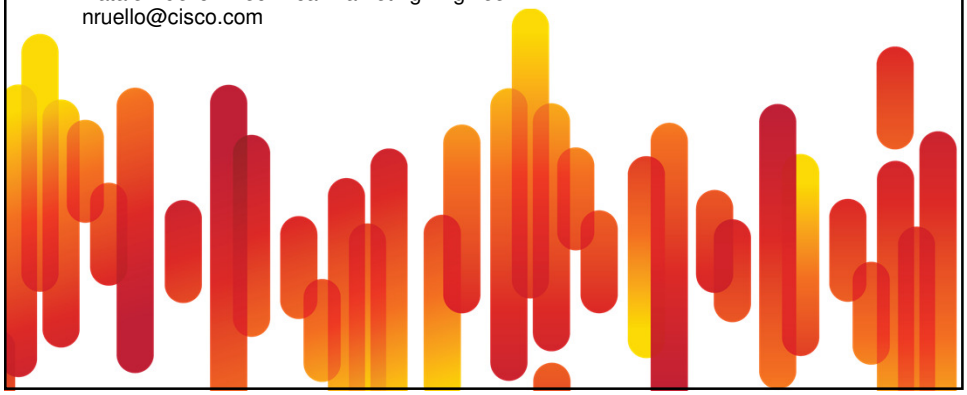
Overlay Transport Virtualization

BRKDCT-2049

Natale Ruello – Technical Marketing Engineer
nruello@cisco.com



Housekeeping

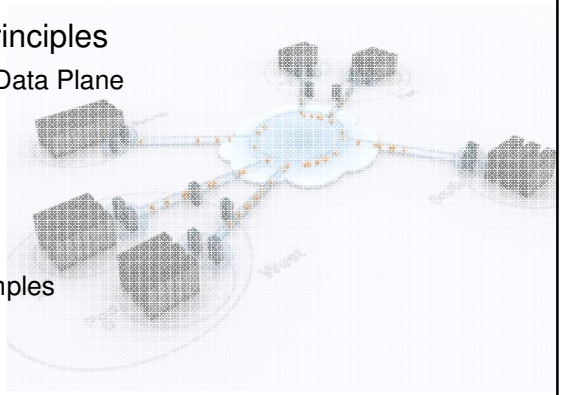
- We value your feedback—don't forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Thursday
- Visit the World of Solutions
- Please remember this is a 'non-smoking' venue!
- Please switch off your mobile phones
- Please make use of the recycling bins provided
- Please remember to wear your badge at all times

Meet the Engineer

- To make the most of your time at Networkers at Cisco Live 2010, schedule a Face-to-Face Meeting with a top Cisco Engineer
- Designed to provide a “big picture” perspective as well as “in-depth” technology discussions, these face-to-face meetings will provide fascinating dialogue and a wealth of valuable insights and ideas
- Visit the Meeting Centre reception desk located in the Meeting Centre in World of Solutions

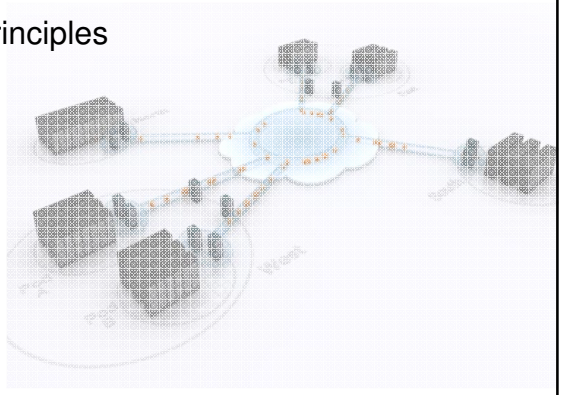
Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
- Use Cases



Distributed Data Centers

Building the Data Center Cloud

Distributed Data Center Goals:

- Seamless workload mobility between multiple datacenters.
- Distributed applications closer to end users.
- Pool and maximize global compute resources.
- Ensure business continuity with workload mobility and distributed deployments.

Distributed Data Centers
Challenges with the Existing Solutions

- Complex operations – Current solutions are complex to deploy and manage.
- Transport dependant – Requires the provisioning of specific transport (MPLS, Dark fiber, etc.).
- Bandwidth management – Inefficient use of bandwidth.
- Failure containment – Failures from one data center can impact all data centers.

Overlay Transport Virtualization (OTV)

OTV delivers a virtual L2 transport over any L3 Infrastructure

O ***Overlay*** - A solution that is *independent of the infrastructure technology* and services, flexible over various inter-connect facilities

T ***Transport*** - Transporting services for *layer 2 and layer 3* Ethernet and IP traffic

V ***Virtualization*** - Provides *virtual* connections, *connections* that are in turn *virtualized and partitioned* into VPNs, VRFs, VLANs

Overlay Transport Virtualization (OTV)

Simplifying Data Center Interconnect

- Ethernet LAN Extension over any Network
  - Works over dark fiber, MPLS, or IP network
  - Multi-data center scalability
- Simplified Configuration and Operation
  - Seamless overlay – no network redesign
  - Single touch site configuration
- High Resiliency
  - Failure domain isolation
  - Seamless Multi-homing
- Maximizes available bandwidth
  - Automated multipathing
  - Optimal multicast replication

Any Workload, Anytime, Anywhere

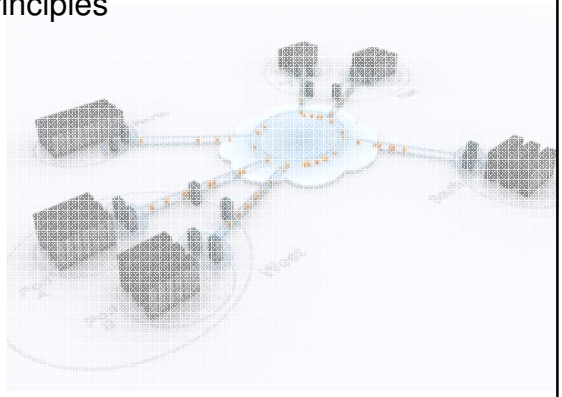
Unleashing the full potential of compute virtualization

OTV Benefits

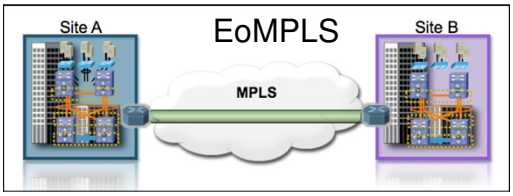
Business Goals	*OTV LAN Extensions Enable*	*Benefit*
99.999% Global Availability	Enable *Distributed Clusters* to improve Application Availability without compromising Network Resiliency	*Application Resiliency*
Service Velocity and On-Demand Capacity	Unleash *Compute Virtualization* beyond a single physical data center *for fast service and capacity additions*	*Geo Diversity and Adaptability*
Maximize Asset Utilization	Supports *migration of workloads* across locations *to avoid power/cooling hot spots or compute/network idleness*	
Streamline Operations and Reduce OPEX	Enables *improved change management* methods across multiple physical locations	*Business Flexibility*

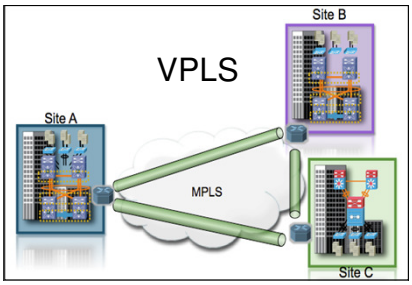
Agenda

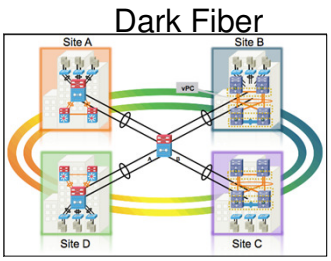
- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
- Use Cases



Traditional Layer 2 VPNs

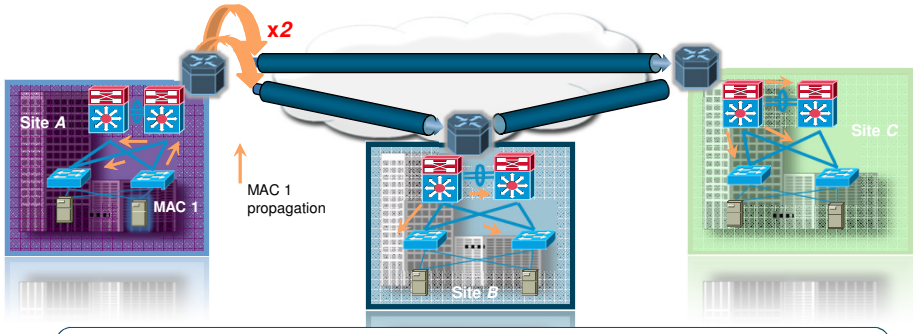






Flooding Behavior

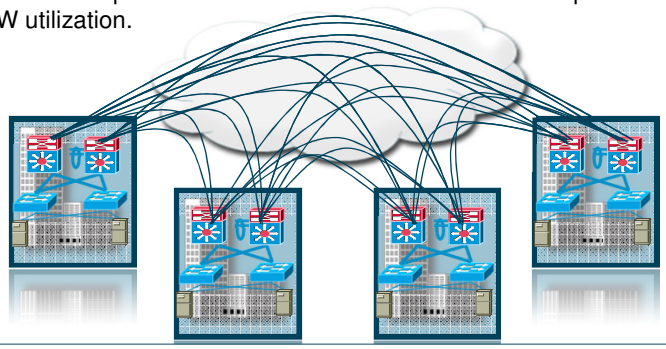
- Traditional Layer 2 VPN technologies rely on flooding to propagate MAC reachability.
- The flooding behavior causes failures to propagate to every site in the Layer 2 VPN.



The new solution should…
provide layer 2 connectivity, yet restrict the reach of the flooding domain in order to contain failures and preserve the resiliency.

Pseudo-Wires Maintenance

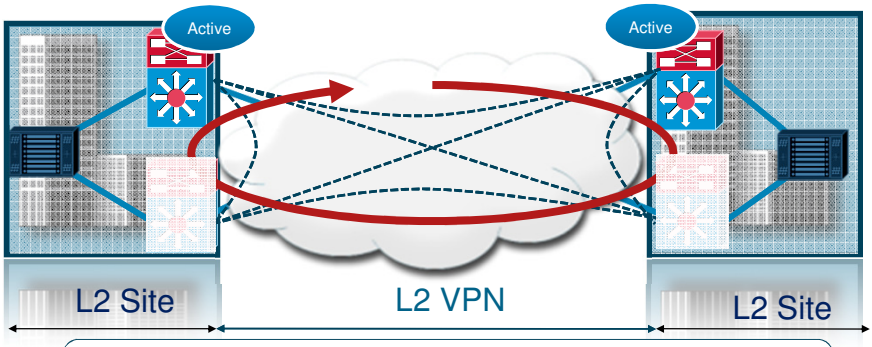
- Before any learning can happen a full mesh of pseudo-wires/ tunnels must be in place.
- For N sites, there will be ***N*(N-1)/2*** pseudo-wires. Complex to add and remove sites.
- Head-end replication for multicast and broadcast. Sub-optimal BW utilization.



The new solution should… provide point-to-cloud provisioning and optimal bandwidth utilization in order to reduce cost.

Multi-homing

- Require additional protocols to support Multi-homing.
- STP is often extended across the sites of the Layer 2 VPN. Very difficult to manage as the number of sites grows.
- Malfunctions on one site will likely impact all sites on the VPN.



The new solution should… natively provide automatic detection of multi-homing without the need of extending the STP domains, together with a more efficient load-balancing.

The new solution will…

- Flooding Based Learning → Control-Plane Based Learning

  Move to a Control Plane protocol that proactively advertises MAC addresses and their reachability instead of the current flooding mechanism.

- Pseudo-wires and Tunnels → Dynamic Encapsulation

  Not require static tunnel or pseudo-wire configuration.

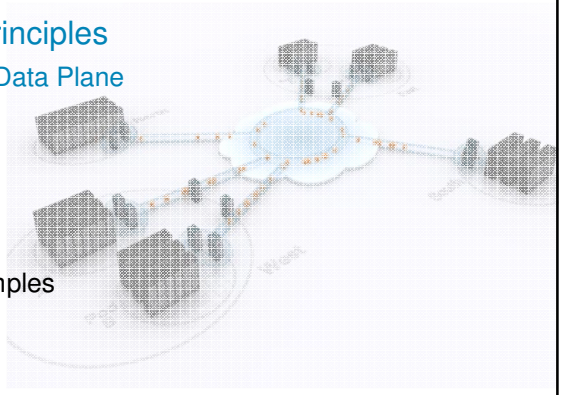
  Offer optimal replication of traffic done closer to the destination, which translates into much more efficient bandwidth utilization in the core

- Multi-homing → Native Built-in Multi-homing

  Allow load balancing of flows within a single VLAN across the active devices in the same site, while preserving the independence of the sites. STP confined within the site (each site with its own STP Root bridge)

Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



Overlay Transport Virtualization

Technology Pillars

OTV is a "MAC in IP" technique to extend Layer 2 domains **OVER ANY TRANSPORT**

Dynamic Encapsulation

- No Pseudo-Wire State Maintenance
- Optimal Multicast Replication
- Multipoint Connectivity
- Point-to-Cloud Model

Nexus 7000

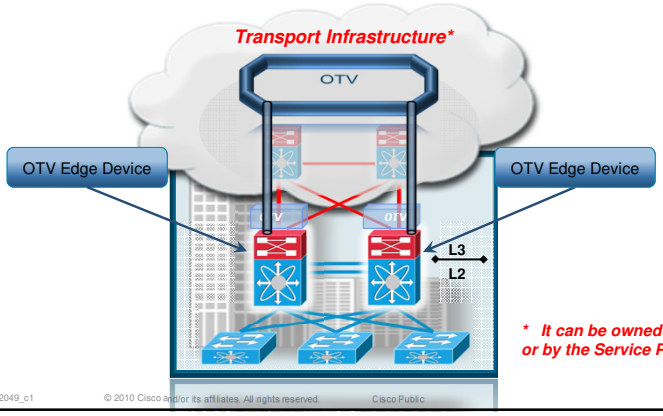
First platform to support OTV starting with 5.0(3) release!

Protocol Learning

- Preserve Failure Boundary
- Built-in Loop Prevention
- Automated Multi-homing
- Site Independence

Terminology: "Edge Device"

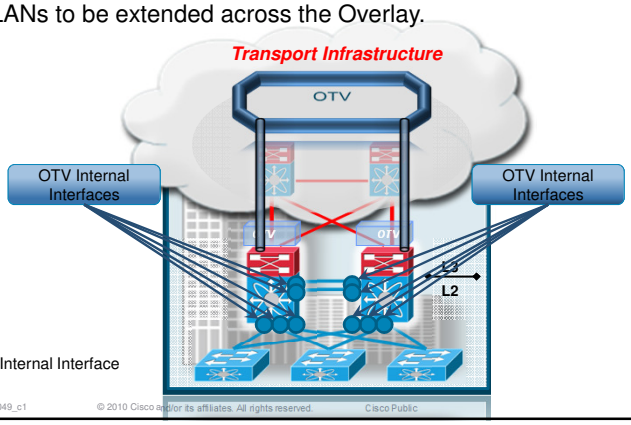
- The *Edge Device* is responsible for performing all the OTV functionality.
- The *Edge Device* can be located at the Aggregation Layer as well as at the Core Layer depending on the network topology of the site.
- A given site can have multiple OTV *Edge Devices (multi-homing).*



** It can be owned by the Enterprise or by the Service Provider*

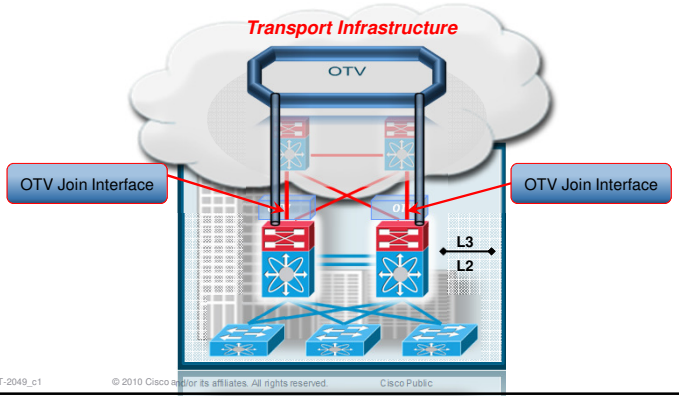
Terminology: "Internal Interfaces"

- The *Internal Interfaces* are those interfaces of the Edge Devices that face the site and carry at least one of the VLANs extended through OTV.
- *Internal Interfaces* behave as regular layer 2 interfaces. No OTV configuration is needed on the OTV *Internal Interfaces.*
- Typically these interfaces are configure as Layer 2 trunks carrying the VLANs to be extended across the Overlay.



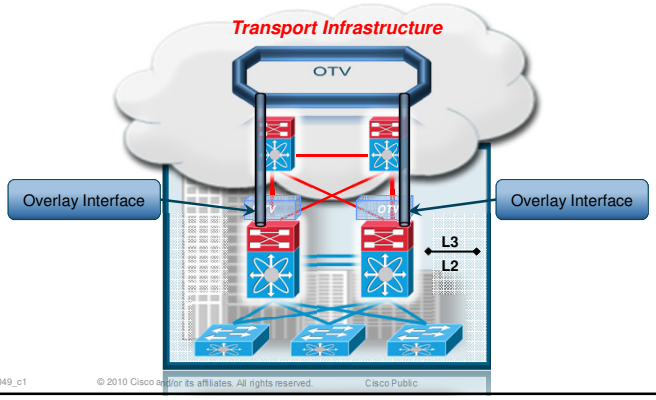
Terminology: "Join Interface"

- The *Join interface* is one of the uplink interfaces of the Edge Device.
- The *Join Interface* is usually a point-to-point routed interface and it can be a single physical interface as well as a port-channel (higher resiliency).
- The *Join Interface* is used to physically "join" the Overlay network.



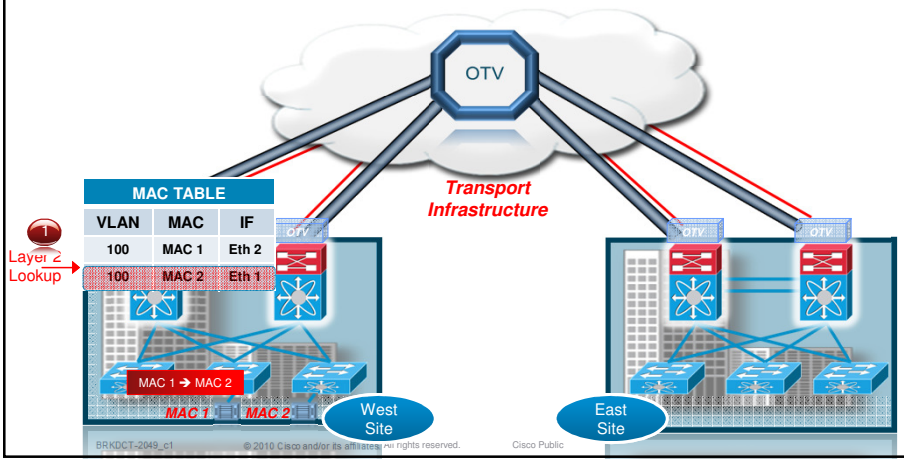
Terminology: "Overlay Interface"

- The *Overlay Interface* is the **virtual** interface where all the OTV configuration is placed.
- It's a logical multi-access multicast-capable interface.
- The *Overlay Interface* encapsulates the site Layer 2 frames in IP unicast or multicast packets that are then sent to the other sites.



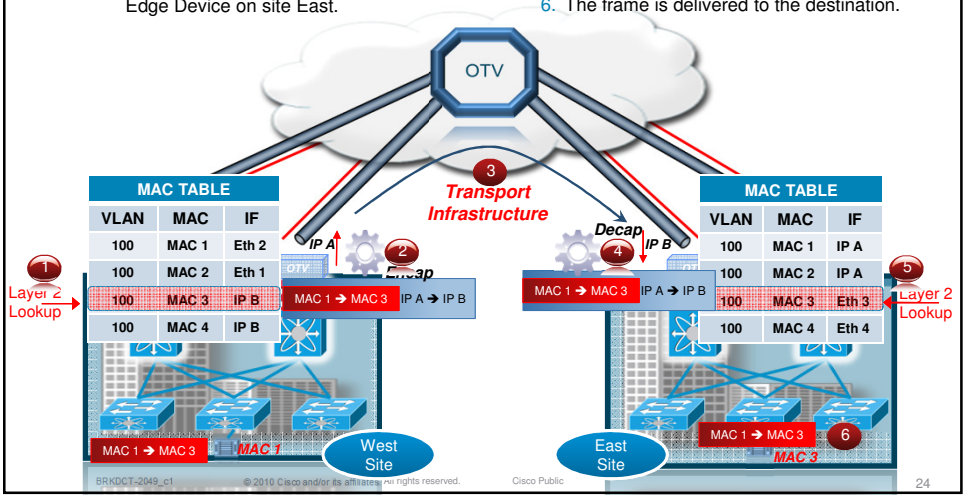
OTV Data Plane: Intra-Site Packet Flow

1. Layer 2 lookup on the destination MAC address.
2. MAC 2 is reachable through Ethernet 1.
3. The frame is delivered to the destination.



OTV Data Plane: Inter-Site Packet Flow

1. Layer 2 lookup on the destination MAC. MAC 3 is reachable through IP B.
2. The Edge Device encapsulates the frame.
3. The transport delivers the packet to the Edge Device on site East.
4. The Edge Device on site East receives and decapsulates the packet.
5. Layer 2 lookup on the original frame. MAC 3 is a local MAC.
6. The frame is delivered to the destination.



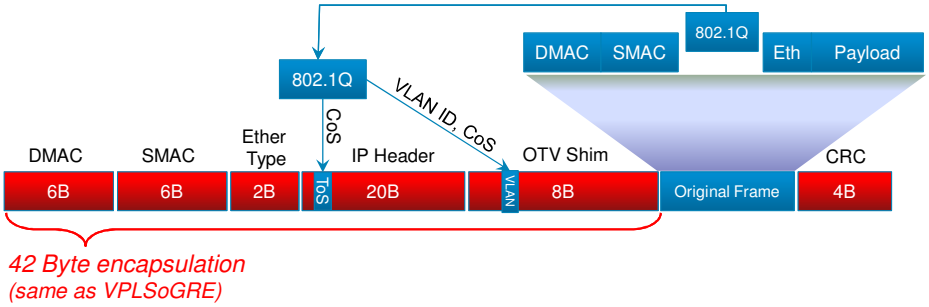
OTV Data Plane: Inter-Site Packet Flow

The frame goes from Server 1 (MAC 1) on site West to Server 3 (MAC 3) on site East:

1. The Layer 2 frame arrives at the West Site Edge Device. A classic Layer 2 lookup on the destination MAC address takes place
   1. The destination MAC address, MAC 3, is reachable through an IP address, which indicates that MAC 3 is not a local MAC. MAC 3 is in fact reachable through IP B, which is the IP address of the join-interface of the Edge Device in site East
2. MAC 3 is reachable through IP B, the Edge Device then encapsulates the original frame into an IP packet where the *IP_SA* is IP A and the *IP_DA* is IP B.
3. The encapsulated packet is now passed to the Core which will deliver it to its destination: the Edge Device on site East.
4. The Edge Device on site East receives and decapsulates the packet. We have now the original Layer 2 frame.
5. Another classic Layer 2 lookup is then performed on the frame. MAC 3 is now reachable through a physical interface. It's in fact a local MAC.
6. The Layer 2 frame is delivered to its destination server.

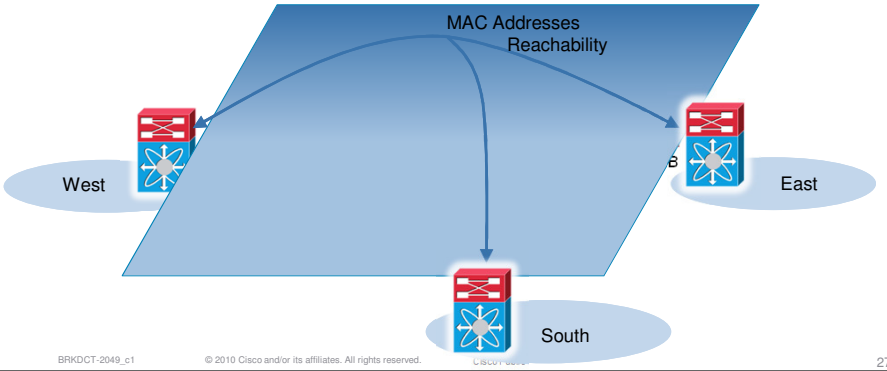
OTV Data Plane Encapsulation

- OTV adds a 42 Byte IP encapsulation.
- The outer IP header is followed by an OTV shim header, which contains information about the overlay (vlan, overlay number, etc).
- The 802.1Q header is extracted from the original frame and the VLAN field copied over into the OTV shim header.
- The OTV Edge Device can also map the 802.1p CoS bits to the outer IP header's DSCP field as well as to the OTV Shim header.



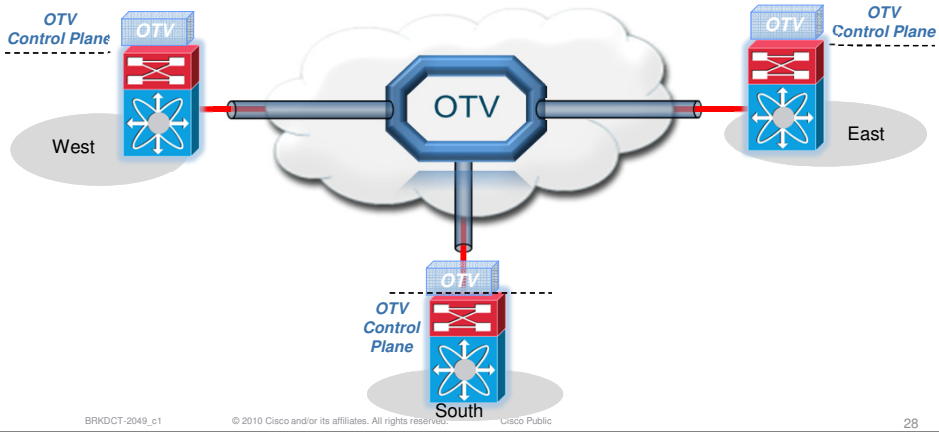
Building the MAC tables
The OTV Control Plane

- The OTV control plane **proactively advertises** MAC reachability (control-plane learning).
- The MAC addresses are advertised in the **background** once OTV has been configured.
- *No protocol specific configuration is required.*



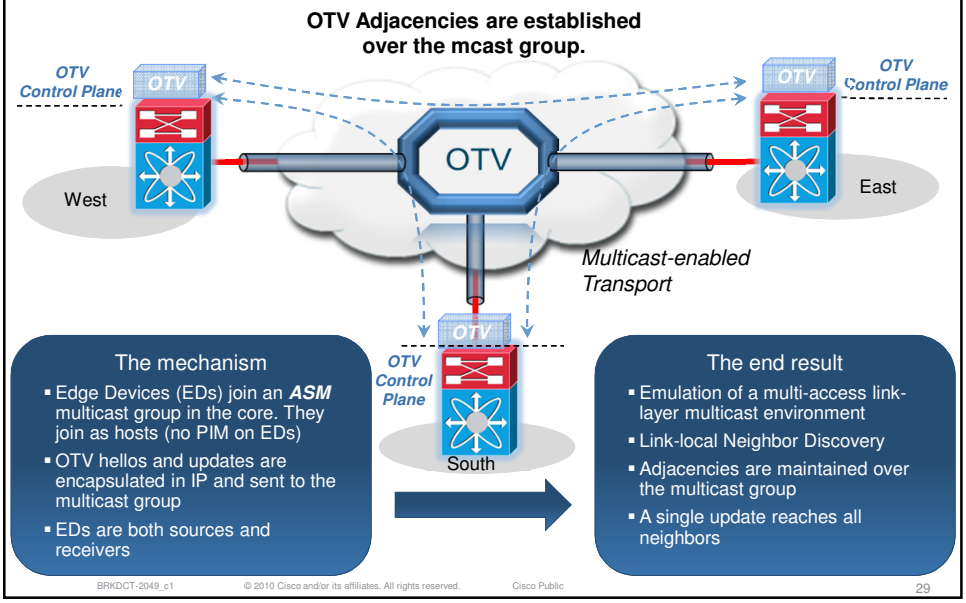
OTV Control Plane
Neighbor Discovery and Adjacency Formation

- The *Edge Devices build a neighbor relationship with each other from the OTV Control Plane perspective*.
- The neighbor relationship can be built over a **multicast-enabled** as well as over an **unicast-only** transport infrastructure.
  OTV supports both scenarios.



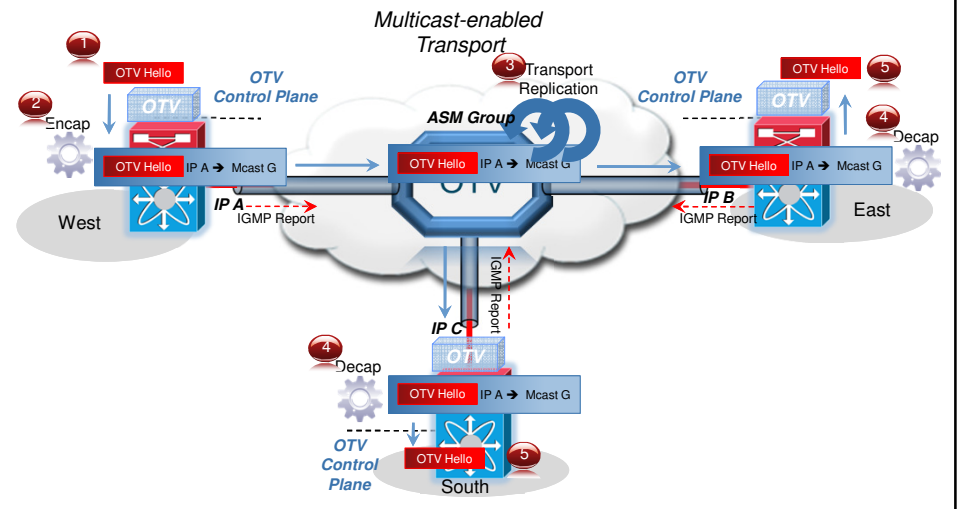
OTV Control Plane

Neighbor Discovery (Multicast-Enabled Transport)



OTV Control Plane

Neighbor Discovery (Multicast-Enabled Transport – 1)



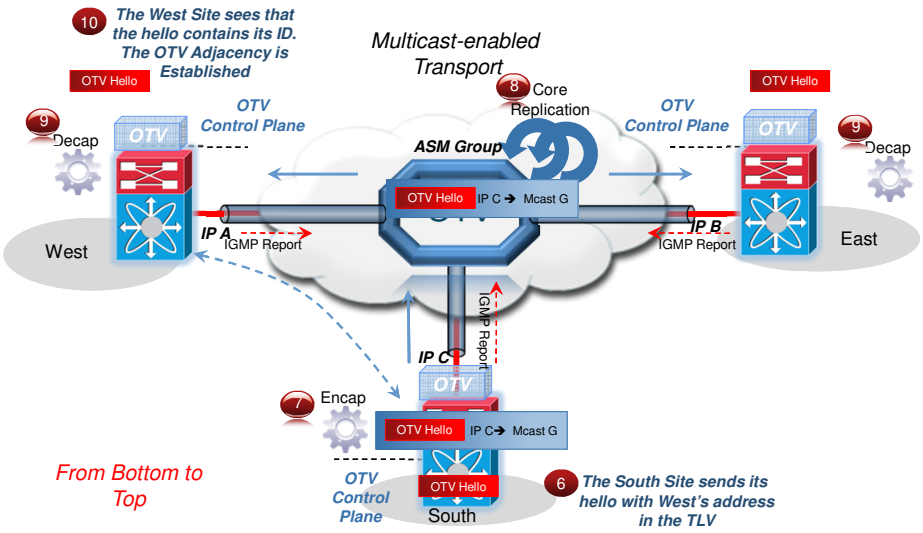
OTV Control Plane

Neighbor Discovery (Multicast-Enabled Transport – 1)

0. The Edge Devices (EDs) join an ASM mcast group in the core. They join has hosts by sending IGMP report for the ASM group.
1. The OTV control plane in the ED of the West site generates an OTV hello.
2. The ED encapsulates the OTV hello into an IP packet where the IP destination address is the ASM mcast group in the core which was previously joined by the ED.
3. The core receives this mcast packet and performs an optimal replication so that all the EDs on the specific Overlay receive the packet.
    1. The ASM group joined by the EDs identifies the Overlay. All the EDs belonging to a specific Overlay will join the same ASM group. Two different Overlays cannot use the same ASM group in the core.
4. The packet is received by the other EDs which will then perform a decapsulation.
5. The original OTV hello is delivered to the OTV control plane.

OTV Control Plane

Neighbor Discovery (Multicast-Enabled Transport – 2)



OTV Control Plane

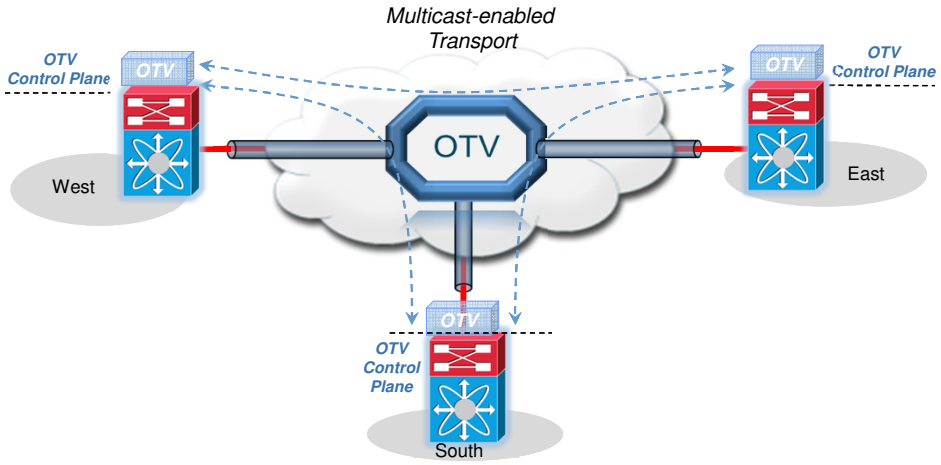
Neighbor Discovery (Multicast-Enabled Transport – 2)

6. The South Site now sends its OTV hello. In the TLV of the OTV hello the South Site will include the West Site ID.
7. The South Site ED encapsulates the OTV hello into an IP packet where the IP destination address is the ASM mcast group which identifies the Overlay.
8. The core receives this mcast packet and performs an optimal replication so that all the EDs on the specific Overlay receive the packet.
9. The packet is received by the EDs belonging to the Overlay, which perform the decapsulation and deliver the original OTV hello to the OTV control plane.
10. The OTV control plane on the West Site ED sees that the OTV hello just received from the South Site contains the West Side ID. This indicates to the OTV control plane that there is a two ways communication between the West and the South sites, which allows the OTV adjacency to be formed.

OTV Control Plane

Neighbor Discovery (Multicast-Enabled Transport)

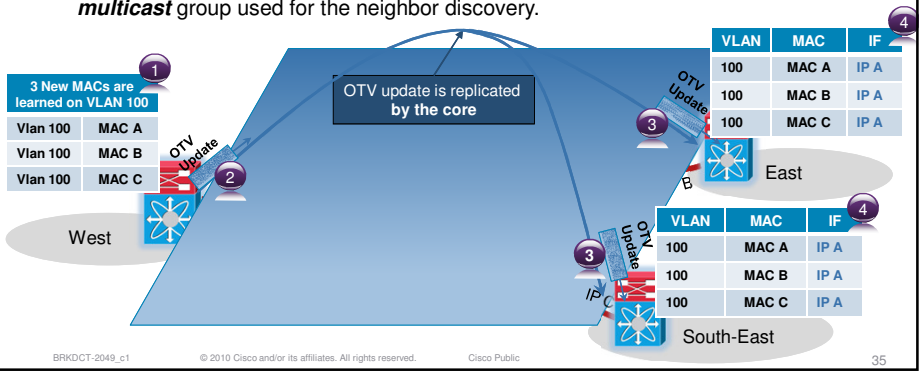
OTV Adjacencies Established over the mcast group in the core



OTV Control Plane

MAC Address Advertisements (Multicast-Enabled Transport)

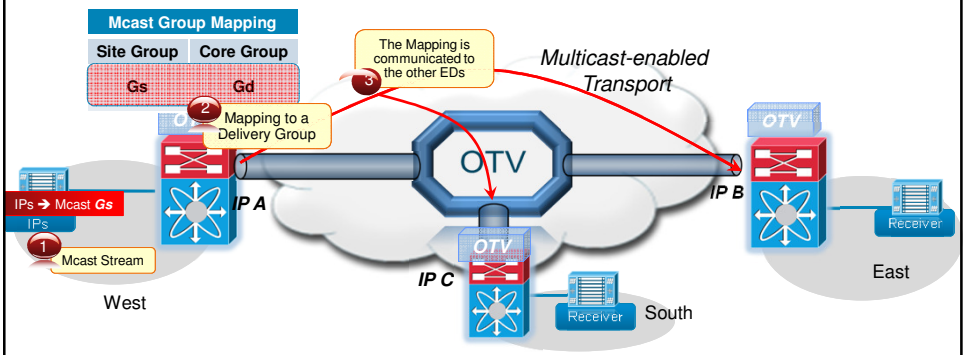
- Every time an Edge Device learns a new MAC address, the OTV control plane will advertise it together with its associated VLAN IDs and IP next hop.
- The IP next hops are the addresses of the Edge Devices through which these MACs addresses are reachable in the core.
- A single OTV update can contain multiple MAC addresses for different VLANs.
- A single update reaches all neighbors, as it is encapsulated in the same ***ASM multicast*** group used for the neighbor discovery.



OTV Data Plane: Multicast Data

Mapping of the multicast groups

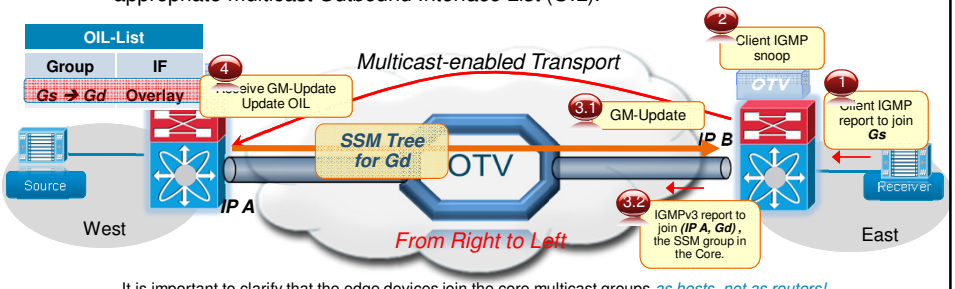
- The site mcast groups are mapped to a ***SSM group range*** in the core.
- This allows the mcast traffic to be transported on the Overlay without the need to run mcast with the core, which could be owned by a Service Provider.



1. The Mcast source starts sending traffic to the group ***Gs.***
2. The West ED maps (S,Gs) to a ***delivery group Gd*** (from the SSM group in the core).
3. The West ED communicates the mapping information (including the source VLAN) to the East ED.

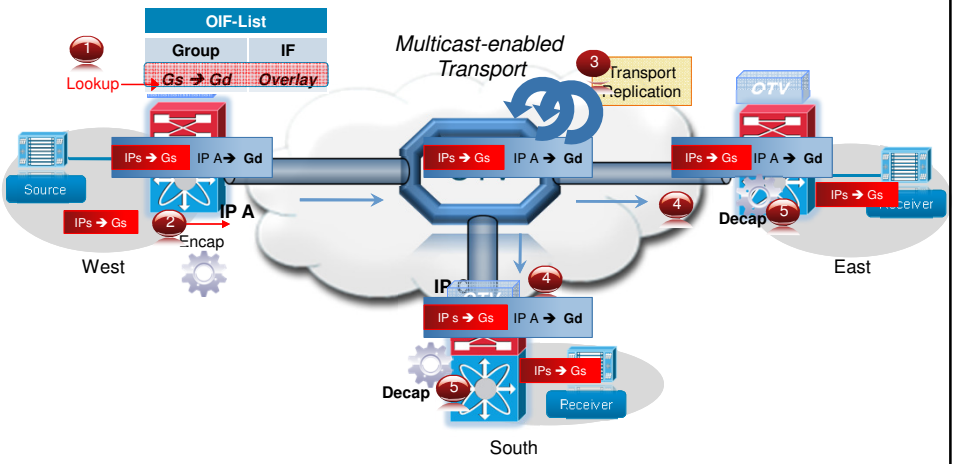
OTV Data Plane: Multicast Data
Multicast State Creation

1. The multicast receivers for the multicast group "Gs" on the East site send IGMP reports to join the multicast group.
2. The *Edge Device* (ED) snoops these IGMP reports, but it doesn't forward them.
3. Upon snooping the IGMP reports, the ED does two things:
   1. Announces the receivers in a Group-Membership Update (GM-Update) to all EDs.
   2. Sends an IGMPv3 report to join the ***(IP A, Gd)*** group in the core.
4. On reception of the GM-Update, the source ED will add the overlay interface to the appropriate multicast *Outbound Interface List* (OIL).



It is important to clarify that the edge devices join the core multicast groups *as hosts, not as routers!*

OTV Data Plane: Multicast Data
Multicast Packet Flow



OTV Data Plane: Multicast Data

Multicast Packet Flow

1. The multicast frame with IP_DA set as the ***Gs*** mcast group reaches the ED. An OIF lookup takes place. The table shows that there are receivers across the Overlay.
2. The ***Gs*** mcast group is mapped to the ***Gd*** group in the core (one of the SSM address from the defined range). The original multicast frame is encapsulated into a multicast packet with ***Gd*** as IP_DA and sent to the core.
3. The core is responsible for the optimal replication and delivery of the packet.
4. The other EDs, in the sites where the receivers for the Gs group are, receive the multicast packet.
5. Decapsulation takes place and the original multicast frame is delivered to the receivers.

Summary of the Multicast Groups used in a Multicast-Enabled Transport

- OTV is able to leverage the multicast capabilities of the core.
- This is the summary of the Multicast groups used by OTV:

  An ***ASM group*** used for neighbor discovery and to exchange MAC reachability.

  A ***SSM group range*** to map the sites internal multicast groups to the mcast groups in the core, which will be leveraged to extend the mcast data traffic across the Overlay.

Unicast-Only Transport?

OTV has a solution for it

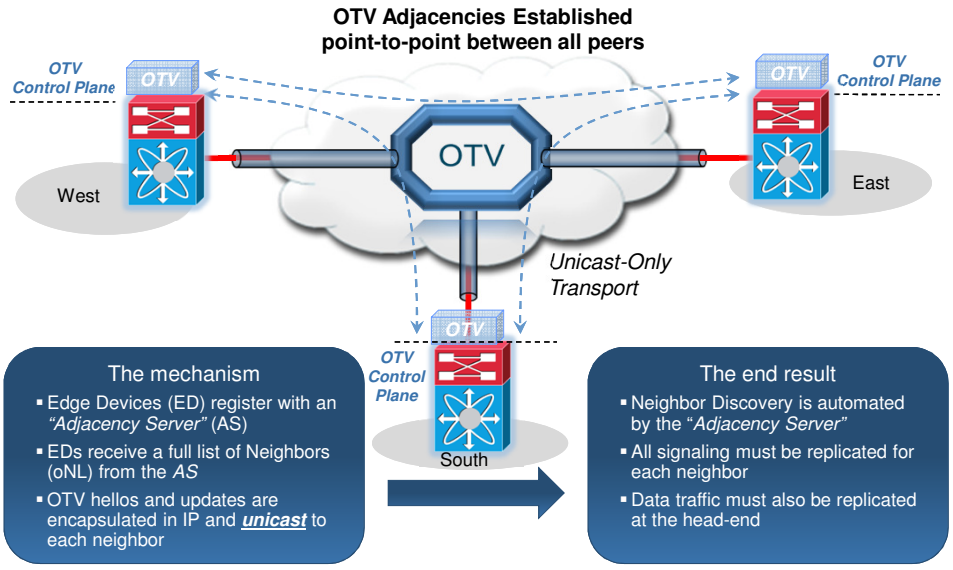
Adjacency Server Mode

- The use of multicast in the core provides significant benefits:
  - Reduces the amount of hellos and updates OTV must issue
  - Streamlines neighbor discovery, site adds and removes
  - Optimizes the handling of broadcast and multicast data traffic
- However multicast support may not always be available.
- The *OTV Adjacency Server Mode* of operation provides the solution for the unicast-only cores.

Supported in the Next Software Release

OTV Control Plane

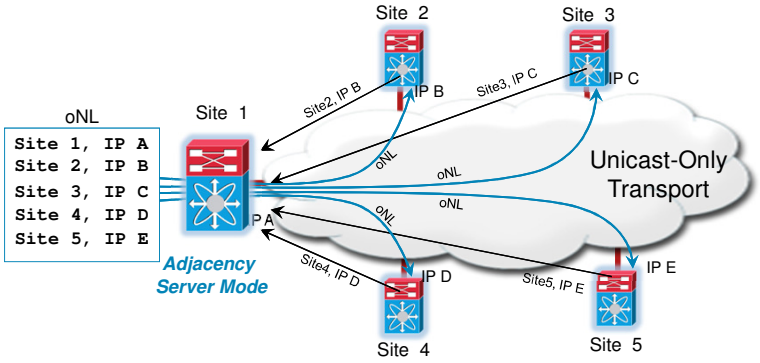
Neighbor Discovery (Unicast-Only Transport)



OTV Control Plane

Neighbor Discovery (Unicast-Only Transport)

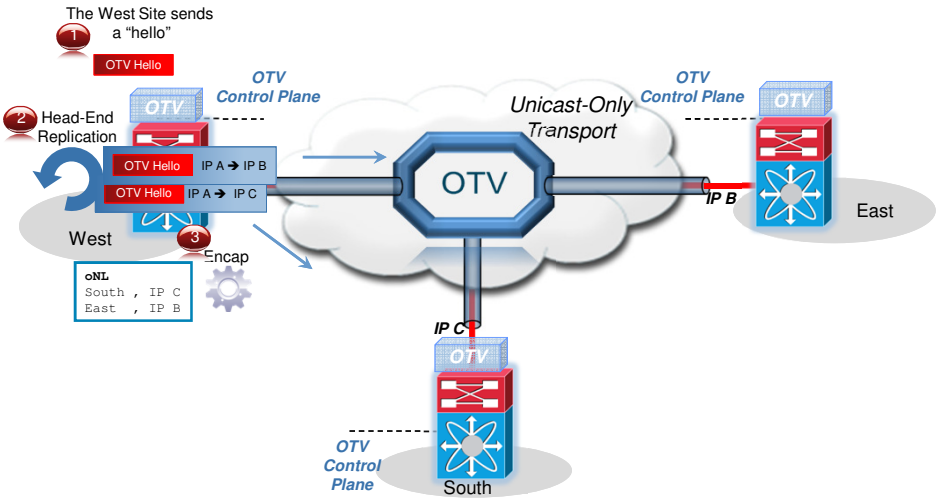
1. One of the OTV Edge Devices (ED) is configured as an Adjacency Server (AS)*.
2. All EDs are configured to register to the AS: send their site-id and IP address.
3. The AS builds a list of neighbor IP addresses: **overlay Neighbor List (oNL).**
4. The AS unicasts the oNL to every neighbor.
5. Each node unicasts hellos and updates to every neighbor in the oNL.



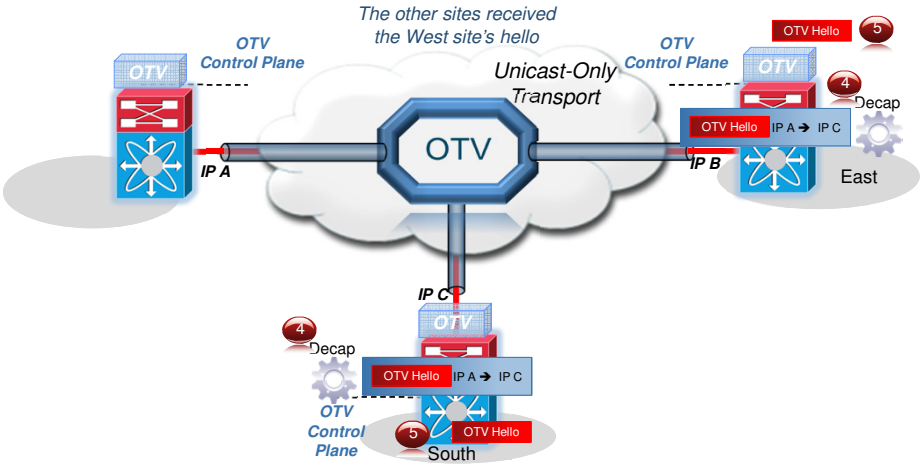
* A redundant pair may be configured

OTV Control Plane

Neighbor Discovery (Unicast-Only Transport)



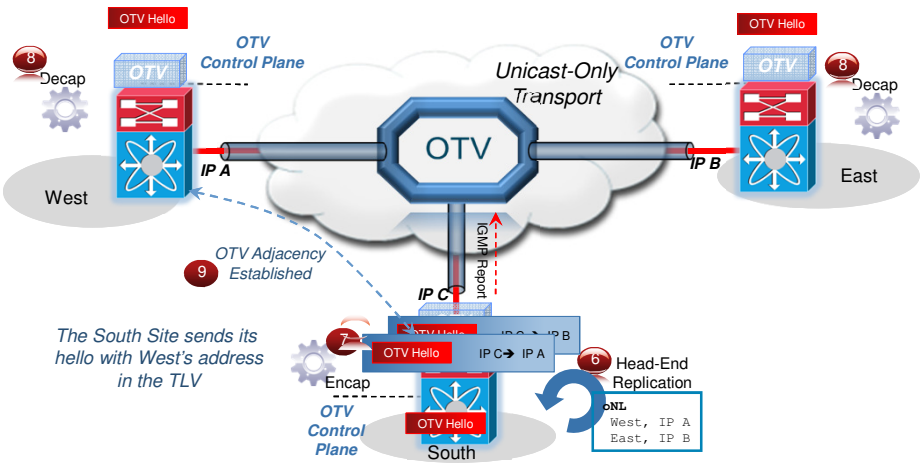
OTV Control Plane
Neighbor Discovery (Unicast-Only Transport)



OTV Control Plane
Neighbor Discovery (Unicast-Only Transport)

1. The West site sends an OTV hello.
2. The Edge Device checks the overlay Adjacency List (oNL) in order to find out how and which neighbors to reach. Once that information is found, the original hello is head-end replicated for the number of destinations that need to be reached.
3. The original hellos are then encapsulated into IP ***unicast*** packets, where the IP source and destination addresses are those of the OTV join interfaces of source and destination sites. The encapsulated packets will now be delivered to their destinations by the core.
4. The unicast encapsulated packets are received by the destination EDs, which perform the decapsulations.
5. The original OTV hello packets are delivered by the EDs to the OTV control plane process.

OTV Control Plane
Neighbor Discovery (Unicast-Only Transport)



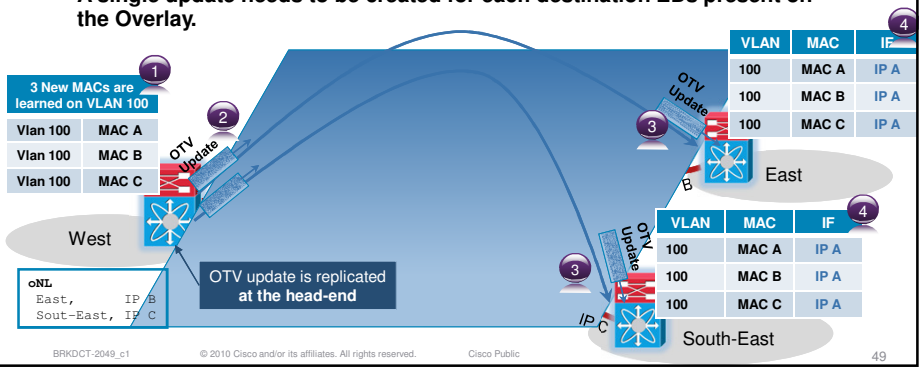
OTV Control Plane
Neighbor Discovery (Unicast-Only Transport)

6. The South Site now sends its OTV hello. In the TLV of the OTV hello the South Site will include the West Site ID. Based on the oNL, the original hello is head-end replicated for the number of destinations that need to be reached.
7. The original hellos are then encapsulated into IP unicast packets, where the IP source and destination addresses are those of the OTV join interfaces of source and destination sites. The encapsulated packets will now be delivered to their destinations by the core.
8. The unicast encapsulated packets are received by the destination EDs, which perform the decapsulations.
9. The OTV control plane on the West Site ED sees that the OTV hello just received from the South Site contains the West Side ID. This indicates to the OTV control plane that there is a two ways communication, which allows the OTV adjacency to be formed.

OTV Control Plane

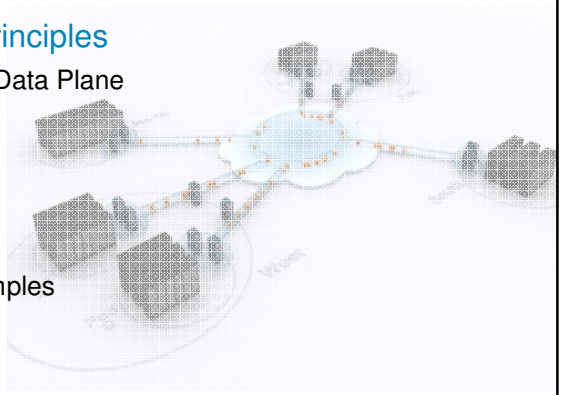
MAC Advertisements (Unicast-Only Transport)

- Every time an Edge Device learns a new MAC address, the OTV control plane will advertise it together with its associated VLAN IDs and IP next hop.
- The IP next hops are the addresses of the Edge Devices through which these MACs are reachable in the core.
- A single OTV update can contain multiple MAC addresses for different VLANs.
- **A single update needs to be created for each destination EDs present on the Overlay.**



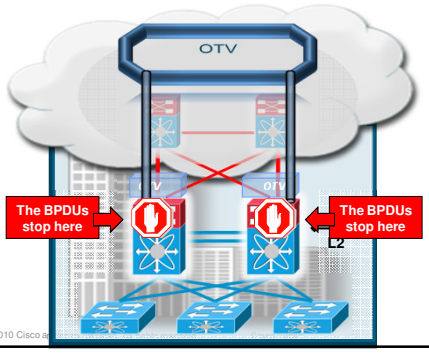
Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



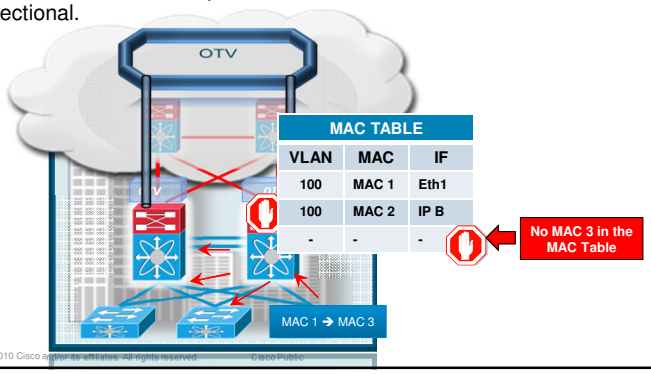
Spanning Tree and OTV
Site Independence

- OTV does not affect the STP topology of the site and in these terms OTV is totally **site transparent**.
- Each site will have its own STP domain, which is separate and independent from the STP domains in other sites, even though all sites will be part of common Layer 2 domain.
- This functionality is built-in into OTV and as such no configuration is required to have it working.
- An Edge Device will send and receive BPDUs ONLY on the OTV Internal Interfaces.



Unknown Unicast and OTV
No longer flooding storms across the DCI

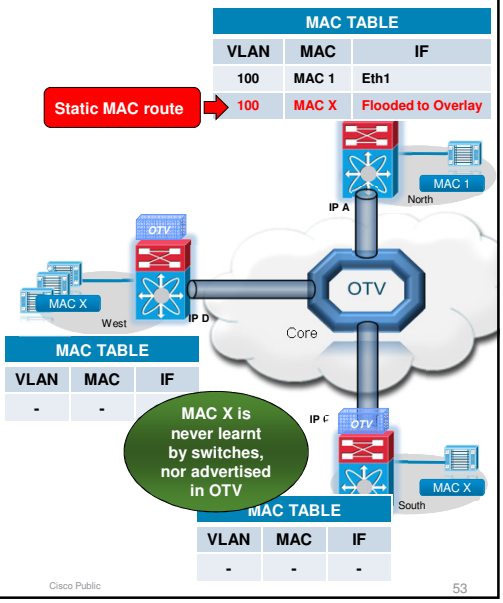
- OTV does not leverage flooding to propagate the learning of the MAC addresses across the overlay.
- No more requirements to forward unknown unicast over the overlay, therefore its forwarding is suppressed.
- Any unknown unicasts that reach the OTV edge device will not be forwarded to the overlay. This is achieved without any additional configuration.
- The assumption here is that the end-points connected to the network are not silent or uni-directional.



Selective Flood of Unknown Unicast

MSFT Network Load Balancing Services (NLBS)

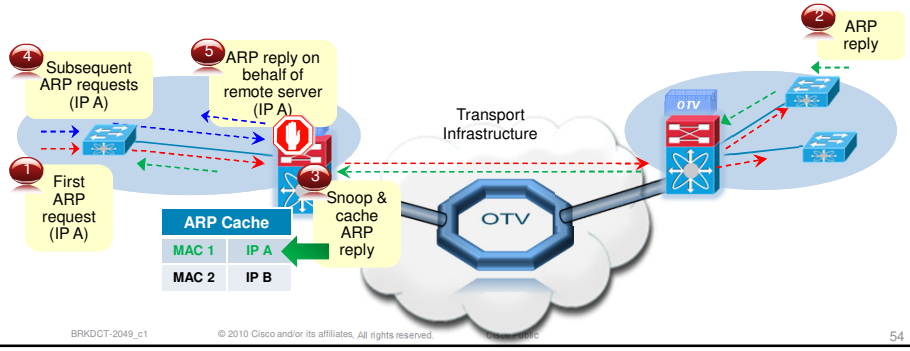
- MSFT Clustering can use unidirectional MAC addresses to force flooding to its cluster members (NLBS).
- By using “listen-only” MAC addresses, learning is prevented and flooding guaranteed.
- OTV can selectively flood traffic for specific MAC addresses in order to support this corner case.
- Flooding can be scoped to specific sites if desired.



Controlling ARP traffic

ARP Neighbor-Discovery (ND) Cache

- An ARP cache is maintained by every OTV edge device and is populated by snooping ARP replies.
- Initial ARP requests are broadcasted to all sites, but subsequent ARP requests are suppressed at the Edge Device and answered locally.
- OTV Edge Devices can thus reply to ARPs on behalf of remote hosts.
- ARP traffic spanning multiple sites can thus be significantly reduced.

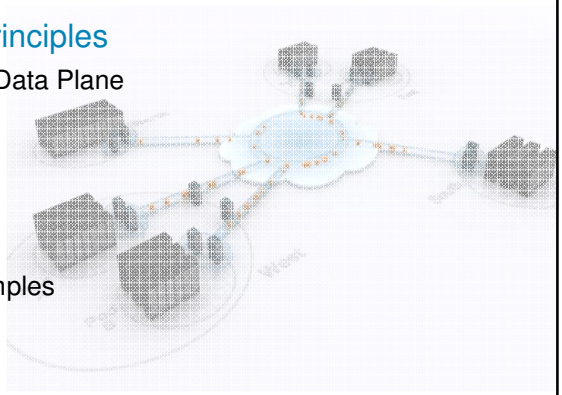


OTV solves Layer 2 Fault Propagation
Summary

- STP Isolation – BPDUs are not forwarded over the overlay.
- Unknown unicasts – not flooded across sites
  - Selective flooding is optional
- Cross site ARP traffic is reduced with ARP ND Cache.
- Broadcast can be controlled based on a white list as well as a rate limiting profile.

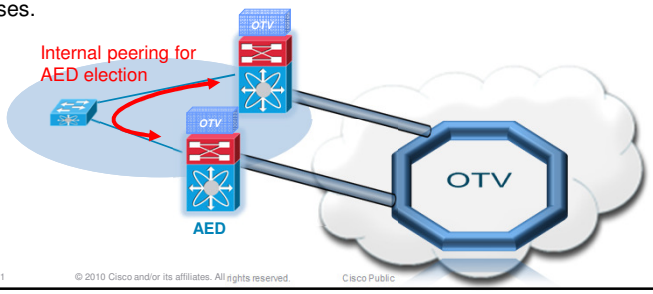
Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



Multi-homing
Per VLAN Authoritative Edge Device

- OTV provides loop-free multihoming by electing a designated forwarding device **per site for each VLAN**.
- This forwarder is known as the ***Authoritative Edge Device*** (AED).
- The Edge Devices at the site peer with each other on the internal interfaces to elect the AED.
- The peering takes place over the OTV *"**site-vlan**"*. It's recommended to use a dedicated VLAN as site-vlan.
- The assignment of the VLANs to a particular AED is all automated (though predictable) in the first release. User control will come later in future software releases.



Multi-homing
Site Merging and Partitioning

Merging:

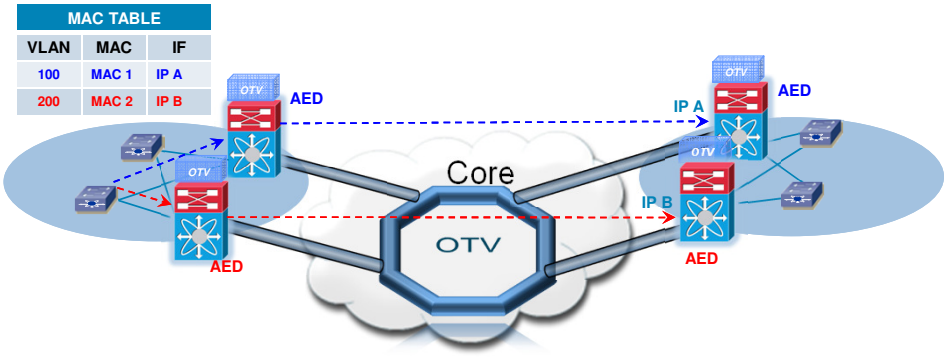
- When two or more sites merge the individual STP domains become one with one new root bridge elected.
- All OTV Edge Devices will notice each other and there will be a new AED election for each VLAN-ID range.
- When an Edge Device was authoritative and then becomes non-authoritative, it needs to remove all MAC entries that point out the overlay network from its MAC table.

Partitioning:

- When a multi-home site partitions, the internal peering will detect an adjacency loss.
- On each site there will be a site-id election.
- The new site-id is advertised by each Edge Device so the other sites detect that those Edge Devices are now in two different sites.

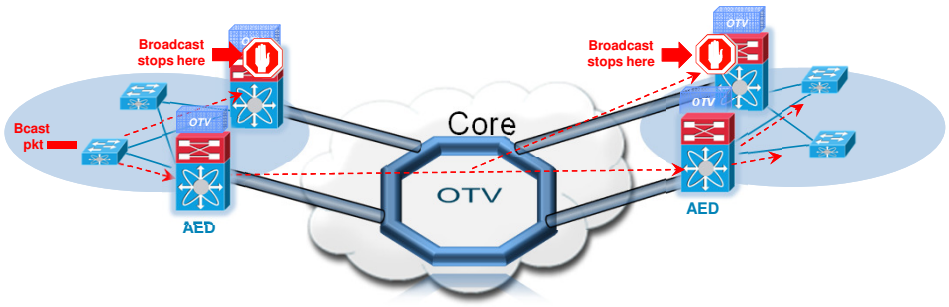
Multi-homing
Per-VLAN Load Balancing

- One AED is elected for each VLAN on each site.
- Different AEDs can be elected for each VLAN to balance traffic load.
- Only the AED forwards unicast traffic to and from the overlay.
- Only the AED advertises MAC addresses for any given site/VLAN.



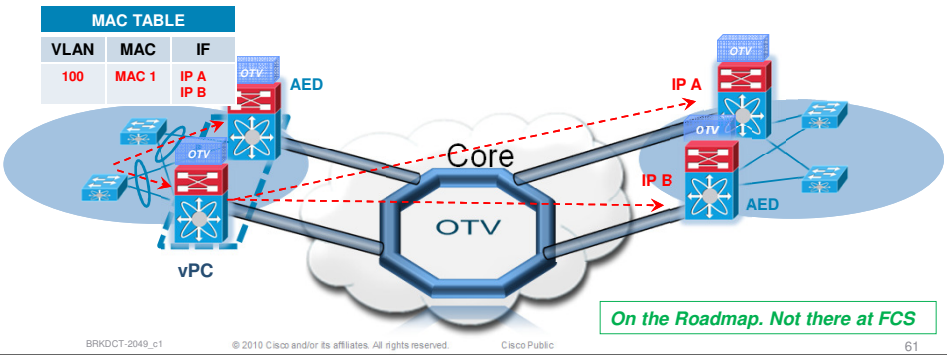
Multi-homing
AED and Broadcast/Multicast Handling

- Broadcast and multicast packets reach all Edge Devices within a site.
- The broadcast/multicast packet is **replicated to all the Edge Devices** on the overlay.
- Only the AED at each remote site will forward the packet from the overlay onto the site.



Multi-homing
Active-Active ECMP and Load Balancing

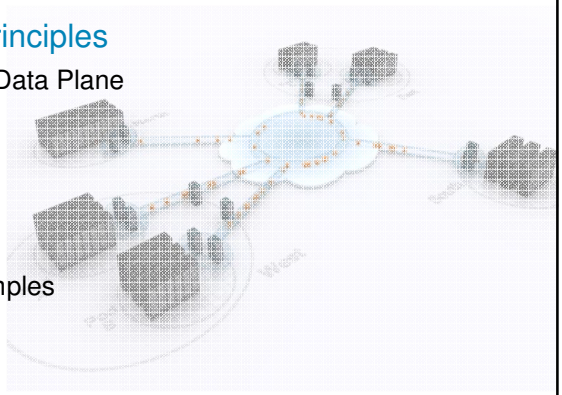
- Within a single VLAN different flows can use different edge devices on a multi-homed site.
- Choice of the edge device (and ECMP route to the remote site) is based on the source/destination addresses of the frames to be forwarded.
- All Edge Devices advertise routes to the site MAC addresses.



On the Roadmap. Not there at FCS

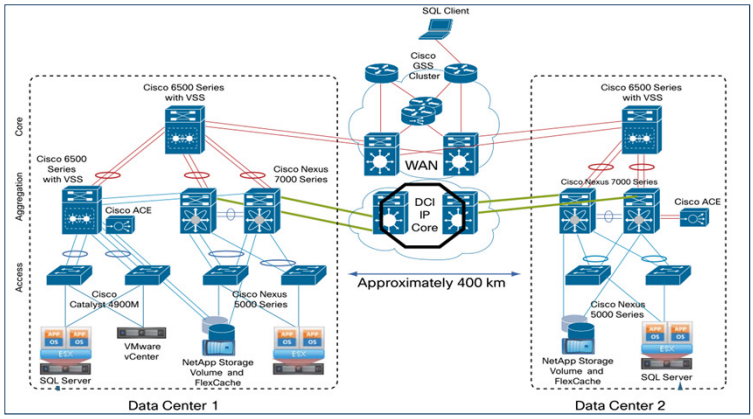
Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



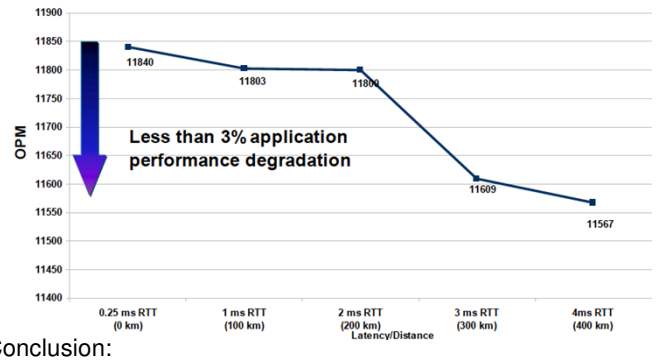
OTV and Long Distance Vmotion

- Cisco, NetApp and VMWare jointly test:
  - Two Data Centers distant ***400 Km*** from each other
  - ***OTV used to extend Layer 2 on 2 pairs of Nexus 7000***



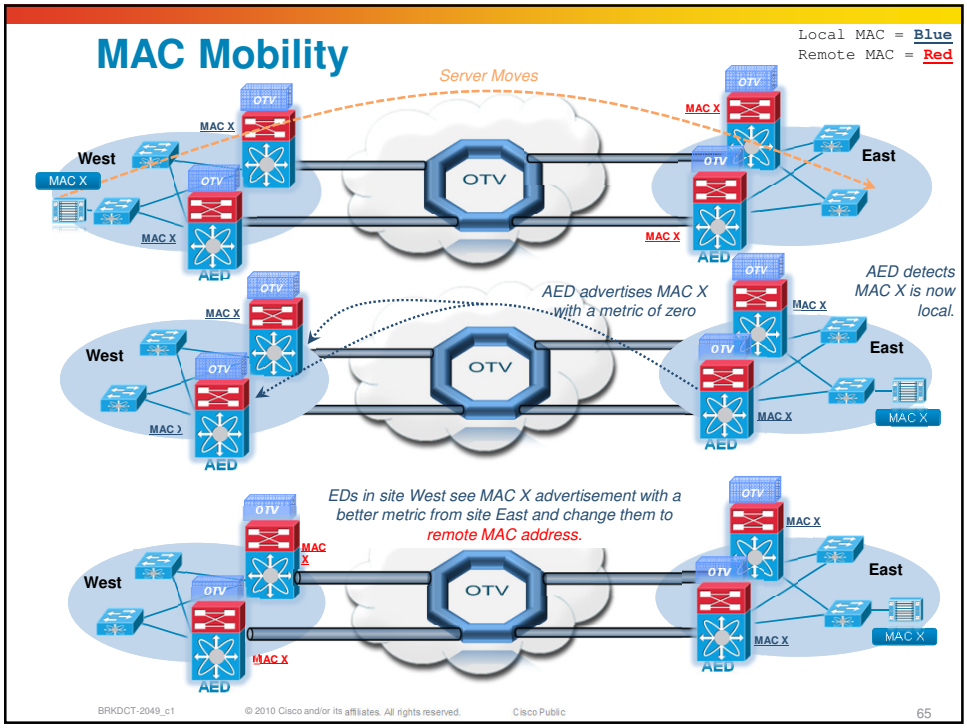
OTV and Long Distance Vmotion

- Testing GoNL:
  - Measurement of the application performance in terms of operations per minute (OPMs) due to Vmotion



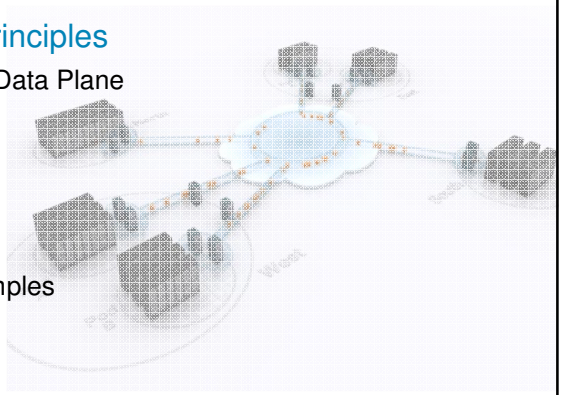
- Conclusion:
  - OTV provides a powerful mechanism for easily and flexibly extending the LAN across any type of transport network without requiring a network redesign.
  - The integration of Cisco GSS and Cisco ACE delivers crucial route optimization functions.

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9402/white_paper_c11-591960.pdf



Agenda

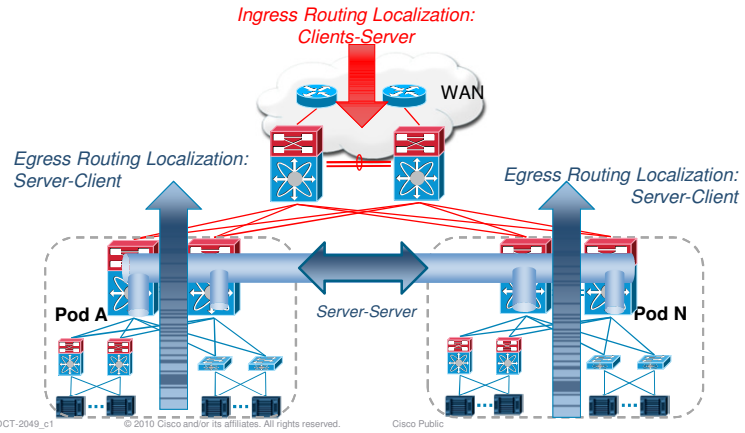
- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



Path Optimization

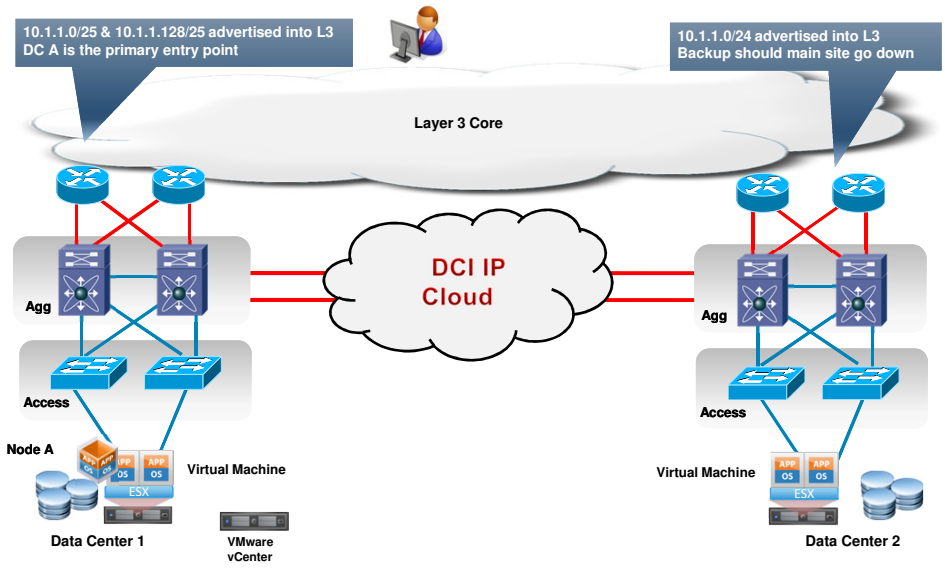
Optimal Routing Challenge

- Layer 2 extensions represent a challenge for optimal routing.
- Challenging placement of gateway and advertisement of routing prefix/subnet.

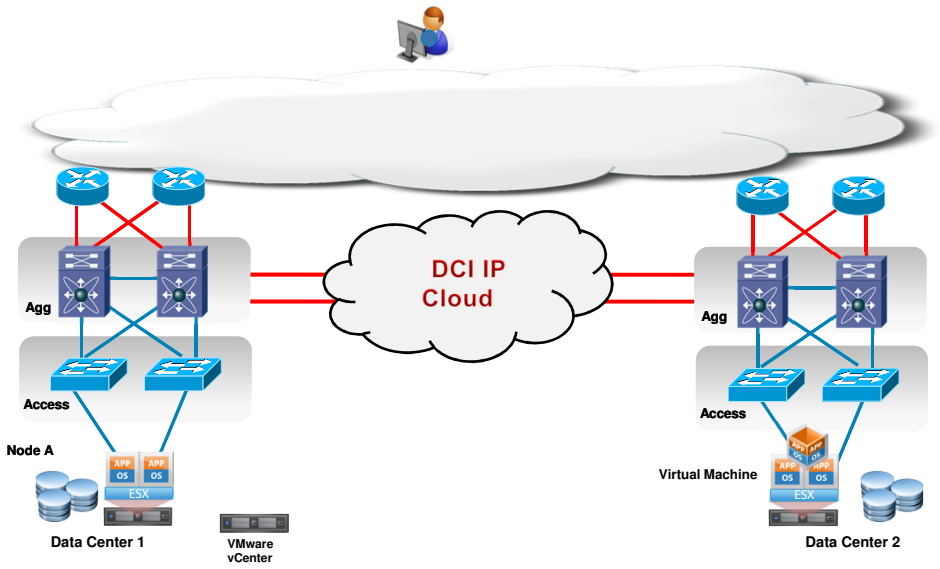


Path Optimization

What is the Problem?



Path Optimization
The Goal



Path Optimization Techniques

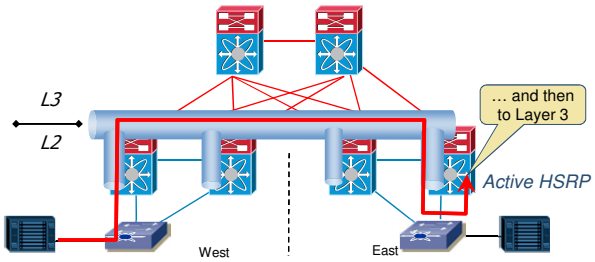
- Egress traffic
  - FHRP isolation*
- Ingress traffic
  - Anycast
    - Active/Standby subnet advertisement
  - Reverse Health Injection (RHI)
    - Host based /32 announcement
  - ACE/GSS*
    - DNS based Global Site Selection
  - Locator/ID Separation Protocol – LISP*
    - Host routing

**** Briefly discussed in this presentation***

Path Optimization

Optimal Egress Routing Challenge

- Outbound routing of traffic, i.e. Server-Client or Server-Server traffic, is dependent on the location of the server's default gateway.
- An extended subnet will have multiple IP gateway candidates distributed across sites. These gateways are all part of the same FHRP/HSRP group.

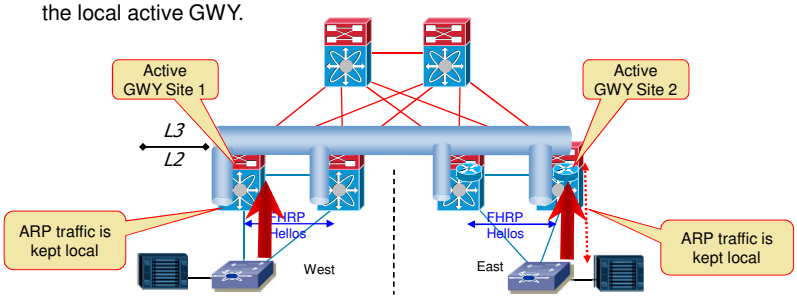


- Goals:
  - To enable site local egress routing, the default gateway must be present in the same site as the server is located.
  - For subnets/VLANs that stretch over multiple locations it means that each location has to have an active gateway.

Path Optimization

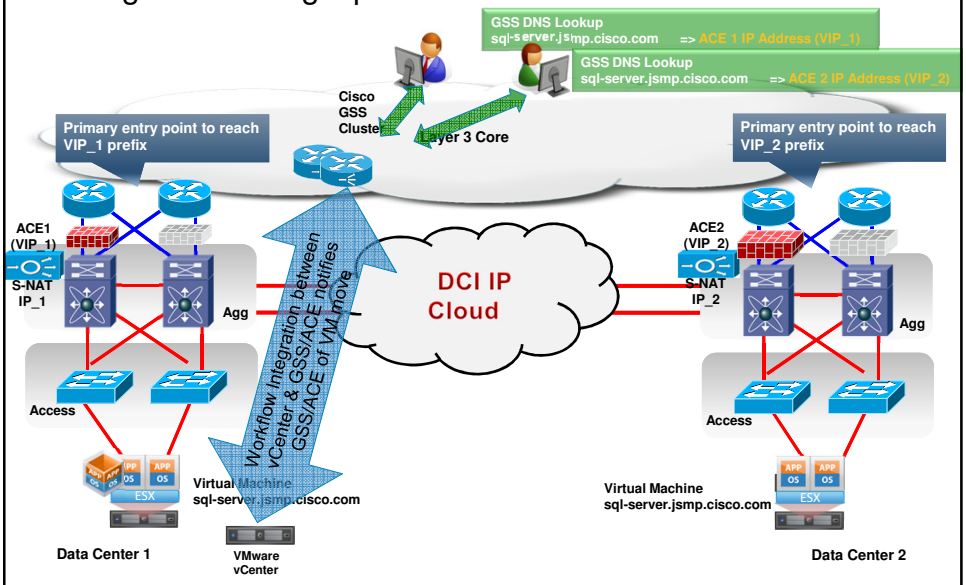
Egress Routing Localization – OTV Solution

- The approach is to use the same HSRP group in all sites and therefore provide the same default gateway MAC address.
- Each site pretends that it is the sole existing one, and provide optimal egress routing of traffic locally.
- **OTV achieves Edge Routing Localization by filtering the HSRP hello messages between the sites**, therefore limiting the "view" of what other routers are present within the VLAN.
- ARP requests are intercepted at the OTV edge to ensure the replies are from the local active GWY.



Path Optimization

Ingress Routing Optimization: ACE and GSS

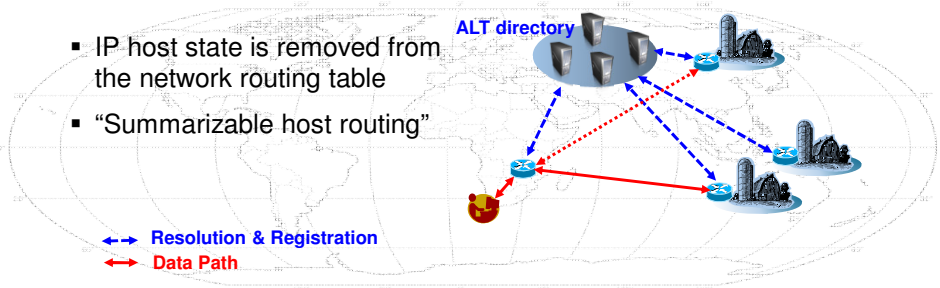


Path Optimization

Locator-ID Separation Protocol (LISP)

- Decouples the Identity of a resource (IP address) from its location
- Routing now focuses on location connectivity, not host reachability
- IP address to Location mappings are kept in a Directory
- IP addresses are resolved to a location by consulting the directory
- Traffic is IPinIP encapsulated and forwarded to the location
- The directory is a distributed and summarizable data base

- IP host state is removed from the network routing table
- "Summarizable host routing"



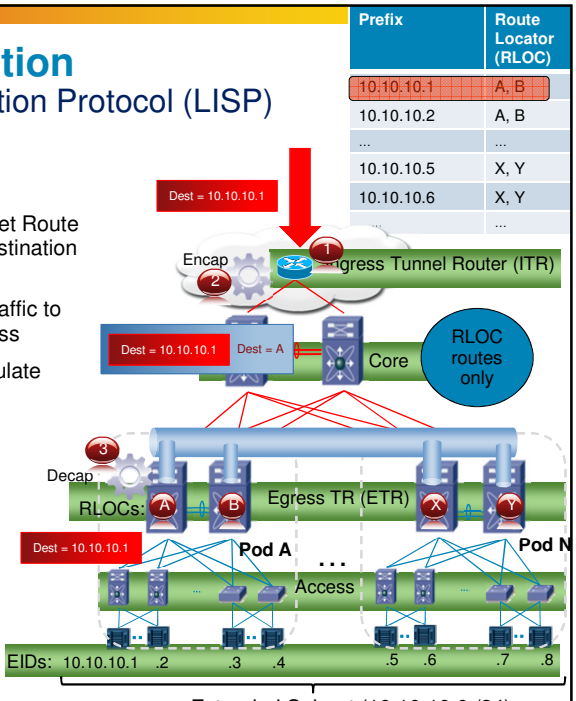
Path Optimization
Locator-ID Separation Protocol (LISP)

Host IP = End-point ID

Router IP = Route Locator

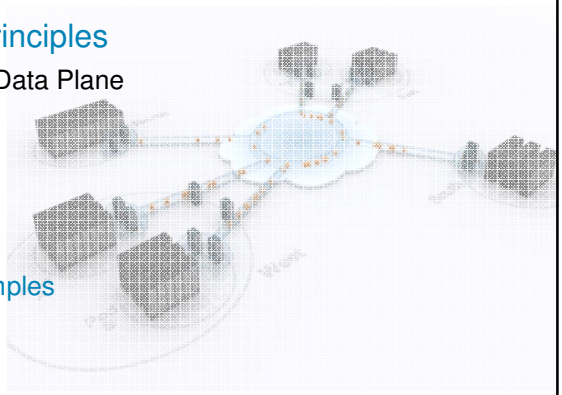
1. ITR consults directory to get Route Locator (RLOC) for the destination End-point ID (EID)
2. ITR IPinIP encapsulates traffic to send it to the RLOC address
3. ETRs receive and decapsulate traffic

- *Granular reachability information for hosts in extended subnet*
- *If a host moves, its mapping is updated*
- *No EID state in routing tables*



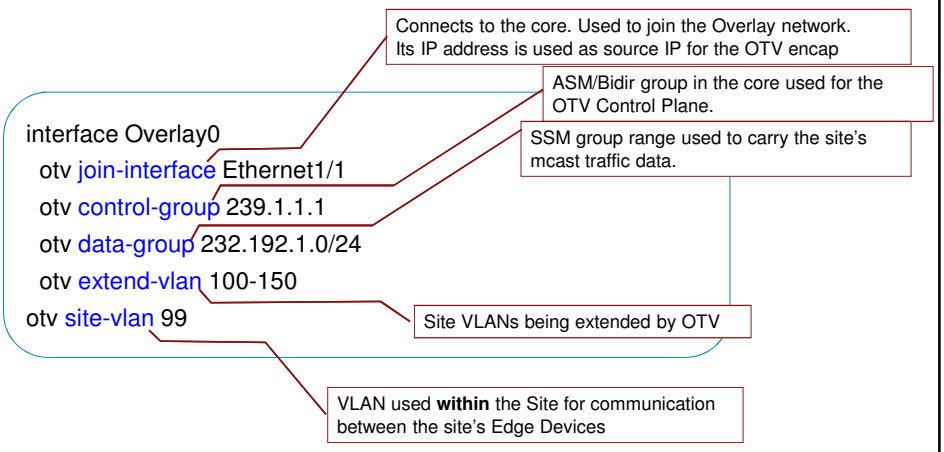
Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - Path Optimization
  - Configuration Examples
- Use Cases



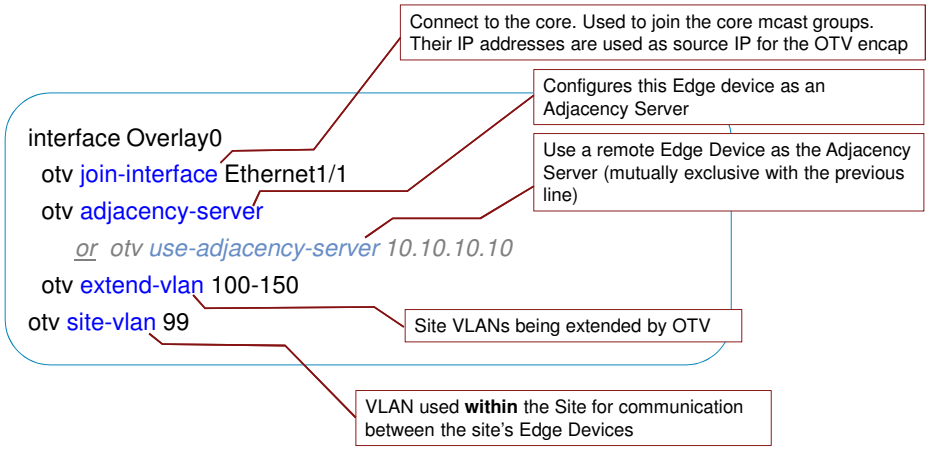
Configuration

OTV CLI Configuration (Multicast-enabled Transport)



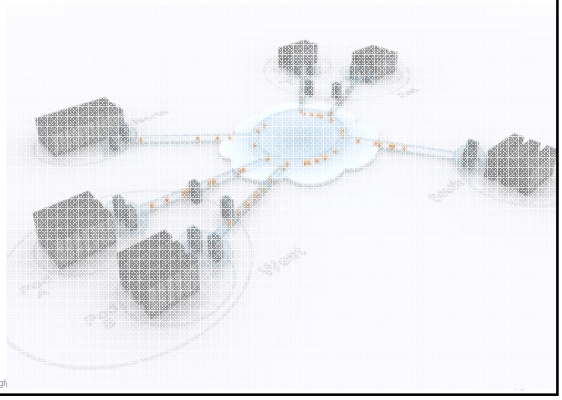
Configuration

OTV CLI Configuration (Unicast-Only Transport)



Agenda

- Distributed Data Centers: Goals and Challenges
- Traditional Layer 2 VPNs
- OTV Architecture Principles
- Use Cases



Applications That Benefit From OTV

vmware *Local Clusters*

vmotion *Local Clusters*

VERITAS BUSINESS WITHOUT INTERRUPTION *Server Clusters*

IBM *HACMP*

ORACLE *RAC –Real App Clusters*

Microsoft *Server Clusters*

EMC² where information lives *Legato Automated Availability Mgr*

NetApp *Metro Cluster*

BACnet *Building Automation Control*

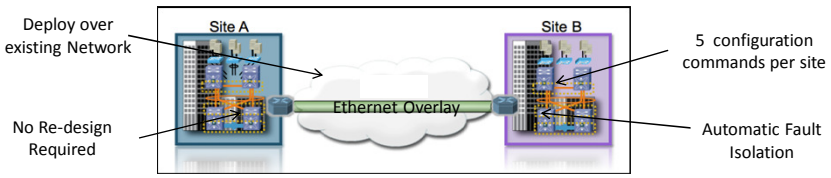
hp invent *Metro Cluster*

OTV Use Case: DC Growth Constraints

Problem = Primary data center maxed out : space, cooling and power

Requirement = Extend clusters and workload across data centers

Challenge = Rapidly establish Data Center Interconnect between data centers

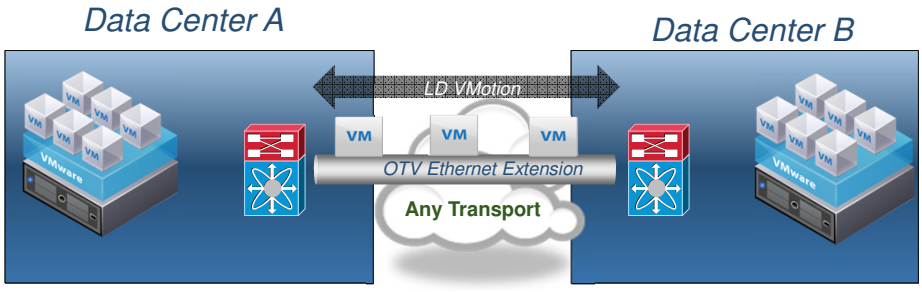


Solution: OTV – Establish DCI in 5 minutes!

- No new transport provisioning required (*Dark fiber, MPLS, etc*)
- Eliminate months of re-design effort
- Significant operations and provisioning cost savings *(no new protocols)*

OTV Use Case: Vmotion

Live migration of VMs from one data center to another

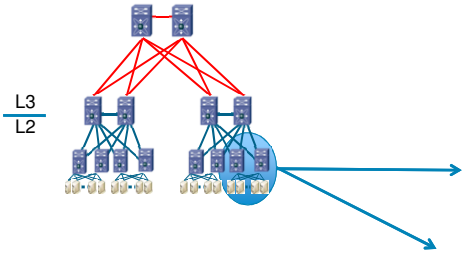


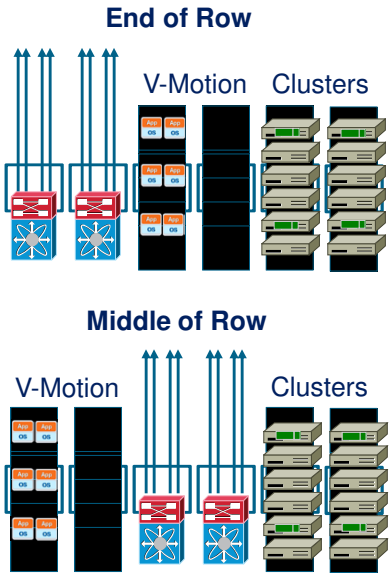
"Moving workloads between data centers has typically involved complex and time-consuming network design and configurations. VMware VMotion™ can now leverage Cisco OTV to easily and cost-effectively move data center workloads across long distances, providing customers with resource flexibility and workload portability that span across geographically dispersed data centers."

"This represents a significant advancement for virtualized environments by simplifying and accelerating long-distance workload migrations."

Ben Matheson, senior director, global partner marketing, VMware.

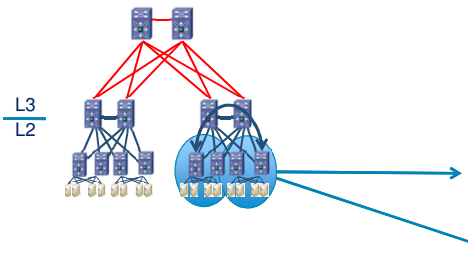
OTV Use Case: Vmotion and Clustering
Bound by Layer 2

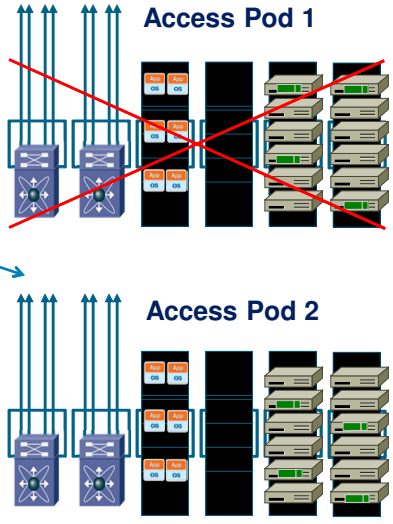




- Clusters and VMotion operate well within Layer 2
- Build larger Layer 2 networks for improved access layer load balance

OTV Use Case: Unbinding Vmotion and Clustering





- Clusters, VMotion require Layer 2 extensions to go across access pods
- Improves Manageability
- Dynamic Annexation
- Portability & Expansion

Conclusion

Now that you know how it works…

- Make sure to learn and follow the Cisco design guidelines to deploy OTV successfully.
- First Step:

  BRKDCT-3060
  Deployment challenges with Interconnecting Data Centers.

  BRKDCT-2840
  Data Center Networking: Taking Risk Away from Layer 2 Interconnects

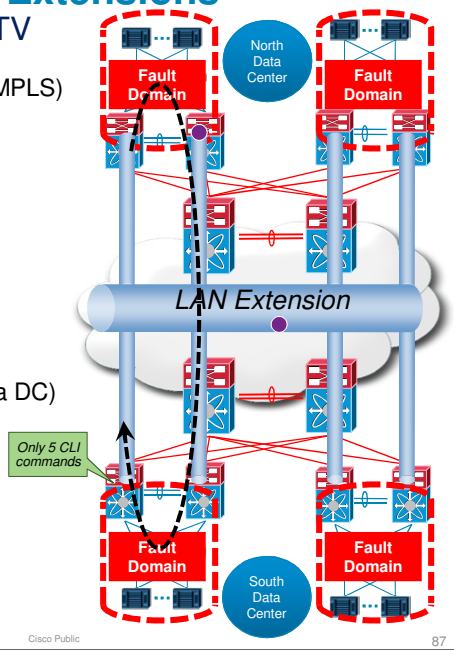
- Next:

  Check out our DCI page on cisco.com:
  http://www.cisco.com/en/US/netsol/ns975/index.html

Challenges with LAN Extensions

Real Problems Solved by OTV

- Extensions over any transport (IP, MPLS)
- Failure boundary preservation
- Site independence / isolation
- Optimal BW utilization (no head-end replication)
- Resiliency/multihoming
- Built-in end-to-end loop prevention
- Multisite connectivity (inter and intra DC)
- Scalability
  - VLANs, sites, MACs
  - ARP, broadcasts/floods
- Operations simplicity



Related Networkers Sessions

- **BRKDCT-3060**
  Deployment challenges with Interconnecting Data Centers.
- **BRKDCT-2840**
  Data Center Networking: Taking Risk Away from Layer 2 Interconnects
- **BRKDCT-2048**
  Deploying Virtual Port Channel in NXOS
- **BRKDCT-1022**
  Introduction Cisco Layer 2 Multipathing (L2MP)

Complete Your Online Session Evaluation

- Give us your feedback and you could win fabulous prizes. Winners announced daily.
- Receive 20 Cisco Preferred Access points for each session evaluation you complete.
- Complete your session evaluation online now (open a browser through our wireless network to access our portal) or visit one of the Internet stations throughout the Convention Center.

Don't forget to activate your Cisco Live and Networkers Virtual account for access to all session materials, communities, and on-demand and live activities throughout the year. Activate your account at any internet station or visit www.ciscolivevirtual.com.

Enter to Win a 12-Book Library of Your Choice from Cisco Press

Visit the Cisco Store in the World of Solutions, where you will be asked to enter this Session ID code

Check the **Recommended Reading** brochure for suggested products available at the Cisco Store