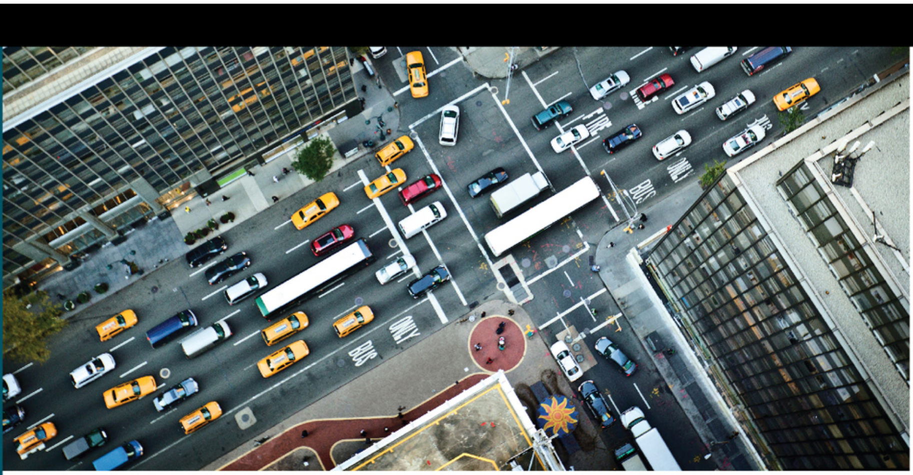

The LISP Network

Evolution to the Next Generation of Data Networks

Victor Moreno
Dino Farinacci

ciscopress.com

The LISP Network

Evolution to the Next Generation of Data Networks

Victor Moreno

Dino Farinacci

Cisco Press

Switching fabrics use network-based overlays to deliver a large number of virtual networks. The indirection between identity and location used in network overlays is key in supporting the required host mobility in the fabric. LISP includes the necessary mechanisms to support fast mobility of hosts while keeping both the mapping database and the Map-Caches up to date as hosts move around the fabric. The overlay data plane headers are leveraged to carry the metadata necessary for the segmentation of the network as well as the metadata required to enforce security policies and define service chains.

LISP is particularly well suited for the role of the overlay control plane in a switching fabric. The overlay control plane has a simple mission: maintain the mappings of endpoint-identity to their location. In the overlay control plane, there isn't a calculation of best routes, redundant paths, or loop resolution. These are all routing tasks that are fully realized in the underlay but are not required in the overlay. Furthermore, it is desirable to extend the database of endpoint-identity and location mappings to include policy, geo-location, and other information that may be relevant to the endpoints. All in all, the overlay control plane is in charge of populating and updating a directory of information that is relevant to the endpoints. At no point does the overlay control plane execute routing calculations. Therefore, it makes sense to use a protocol designed to handle such a directory versus trying to extend a routing protocol to do this database handling. The requirements of the overlay control plane call for a directory service such as LISP while all routing intelligence remains in the underlay, where well-established routing protocols like Open Shortest Path First (OSPF) and Border Gateway Protocol (BGP) excel at optimizing the reachability between routing locators (RLOCs). One salient characteristic of LISP in this area is the fact that it is demand based, allowing network nodes to request information only when they need it, and allowing the network nodes to be specific about the information they request. This capability has obvious scalability benefits but also leads to an interesting notion of contextual mapping database lookups in which the response from the LISP Mapping Database System varies depending on the context of the requestor. For example, if a mapping for an endpoint is requested from one location versus another, the reply may be different in each case. This notion of contextual lookups is discussed later in this chapter in the section, “Policy: The Network as an Enforcer.”

Due to operational familiarity, the industry's first wave of fabric development gravitated toward the use of traditional routing protocols for the overlay control plane. One popular overlay being promoted in the data center uses BGP with the EVPN address family to create virtual networks with mobility. As you will learn throughout the book, LISP presents clear benefits when used to create overlays in fabrics, and many implementations are tapping into these benefits. Cisco Application Centric Infrastructure (ACI) uses a mechanism with demand properties closer to those provided by LISP, and Cisco offers switching fabric implementations for the access networks in campus and branch that are fully based on LISP. Some of the benefits achieved by the demand-based model include scale, high rate mobility, embedded policy, and the simplification of the network control plane.

Optimizing Connectivity to the Data Center with LISP

LISP is frequently used to steer traffic to and from the data centers. It is common practice to deploy data centers in pairs to provide resiliency. When data centers are deployed in pairs, both facilities are expected to actively handle client traffic, and application workloads are expected to move freely between the data centers. The same situation arises in hybrid cloud deployments where workloads are relocated between different data center facilities. Figure 3-3 illustrates common multi–data center scenarios. Note that disaster recovery facilities also require traffic to be routed to them.

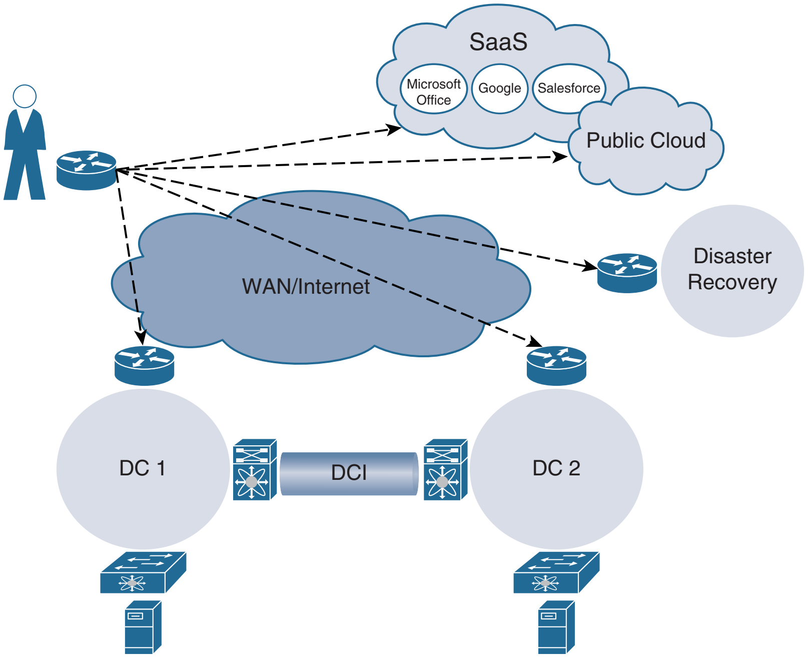

Figure 3-3 *Multi–Data Center Deployments*

Regardless of the type of multi–data center deployment, the IP address of the hosts for the applications does not change when the application workload is moved, so some hosts of a subnet are located in one data center and other hosts in another. This means that the IP subnet in which a host resides is no longer representative of one data center location or another. To maintain an acceptable application response time for users, the connection between the client and the application must follow the most direct path to the data center where the active host for the application resides. Figure 3-4 illustrates the optimal path to the active workload in contrast with suboptimal triangulated paths that result without host granularity in the routing.

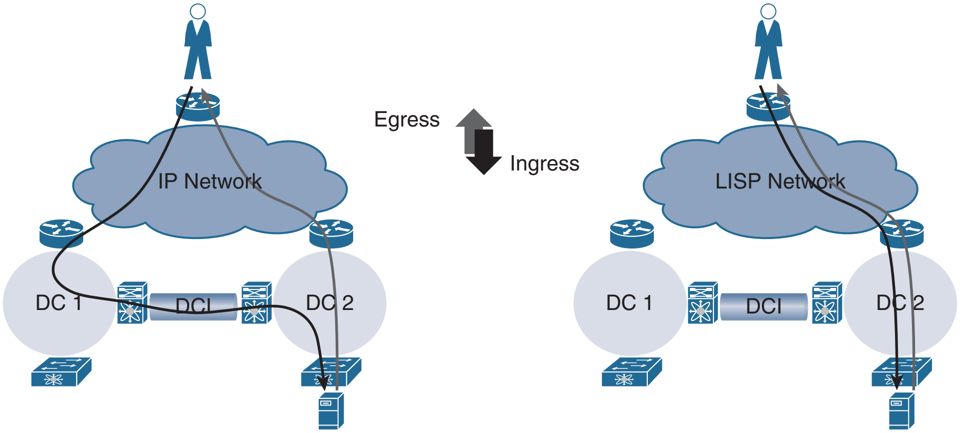

Figure 3-4 *Optimal versus Suboptimal Path to the Data Center*

The triangular pattern shown in Figure 3-4 assumes that the Layer 2 segments hosting the subnets are stretched across the different data centers. This configuration in itself poses an operational and design challenge that is addressed in the following section, where we make the case for mobility without subnet extension. Some network architects believe that the latency added by the triangular traffic path shown in Figure 3-4 may be negligible and therefore not worth addressing. In reality, the latency is the compound effect of the reachability at every layer of the application. Figure 3-5 shows the resulting traffic path when an entire application stack is not maintained in a common location. Figure 3-5 illustrates how the added latency from the triangular traffic path is multiplied. In a pair of data centers that are 10 ms apart, the latency impact on application response multiplies rapidly. This situation must be avoided, and to do so, you need to be able to move all members of an application stack to any data center without relying on triangulation to reach the workloads.

The implication is that traffic must be granularly routed to the different data centers based on host-specific location information. So, the question has been: how do you achieve host granularity in routing at scale? Traditional routing protocols require the dissemination of this host-level state across a large portion of the network, impacting scale and convergence severely. LISP provides the distinct advantage of handling granular host state for the applications in the data center by distributing the host state selectively only to the locations that need the state. By using LISP to handle host reachability outside the data center, address any scale and convergence concerns related to handling a large amount of host routes in the access and wide-area network while optimally routing to the data center where the application actually resides.

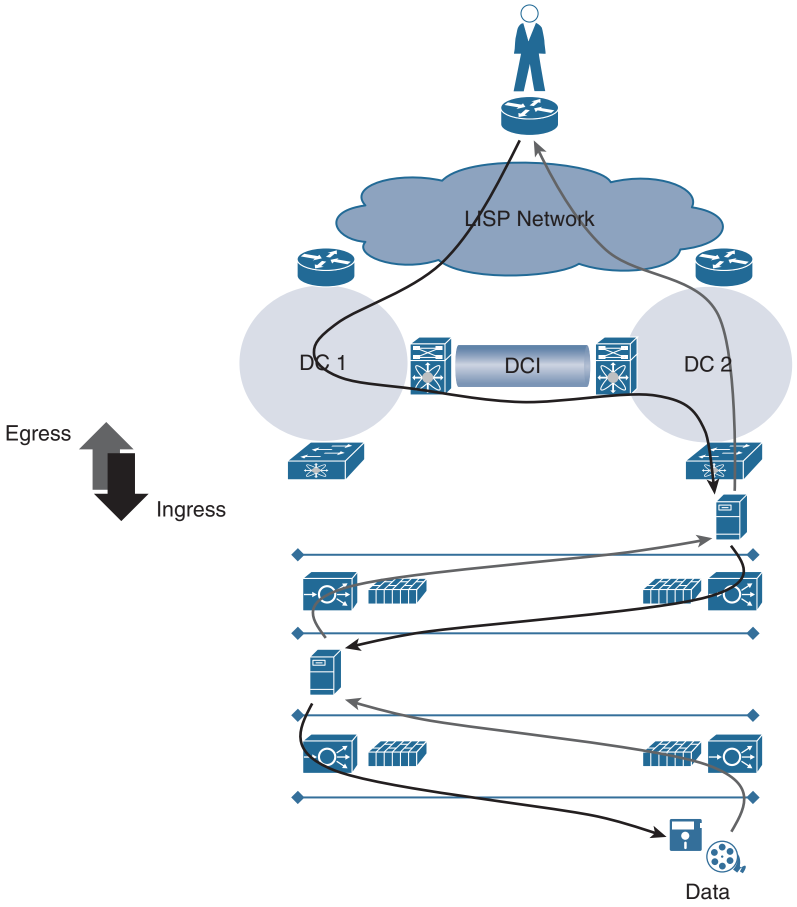

Figure 3-5 *Suboptimal Traffic Across Application Layers Between Data Center Facilities*

Some applications require very high availability. They are deployed in resilient data centers that are close enough to each other to allow the high availability models of the application to be deployed across the multiple data centers while adding resiliency to the network and facilities layer where the two data centers do not share any fate from the networking perspective. As the active application instance moves from a host in one data center to a host in another, the sessions from the clients consuming the applications must be preserved before and after the move. So not only do you need to preserve IP addresses for the moving hosts, but you also need to provide consistent Layer 2 semantics across the different network locations so that the Address Resolution Protocol (ARP)/Neighbor Discovery (ND) state of the moving host remains valid after the move and connections are maintained. One simple way of doing this is to stretch the Layer 2 environments across the different locations, but let's look at how to avoid extending Layer 2 and simplify the way to handle moving hosts.

Mobility: Subnets Really Don't Work

As applications are migrated and load balanced within the data center or across multiple data centers, the IP address of the application hosts must be preserved. Thus, any IP address should be reachable anywhere it connects. With enough entropy, any IP address may show up anywhere, and IP addresses are no longer summarized into prefixes that align with the network topology. So, a switching fabric and the network connecting the clients to the data center must be able to handle a large amount of nonsummarizable host information. Furthermore, LISP allows the preservation of these IP addresses without requiring the extension of the Layer 2 domains that traditionally support the dissemination of these addresses to multiple locations.

The demand nature of LISP allows it to handle reachability state with host-level granularity while minimizing the amount of forwarding table memory required on the LISP tunnel routers (xTRs). At the same time, the underlay remains stable and unaffected while all this mobility takes place. The large amounts of state required to handle mobility are maintained in the LISP Mapping Database System, which uses low-cost memory and scales horizontally as required. Forwarding memory at an xTR is used only to cache the reachability information relevant to flows established with its locally connected hosts.

To support mobility, a LISP site must support the following:

- The moved host must be detected.
- The local first hop routing must be adjusted to handle the new location of the host.
- Registrations for the host in the LISP Mapping Database System must be updated with the new location of the host.
- The xTRs that have cached mappings for the endpoint identifier (EID) of the host must also be updated.

A moved host is detected through many mechanisms. In a LISP network, the local xTR must become aware of the move. When hosts are connected directly to the LISP xTRs, an xTR may glean data plane events and through a simple heuristic determine whether the packets correspond to a moved host. Alternatively, the host can be connected to a non-LISP fabric (for example, a BGP-EVPN fabric), and host routes can be reported to the LISP xTR that had not been previously reported to the particular xTR, thus indicating a move (or a new host). It is also possible to identify the mobility of a host by connecting to the APIs of the host orchestration software, which often signals the different stages of a host move.

When a host move is detected, the xTR registers the host with the LISP Mapping Database System. The LISP Mapping Database System also notifies the old xTR where the host used to be located. In any mobility solution, it is critical to signal the site from which the mobile host departed so that the site can clean up its forwarding tables accordingly and adjust its state to reflect the fact that a host that was otherwise local is now remote. In the particular case of LISP, the knowledge of the move at the old site is used to notify any sender who is still sending traffic to the old location that it needs to update its information. This is how LISP updates the Map-Cache of any active senders to the moving destination. In some of the existing implementations, the data traffic from the sending ingress tunnel router (ITR) triggers a solicit-map-request (SMR) message from the old egress tunnel router (ETR). This message tells the ITR to issue a new Map-Request for the destination and the lookup process eventually results in the refresh of the ITR's Map-Cache. Alternative mechanisms may be used in LISP implementations that allow subscriptions to mappings. In such systems, the Mapping Database System keeps a log of which ITRs are sending traffic to specific destinations and updates the ITRs if the registration for the destinations changes.

When a host move is detected, the xTR at the arrival site also installs a local route to the host so that traffic destined to the host and received over LISP is forwarded to the host after it is de-encapsulated. This route is a directly connected route or a route learned over a dynamic routing protocol, depending on whether the host is directly connected to the xTR or if the host was learned from an adjacent non-LISP fabric. Figure 3-6 illustrates this concept of directly connected hosts versus hosts connected to the xTR through a fabric.

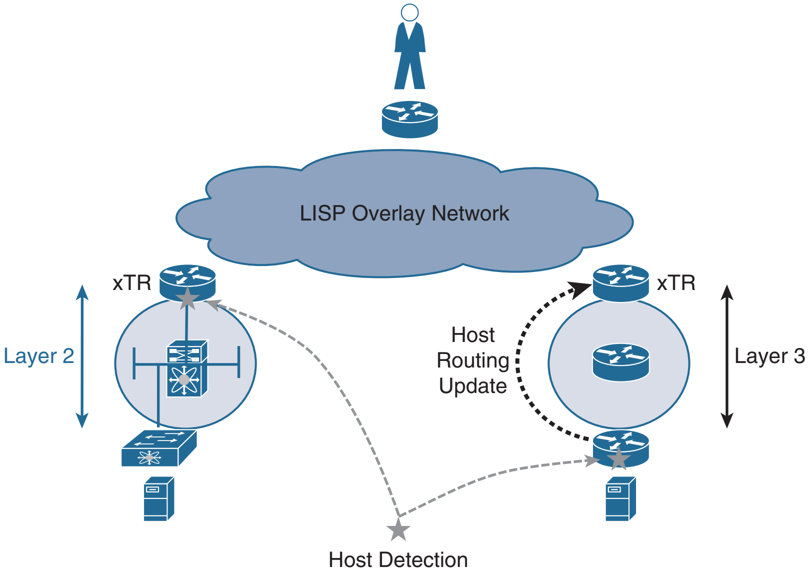

Figure 3-6 *Host to xTR Connectivity Options*

When the hosts are directly connected to the LISP xTR, the xTR may rely on a Layer 2 extension to address reachability of remote hosts on the same subnet as the moved host. This basically means that the traffic is switched into an L2 connection. This L2 connection may be in the form of one of the following:

- A VLAN
- An OTV extension
- A VXLAN tunnel
- An L2 LISP tunnel