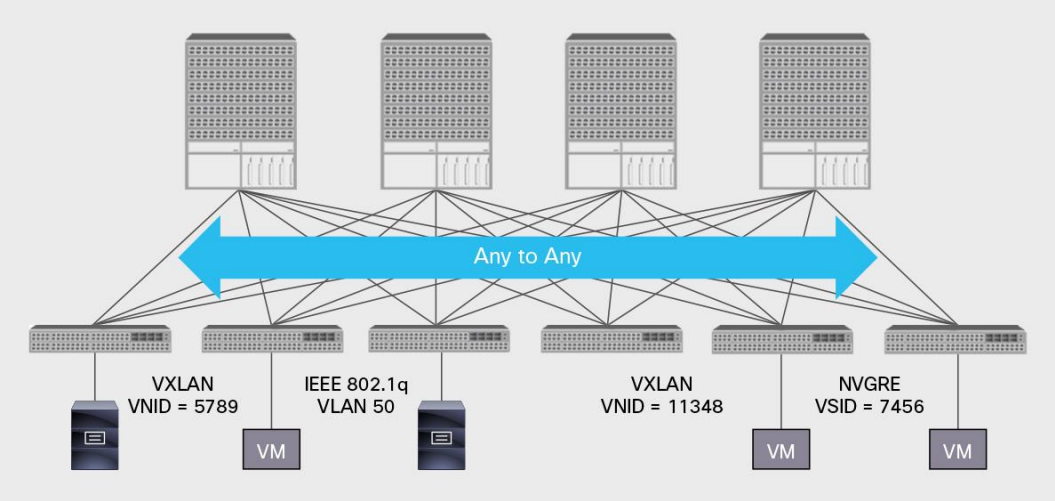
The policies that are used to describe the communication, services insertion, and QoS and SLAs that are embedded between EPGs are referred to as contracts. A contract is a set of network requirements that describes how EPGs can communicate with each other and with the outside world. The Cisco ACI contract defines a filter, which includes a Layer 4 ACL and an associated action that dictates whether the traffic is permitted, denied, logged, marked, redirected, or copied. Cisco ACI security uses a whitelist model, which denies all traffic by default. Administrators must explicitly allow communication between EPGs.

For each tenant, EPGs and policies are summarized in an application network profile (ANP). These ANPs are the logical representation of the application infrastructure requirements. When the application is ready to be provisioned, Cisco APIC can push the ANP and provision the entire stateless Cisco ACI fabric instantaneously.

Cisco ACI Integration with Multiple Hypervisors

An important benefit of Cisco ACI, is the capability to be hypervisor independent. Cisco ACI is independent to the tenant traffic; it can be tagged with IEEE 802.1q (VLAN), VXLAN, or NVGRE. Traffic forwarding is not limited to nor constrained within the encapsulation type or encapsulation overlay network (Figure 2).

Figure 2. Cisco ACI Is Hypervisor Independent



Cisco ACI uses an extended VXLAN encapsulation frame format within the fabric. All tenant traffic within the fabric is tagged at the first hop leaf ingress port with an extended VXLAN header that identifies the policy attributes of the application endpoint within the fabric. The VXLAN header carries the virtual network ID (VNID) along with the EPG policy. As a result, the policy attributes are carried in every packet. As workloads move within the virtual environment, the policies attached to the workloads are enforced transparently and consistently within the infrastructure. When the packet leaves the fabric, the VXLAN header is deencapsulated, and the packet is encapsulated with any tag of the tenant's choice: VLAN, VXLAN, or NVGRE.

Virtual machine management (VMM) is used in Cisco ACI to define a hypervisor management system that has control over the virtual machines. A Cisco ACI fabric can have multiple VMM domains across hypervisors from a variety of vendors. Each VMM domain has local significance to the VLAN or VXLAN and contains its own mobility domain, so that mobility is restricted to that VMM domain. Across the different VMM domains, the VLAN or VXLAN namespace can be reused. At FCS, the VLAN ID has local significance to the leaf node. In the future, the VLAN ID will have local port significance. As a best practice, do not overlap VLAN namespace for a given leaf node. Designate a few leaf nodes and put them in a VMM domain; they can own the VLAN namespace. With VMM domains, a customer can scale beyond the 4000 VLANs. In a large-scale cloud environment, 4000 VLANs or identifiers is not enough to uniquely identify each EPG or network. Each VMM domain can scale up to 4000 VLANs. As a result, you can span and scale horizontally using VMM domains and achieve more than 4000 EPGs, each mutually exclusive. The only restriction, as mentioned previously, is that mobility is restricted to the VMM domain.

The location and reachability of VMware virtual endpoints can be learned in several ways:

- Control-plane learning
  - Out-of-band (OOB) handshake: The OOB control protocol is used to communicate between the VMM and Cisco APIC. The VMware vCenter knows where the hosts are because it performs virtual host placement. You can use that information to identify the location where the host resides.
  - In-band handshake: Identify OpFlex-enabled Cisco AVS hosts.
  - Link-Layer Directory Protocol (LLDP) and Cisco Discovery Protocol: Identify the virtual host ID of the attached port on the leaf node.

- Data-plane learning
  - Distributed switch learning: Use when you want to distribute and program policy in hardware.

When a virtual endpoint is discovered, the policy is pushed and programmed to the leaf nodes based on resolution immediacy and instrumentation immediacy, respectively. In both cases, there is an immediate and on-demand (default) option that is defined when the VMM is associated on Cisco APIC. The on-demand option conserves resources and uses the reserved space in the policy content-addressable memory (CAM) when needed.

Resolution Immediacy

The first option to push a policy is immediately. All policies (VLAN, NVGRE, and VXLAN), bindings, contracts, and filters are pushed to the leaf node when the hypervisor physical NIC (pNIC) is attached. With the on-demand option, policies are pushed to the leaf node when the pNIC and vNIC are attached to the port group (EPG).

Deployment Immediacy

Deployment immediacy defines when the policy is programmed in hardware. If the immediate option is chosen, the policies are programmed in the policy CAM after they are received by Cisco APIC. The on-demand option programs policies in the hardware policy CAM only when reachability is learned through the data path.

Integrating Cisco ACI with Cisco AVS

Cisco AVS is a distributed, Layer 2 virtual switch that extends across many virtualized hosts. Cisco AVS manages and control data center virtual switching by integrating with different Hypervisor and their management servers for example VMwae vCenter Server. Cisco AVS is compatible with any upstream physical access layer switch that is Ethernet standards compliant, including the Cisco Catalyst® 6500 Series Switches, Cisco Nexus® switches, and switches from other network vendors. Cisco AVS is compatible with any server hardware listed in the VMware Hardware Compatibility List (HCL).

Cisco AVS uses the VMware vNetwork Distributed Switch (vDS) API, which was developed jointly by Cisco and VMware, to provide advanced networking capability to virtual machines. This solution offloads the configuration of the virtual switch and port groups to the APIC Controller to enforce a consistent data center network policy. At the time of this writing, Cisco AVS supports the VMware hypervisor only for Cisco ACI. Cisco AVS is supported as a vLeaf for Cisco APIC with the VMware ESX and ESXi Release 5.1 or later hypervisor.

Each hypervisor is embedded with one virtual Ethernet module (VEM), which is a lightweight software component that replaces the virtual switch by performing the following functions:

- Advanced networking and security
- Switching between directly attached virtual machines
- Uplink to the rest of the network

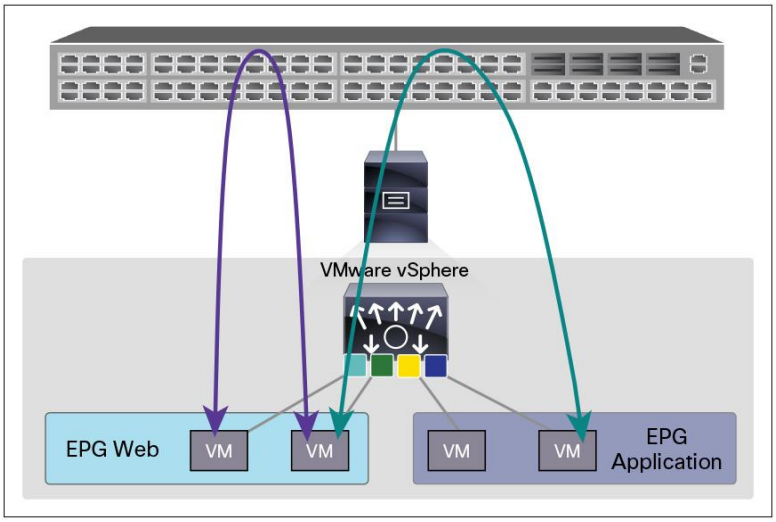
In Cisco AVS, traffic is switched between virtual machines locally at each Cisco AVS instance. Each Cisco AVS also interconnects the local virtual machine with the rest of the network through the upstream access-layer network switch (blade, top-of-rack, end-of-row, and so forth).

In Cisco AVS, the module slots are for primary module 1 and secondary module 2. Either module can act as the active or standby device. The first server or host is automatically assigned to module 3. The NIC ports are 3/1 and 3/2 (pNIC0 and pNIC1 on the VMware ESX or ESXi host). The ports to which the vNIC interfaces connect are virtual ports on Cisco AVS, where they are assigned a global number.

At Cisco ACI FCS, Cisco AVS will support two modes of operation:

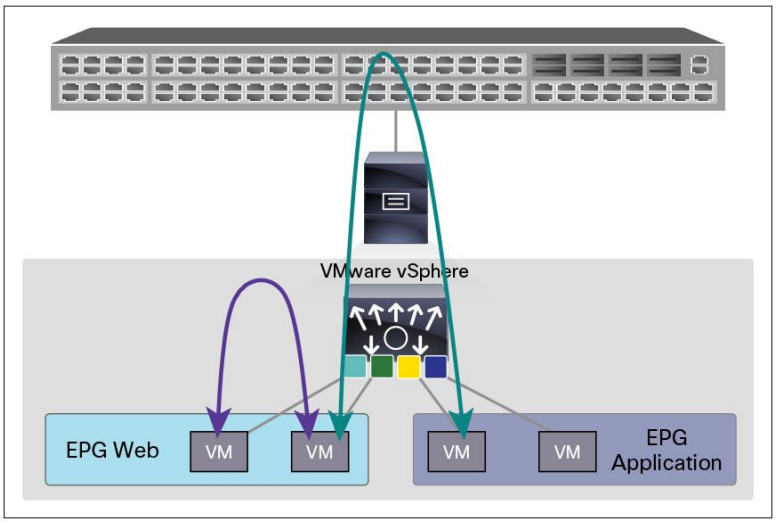
Fabric Extender (FEX) enabled mode: Cisco AVS sends all packets received from vNICS to Cisco ACI for all forwarding. Essentially, any intra-EPG and inter-EPG policies are enforced at the Cisco ACI leaf (Figure 3).

Figure 3. Cisco AVS in FEX Enabled Mode



Local Switching Mode: Cisco AVS can perform local switching (similar VMware ESX Virtual Distributed Switch [VDS]) within EPGs on the same host. In fabric extender disabled mode, Cisco AVS can perform do intra-EPG switching; all inter-EPG or external policies are still enforced at the Cisco ACI leaf (Figure 4).

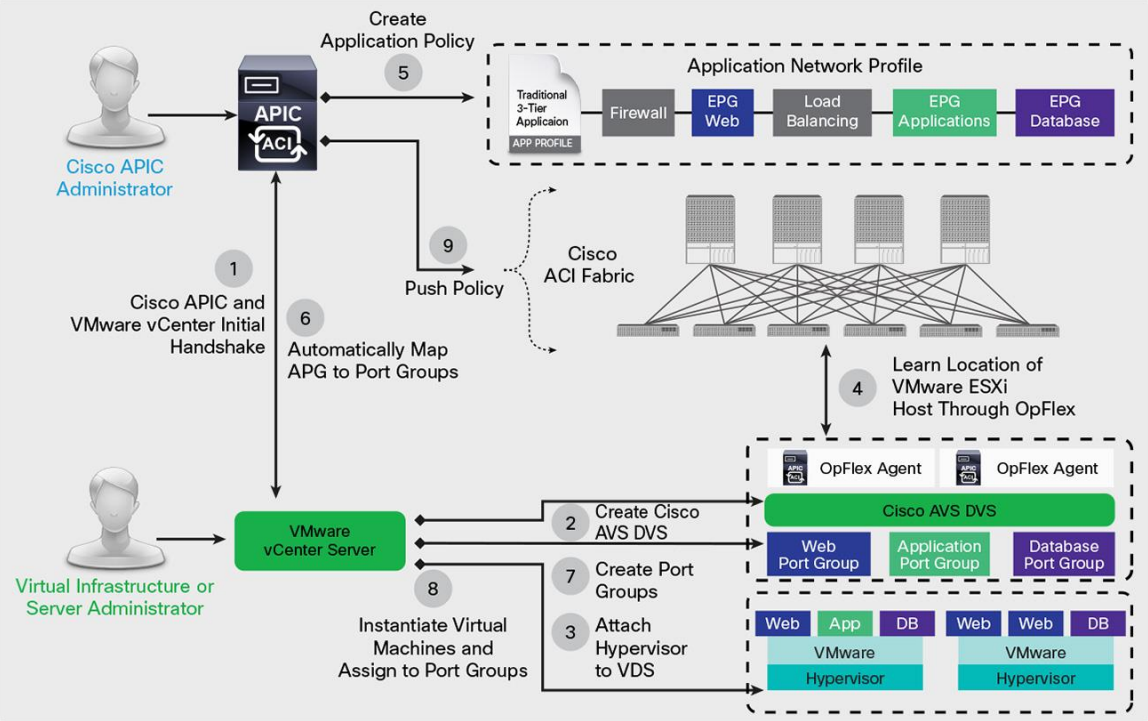
Figure 4. Cisco AVS in FEX Disabled Mode



Cisco ACI and AVS Workflow

Figure 5 shows the Cisco ACI and Cisco AVS workflow.

Figure 5. Workflow



For More Information

- Cisco ACI: http://www.cisco.com/go/aci
- VMware: http://www.vmware.com

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

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

Printed in USA C11-731999-00 08/14