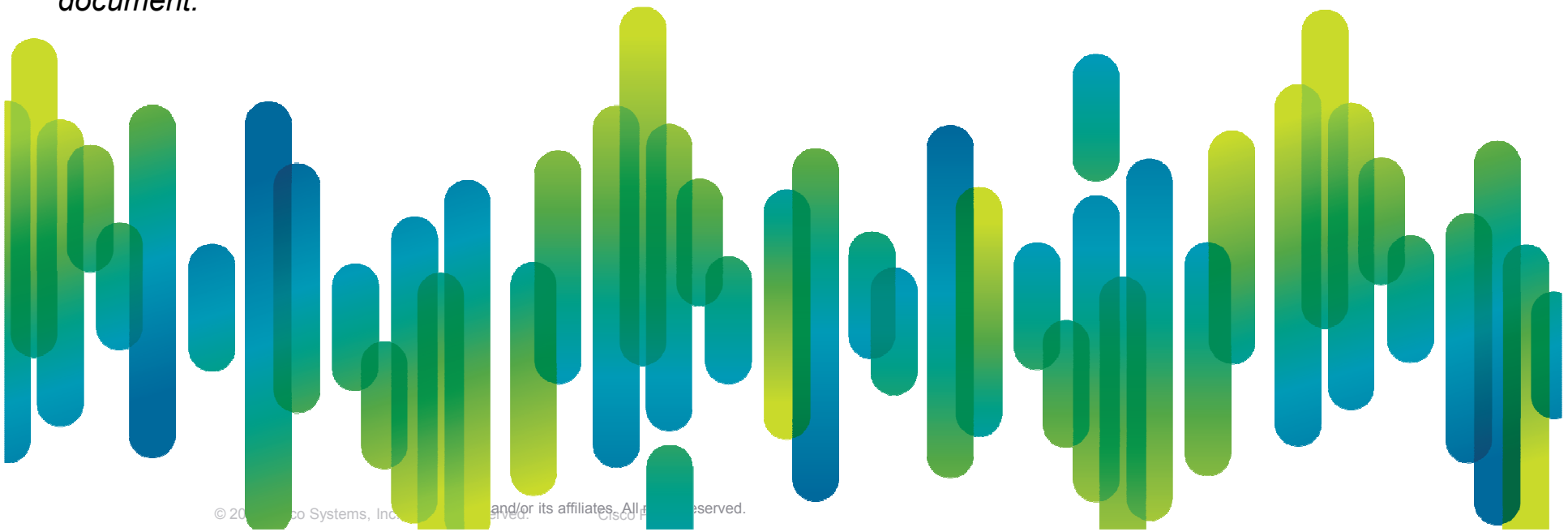
An Architectural Solution to Multi-Homing, Traffic Engineering, and Internet Route Scaling

Unter der Verwendung von Folien von
Dino, Gregg Schudel, Viktor Moreno,
Darrel Lewis und weiteren Kollegen

LISP

Dipl.-Ing. Andreas la Quiante, TSA CCIE 7817

“Many of the products and features described herein remain in varying stages of development and will be offered on a when-and-if-available basis. This roadmap is subject to change at the sole discretion of Cisco, and Cisco will have no liability for delay in the delivery or failure to deliver any of the products or features set forth in this document.”



Agenda

- Der Anfang und etwas Motivation
- Das Konzept
- Die “unicast data plane “
- Der “mapping database” Mechanismus
- Die Erreichbarkeit der Locator IDs
- Sicherheit und Management
- Spec References
- *** Use Cases ***
- *** Implementation and Deployment Status ***
- Q & A

Problem Statement

- What provoked this?

  Stimulated by problem statement effort at the Amsterdam IAB Routing Workshop on October 2006

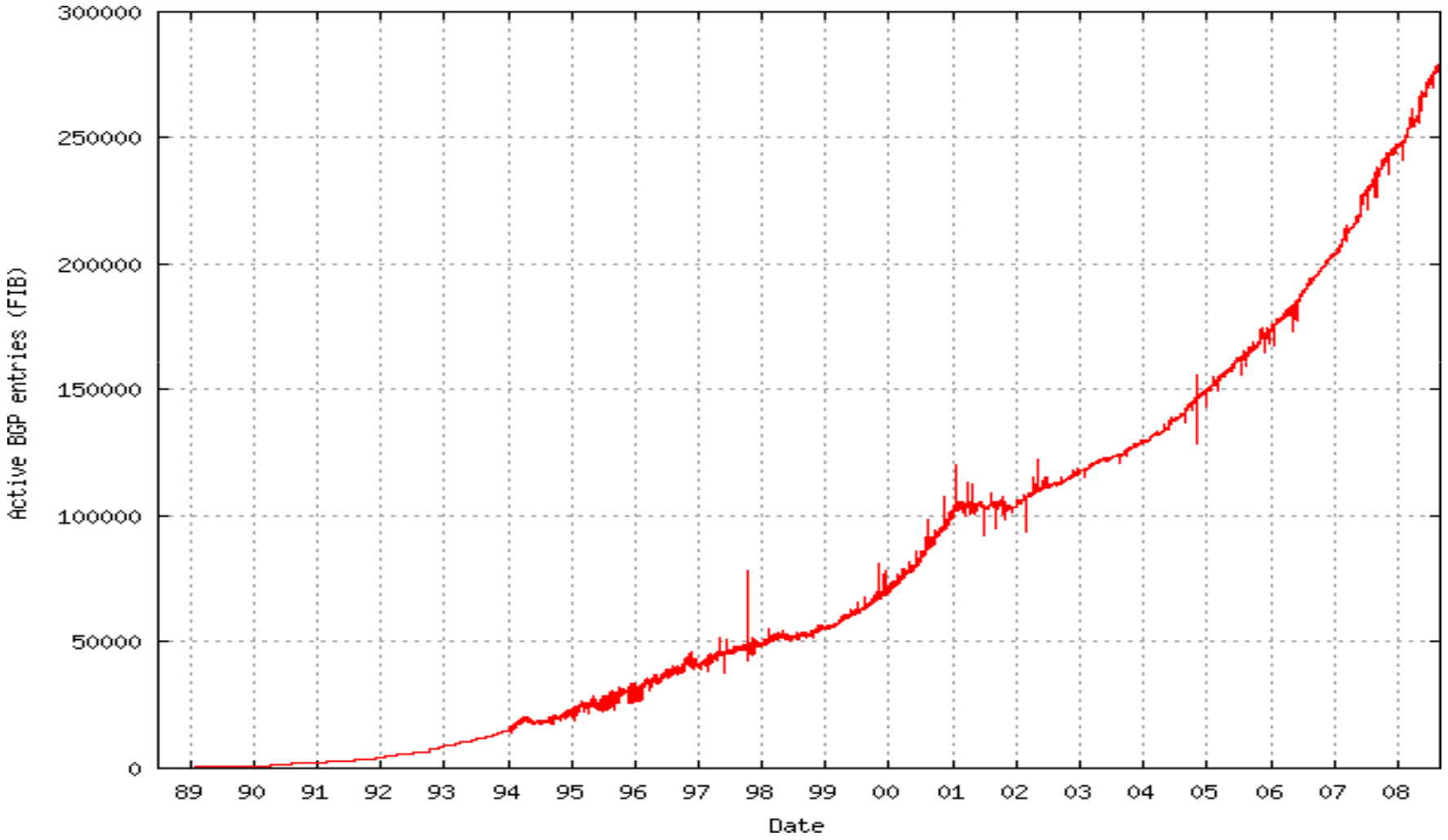
  `RFC 4984`

  More info on problem statement:

  `http://www.vaf.net/~vaf/apricot-plenary.pdf`

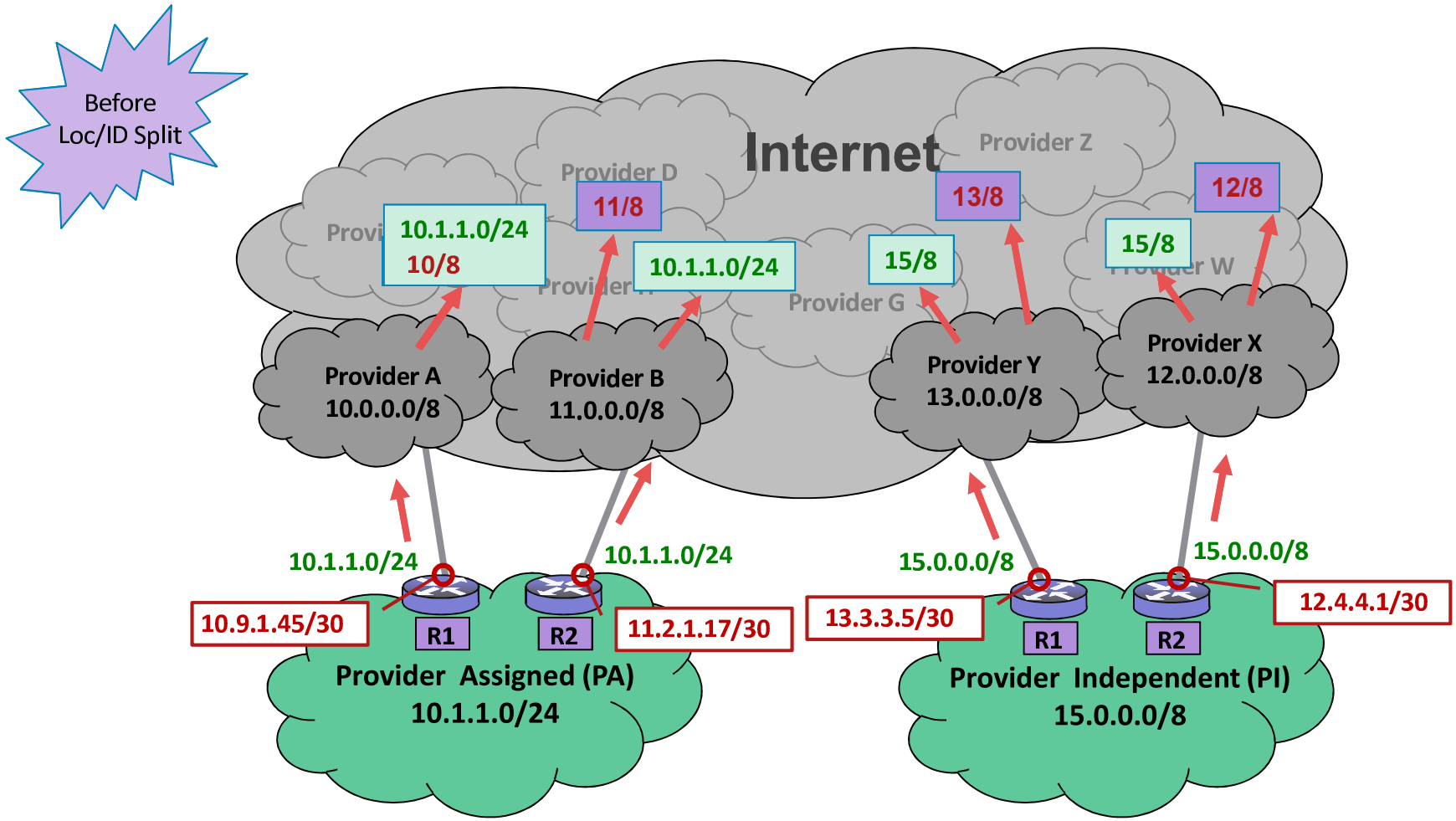
- First and foremost - scale the Internet

Scaling Internet Routing State



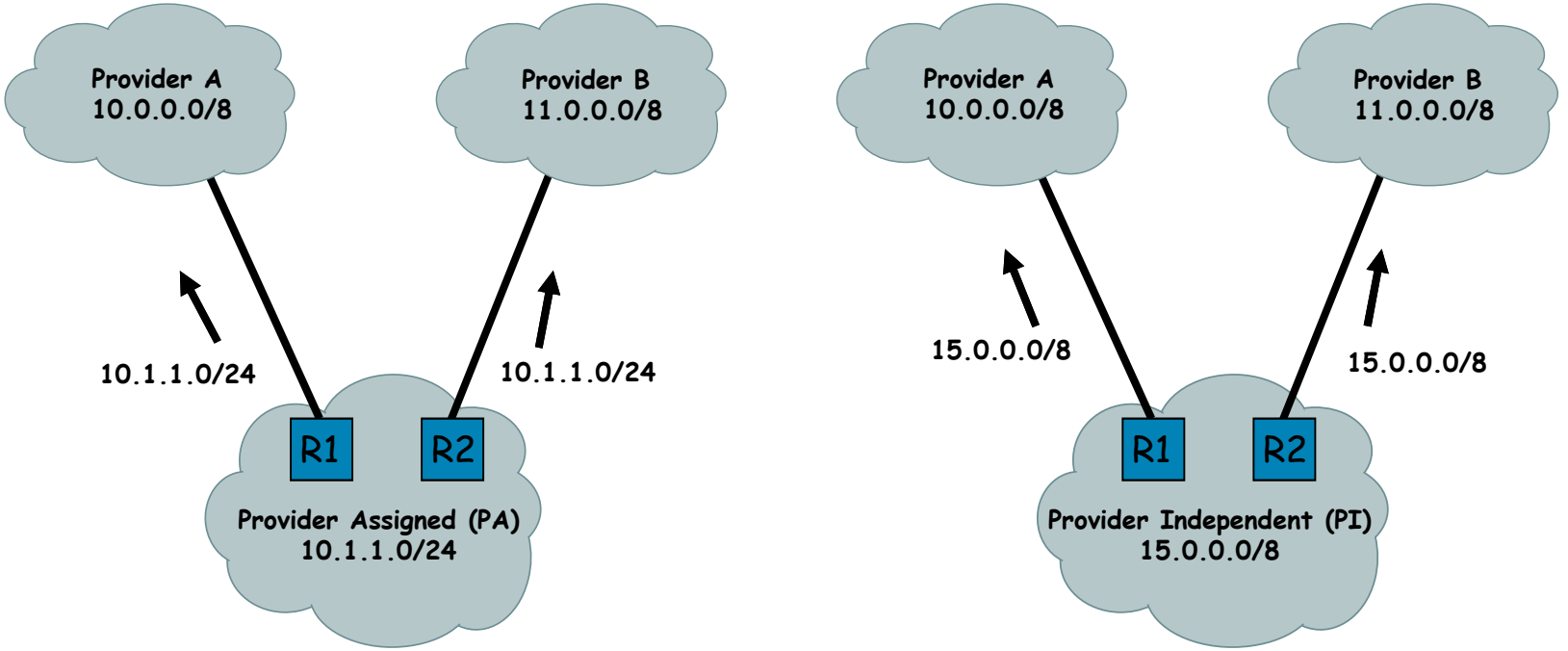
LISP Overview

What pollutes the DFZ today?

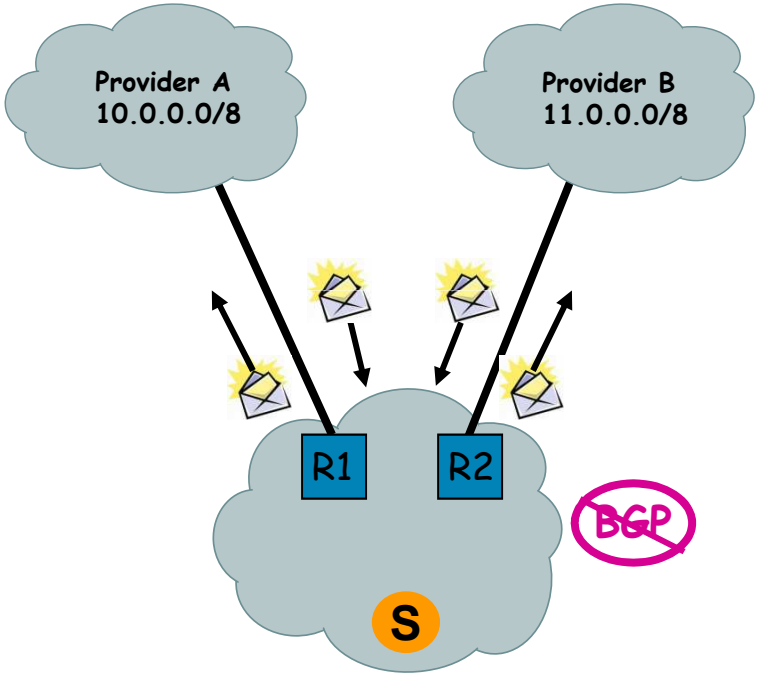


- Addresses at sites, both PA and PI, can get de-aggregated by multi-homing
- Aggregates for infrastructure addresses (e.g. CE-PE links) get advertised as well

What Pollutes the Internet



Foster Growth in Multi-Homing



1. Improve site multi-homing
   a. Can control egress with IGP routing
   b. Hard to control ingress without more specific route injection
   c. Desire to be low OpEx multi-homed (avoid complex protocols, no outsourcing)
2. Improve ISP multi-homing
   a. Same problem for providers, can control egress but not ingress, more specific routing only tool to circumvent BGP path selection

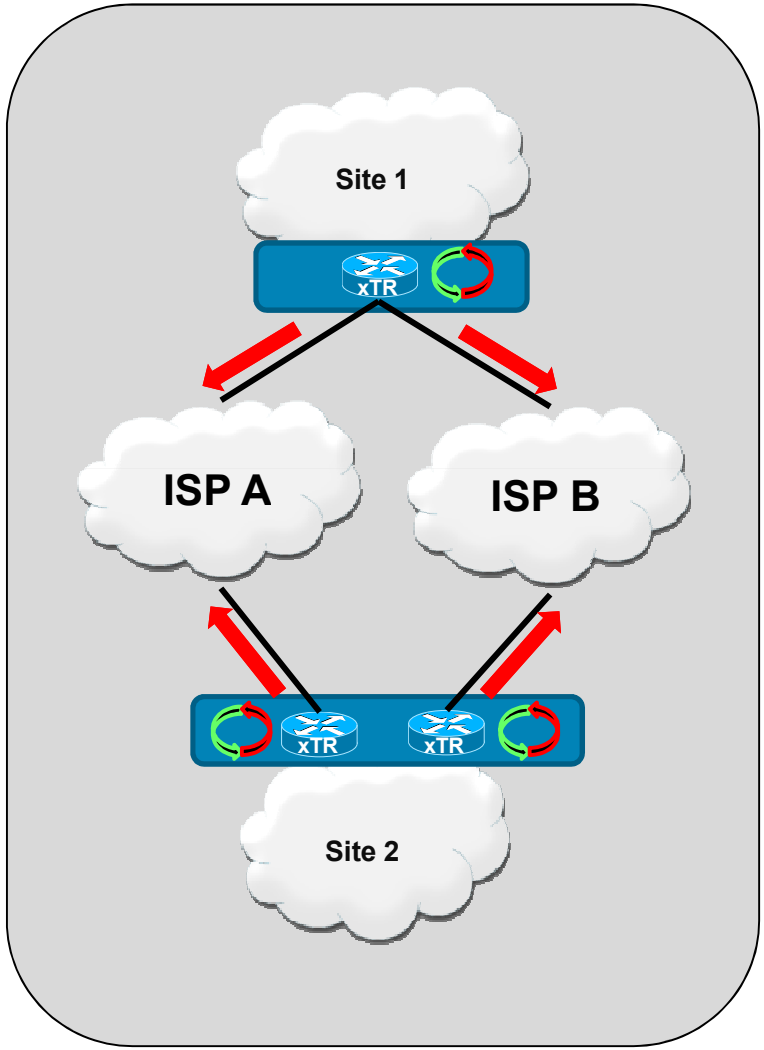
LISP Use-Cases – Multi-Homing/Redundancy

Use-Case Description

- **Needs:**
  - Site connectivity to multiple providers
  - Low OpEx/CapEx
- **LISP Solution:**
  - LISP provides a streamlined solution for handling multi-provider connectivity and policy without BGP complexity

Benefits:

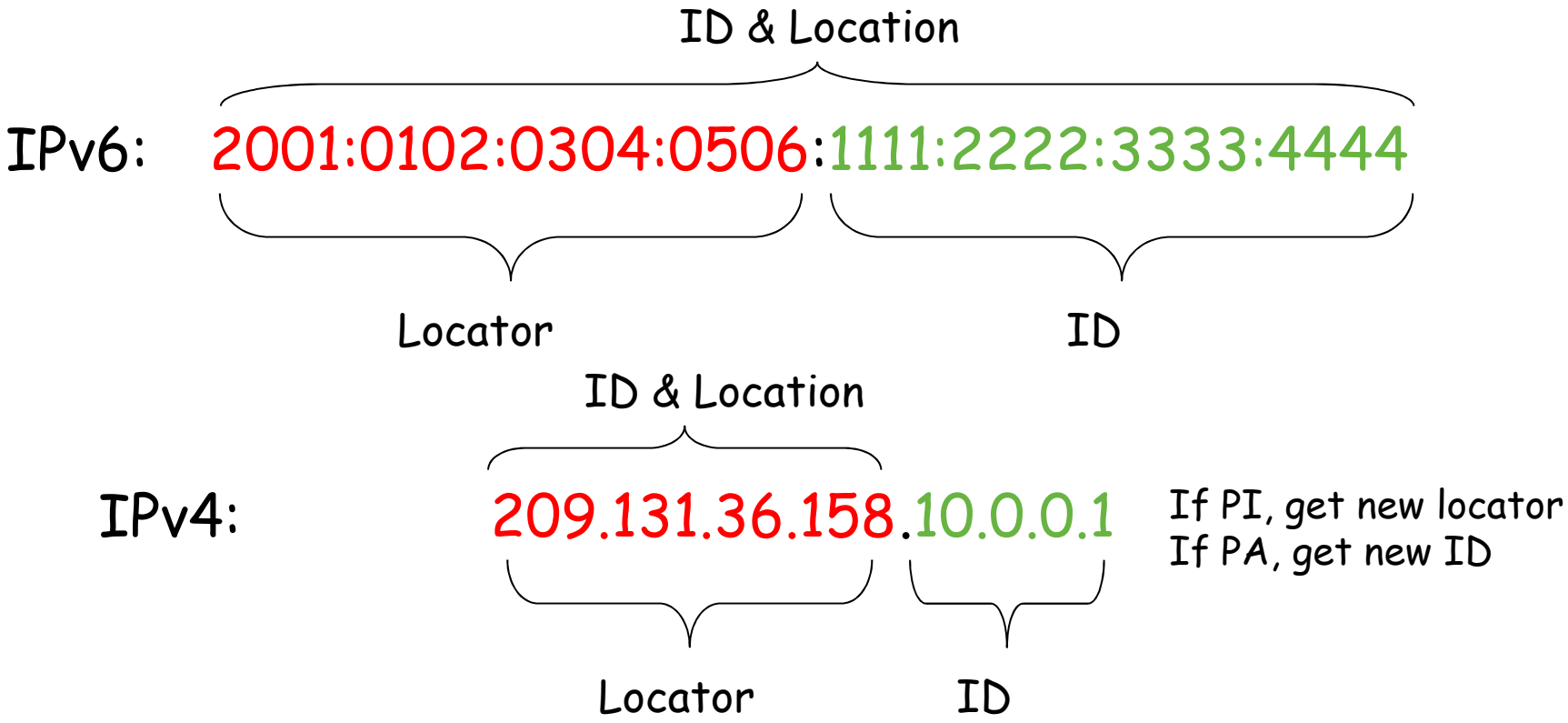
- Multi-homing across different providers
- Simple policy management
- Ingress Traffic Engineering
- Egress Traffic Engineering



LISP encap/decap

Separating (or Adding) an Address

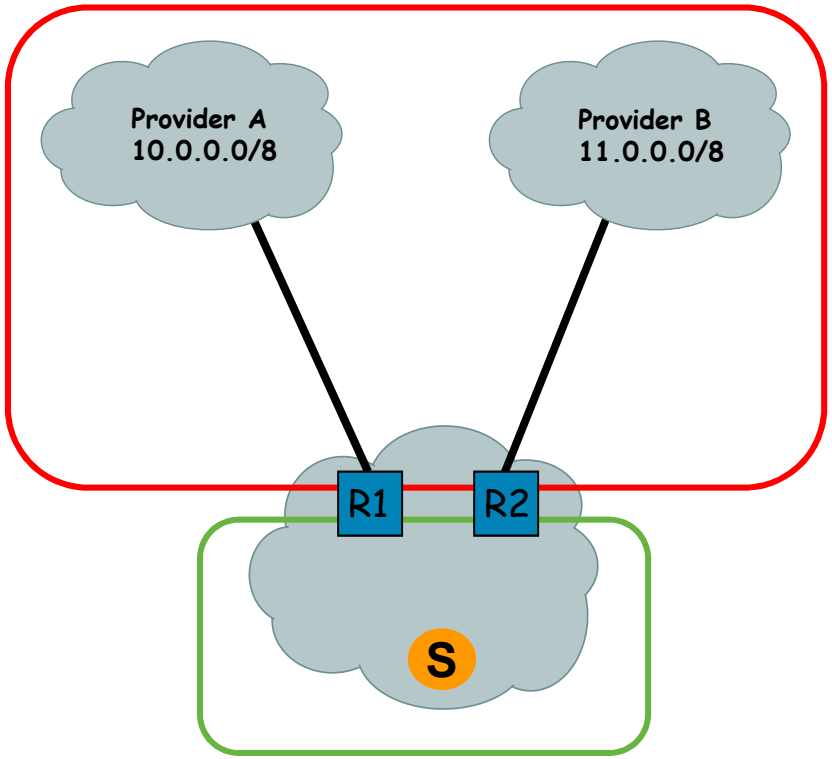
Changing the Semantics of the IP Address



Some Brief Definitions

- IDs or EIDs
  - End-site addresses for hosts and routers at the site
  - They go in DNS records
  - Generally not globally routed on underlying infrastructure
  - New namespace
- RLOCs or Locators
  - Infrastructure addresses for LISP routers and ISP routers
  - Hosts do not know about them
  - They are globally routed and aggregated along the Internet connectivity topology
  - Existing namespace

Multi-Level Addressing



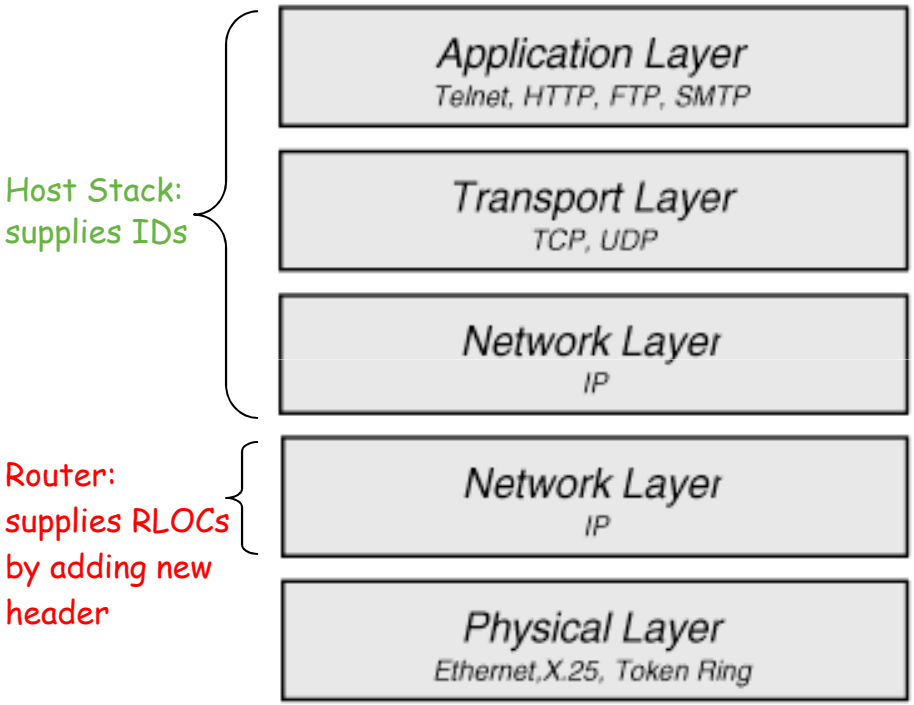
RLOCs used in the core

EIDs are inside of sites

Was ist  ?

- Locator/ID Separation Protocol
- Ground rules for LISP
  - Network-based solution
  - No changes to hosts whatsoever
  - No new addressing changes to site devices
  - Very few configuration file changes
  - Imperative to be incrementally deployable
  - Support for IPv4 and IPv6 EIDs and RLOCs

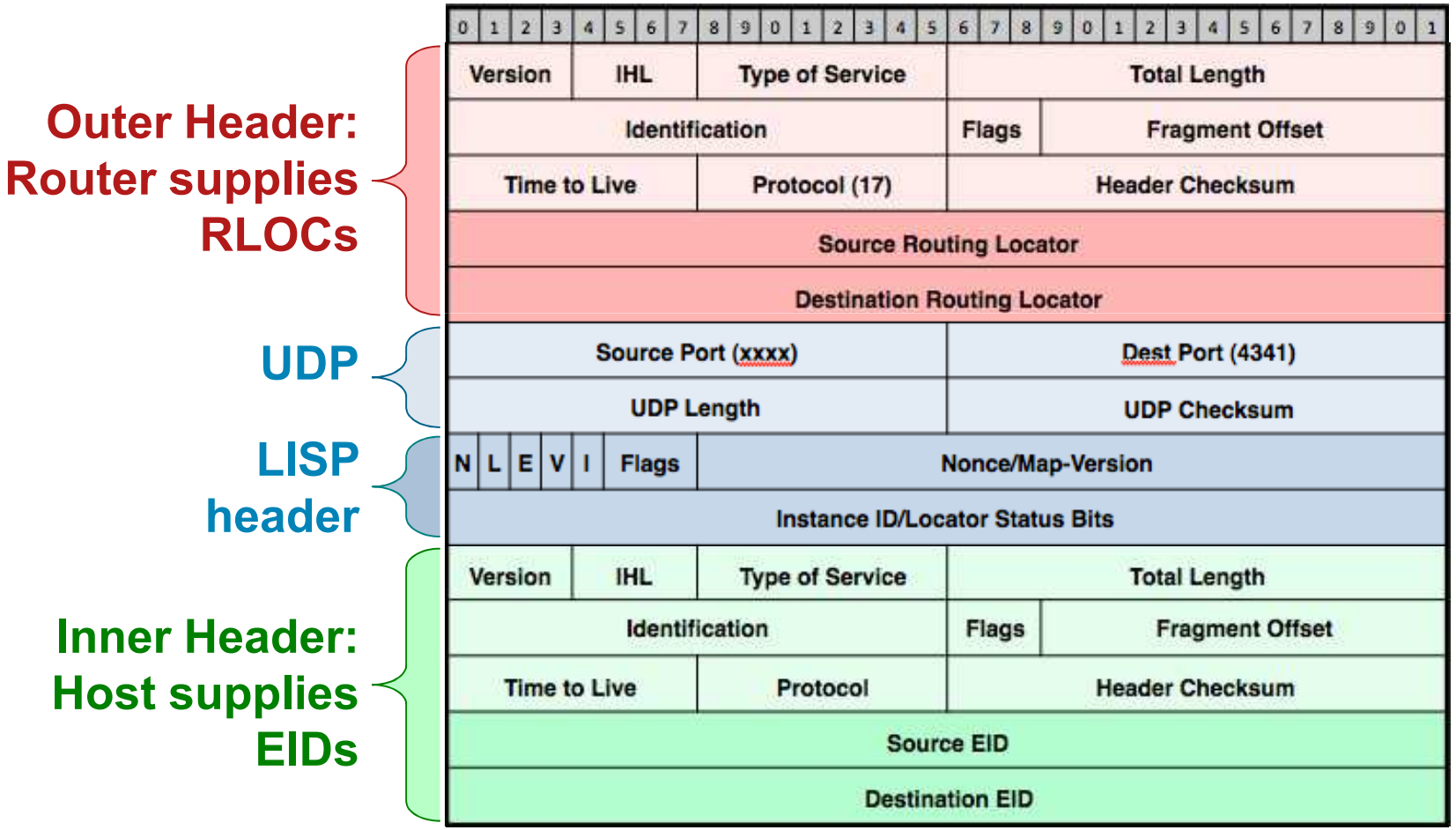
What Is LISP?



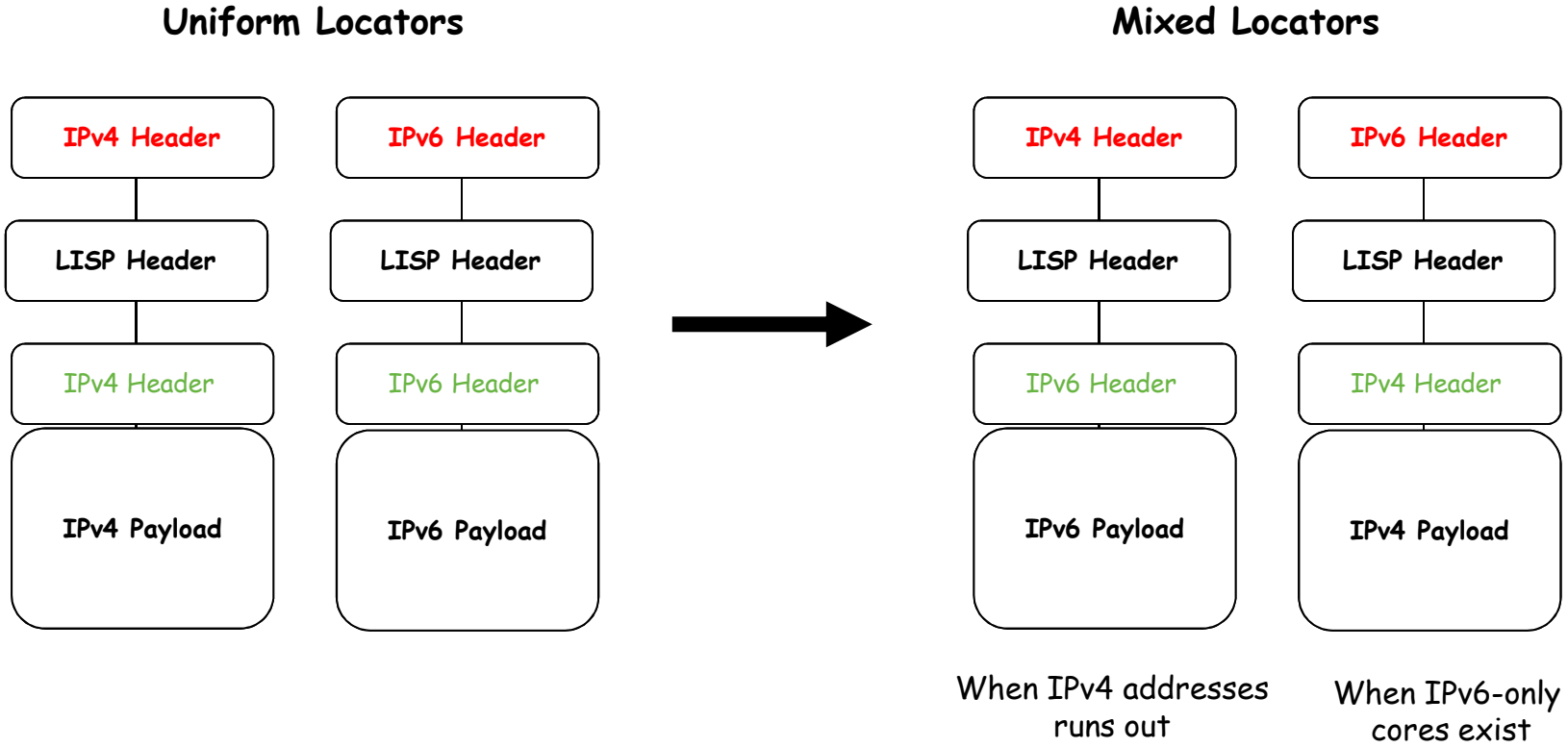
"Jack-Up" or "Map-n-Encap"

LISP – Data Header Format

draft-ietf-lisp-07



LISP for IPv6 Transition

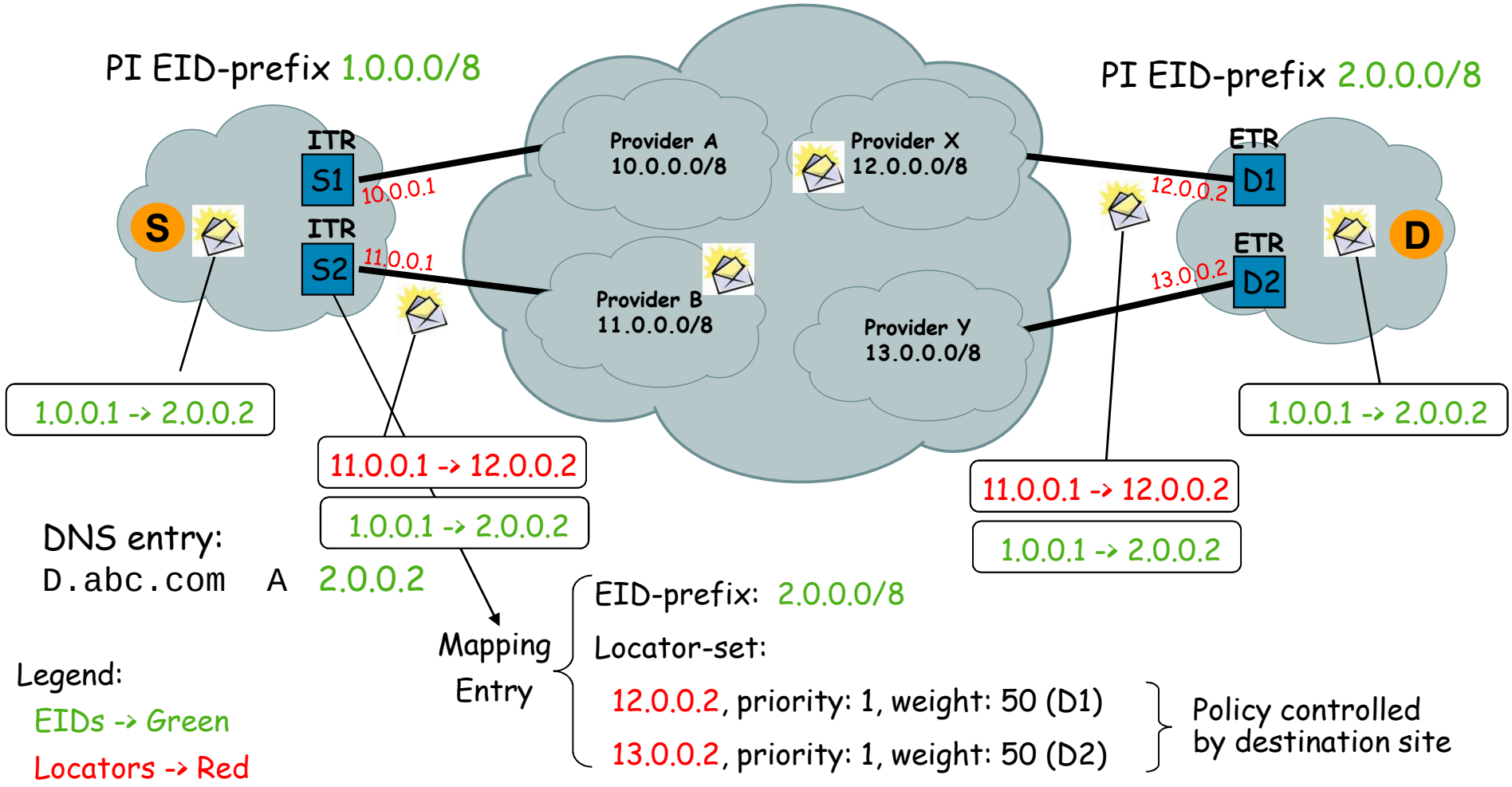


Legend: EIDs -> Green, Locators -> Red

What is the LISP Data-Plane?

- Design for encapsulation and router placement
- Design for locator reachability
- Data-triggered mapping service
  - Map-Request messages
  - Map-Reply messages
  - Map-Register messages

Unicast Packet Forwarding



What is the LISP Control-Plane?

- Definition for the “mapping cache” and “mapping database”
- Design for a modular scalable mapping service
- Examples are: ALT, CONS, EMACs, NERD
- Map-Servers and Map-Resolvers

  Interface LISP sites to mapping database service
- User tools for querying the mapping database

Mapping Database vs Mapping Cache

- LISP Mapping Database
  - Stored in all ETRs of each LISP site, not centralized
  - Authoritative Map-Replies sent from ETRs
  - Hard to DoS attack
- LISP Map Cache
  - Map-cache entries obtained and stored in ITRs for the sites they are currently sending packets to
  - ITRs must respect policy of Map-Reply mapping data
    - TTLs, RLOC up/down status, RLOC priorities/weights
    - ETRs can tailor policy based on Map-Request source

Building a Scalable Database Service

- Need a scalable EID -> RLOC mapping service
  - 10^{10} entries
- The Internet has only 2 large databases
  - BGP - pushes all information everywhere
  - DNS - pulls data on-demand from servers
- Scaling techniques
  - BGP summarizes routing information where it can
  - DNS caches information when needed
- Choose your poison
  - Trading off (`state * rate`)
  - `state` will be large, `rate` will have to be small
- We have designed several mapping database protocols
  - Tradeoff **push** versus **pull** benefit/cost
  - Did I say it needs to be scalable to 10^{10} entries :-)

Mapping Database Designs

- NERD - pure push
  - Documented but deprecated
- EMACs - pure pull
  - EID-prefixes hash to multicast groups
  - Pull mappings by sending Map-Request on multicast tree
  - Documented but deprecated
- CONS - hybrid push/pull
  - Push EID-prefixes using link-state at each hierarchical level of alternate topology
  - Pull mappings
  - Documented and deprecated
- ALT - hybrid push/pull
  - Push EID-prefixes using BGP on alternate topology of GRE tunnels
  - Pull mappings
  - ALT has the most promise
  - We are deploying ALT

Service Interface for the Mapping Database

- ETRs register site EID-prefixes with Map-Servers

  Securely with pair-wise trust model (no PKI needed)

  Policy can be applied on Map-Servers before EID-prefix accepted into mapping service

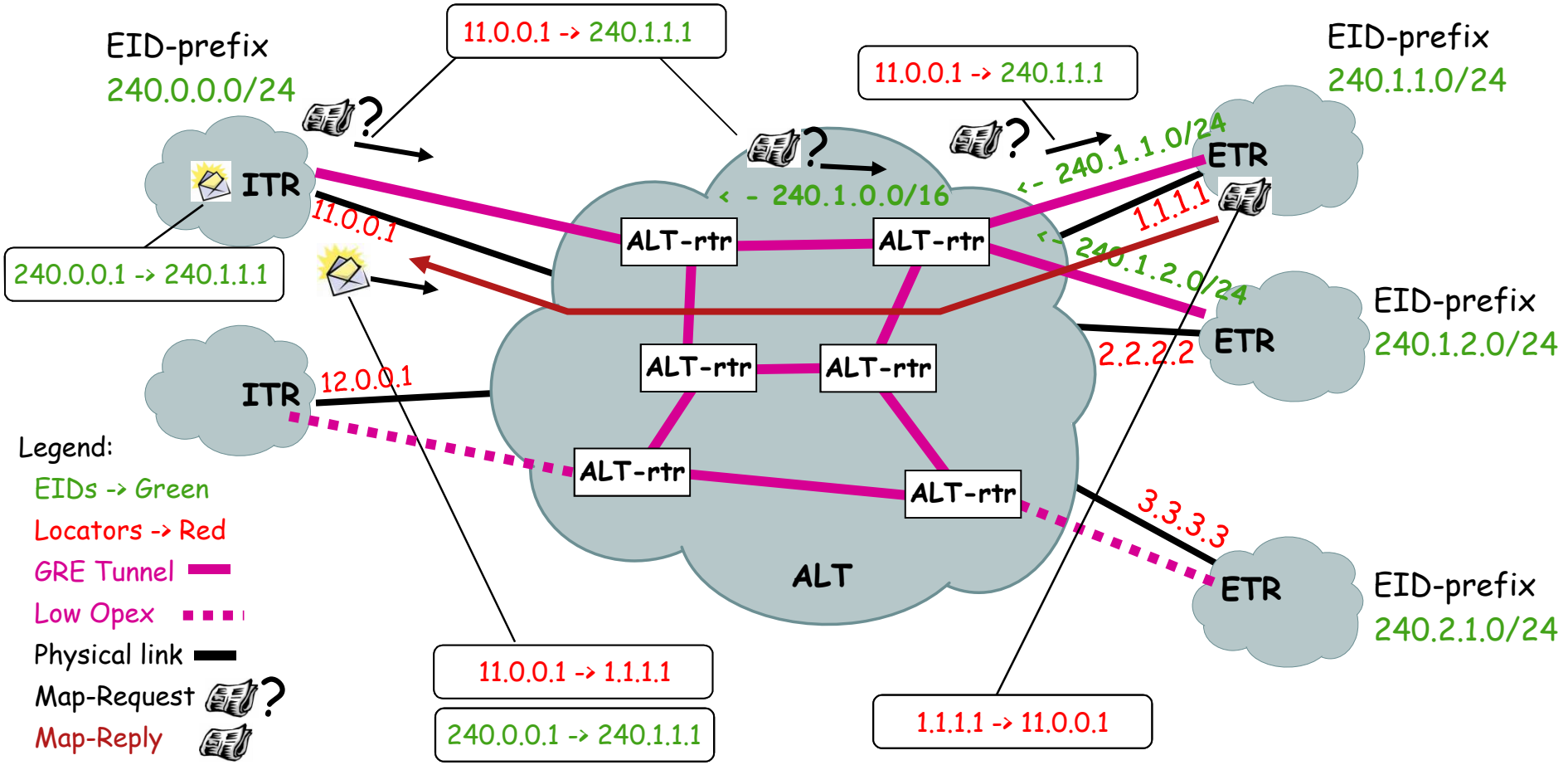
- ETRs (at the site) are authoritative for their own database mappings
- When ALT is used, Map-Servers advertise EID-prefixes

What Is LISP-ALT?

- Advertise EID-prefixes in BGP on an alternate topology of GRE tunnels
- An ALT Device is:
  - xTRs configured with GRE tunnels
  - Map-Servers
  - Map-Resolvers
  - Pure ALT-only router for aggregating other ALT peering connections
- An ALT-only device can be off-the-shelf gear:
  - Router hardware
  - Linux host
  - Just needs to run BGP and GRE

How LISP-ALT Works

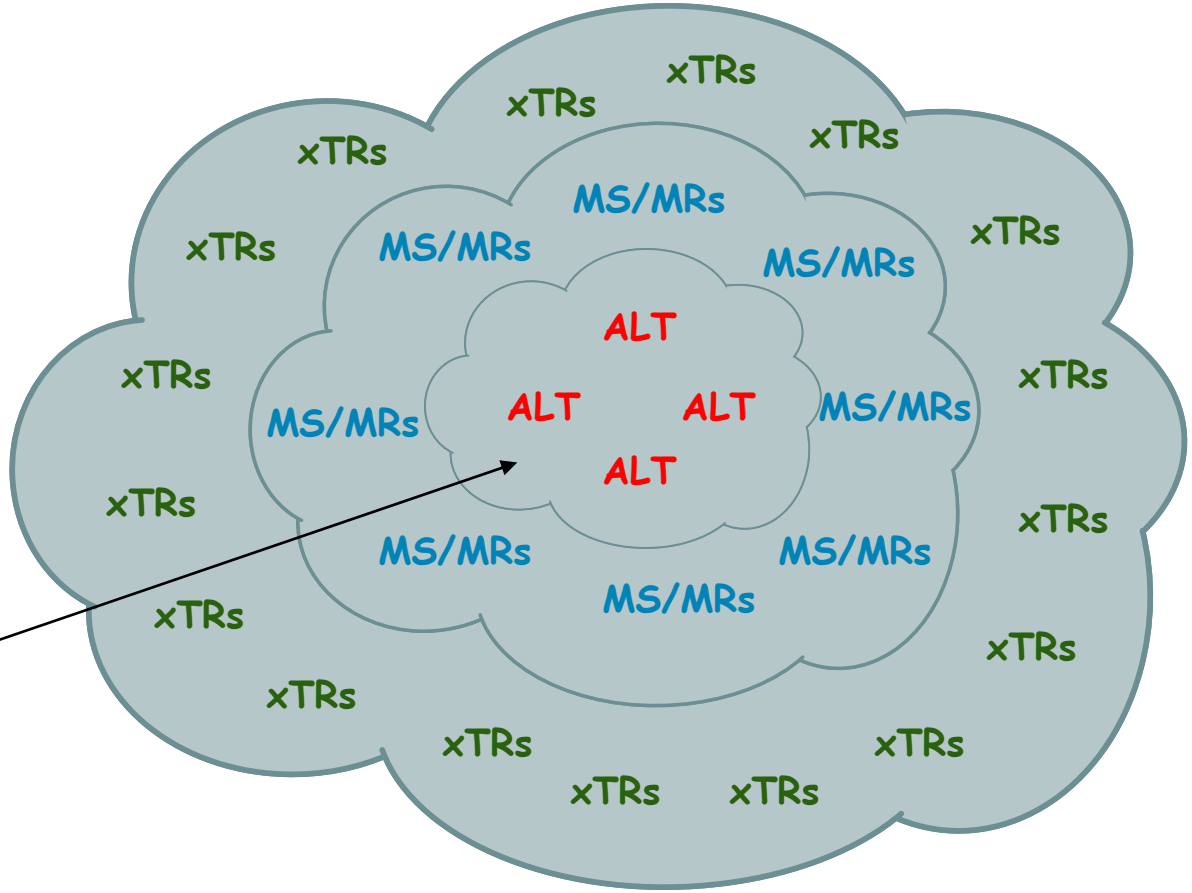
When sites are attached to the ALT with GRE tunnels



Modular Mapping Database Infrastructure

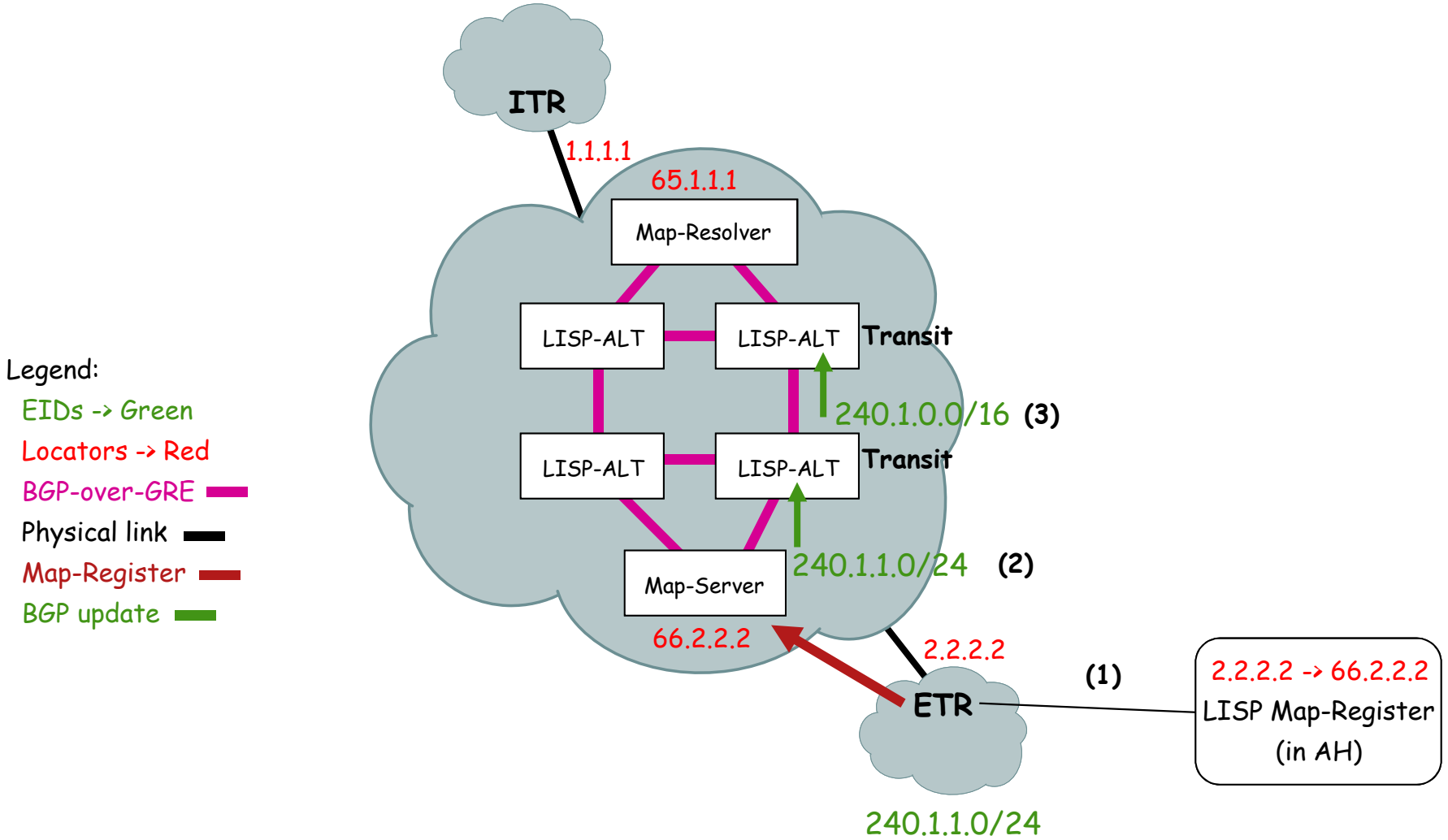
Legend:

- LISP Sites -> green
- 1st layer access infrastructure -> blue
- 2nd layer core infrastructure -> red

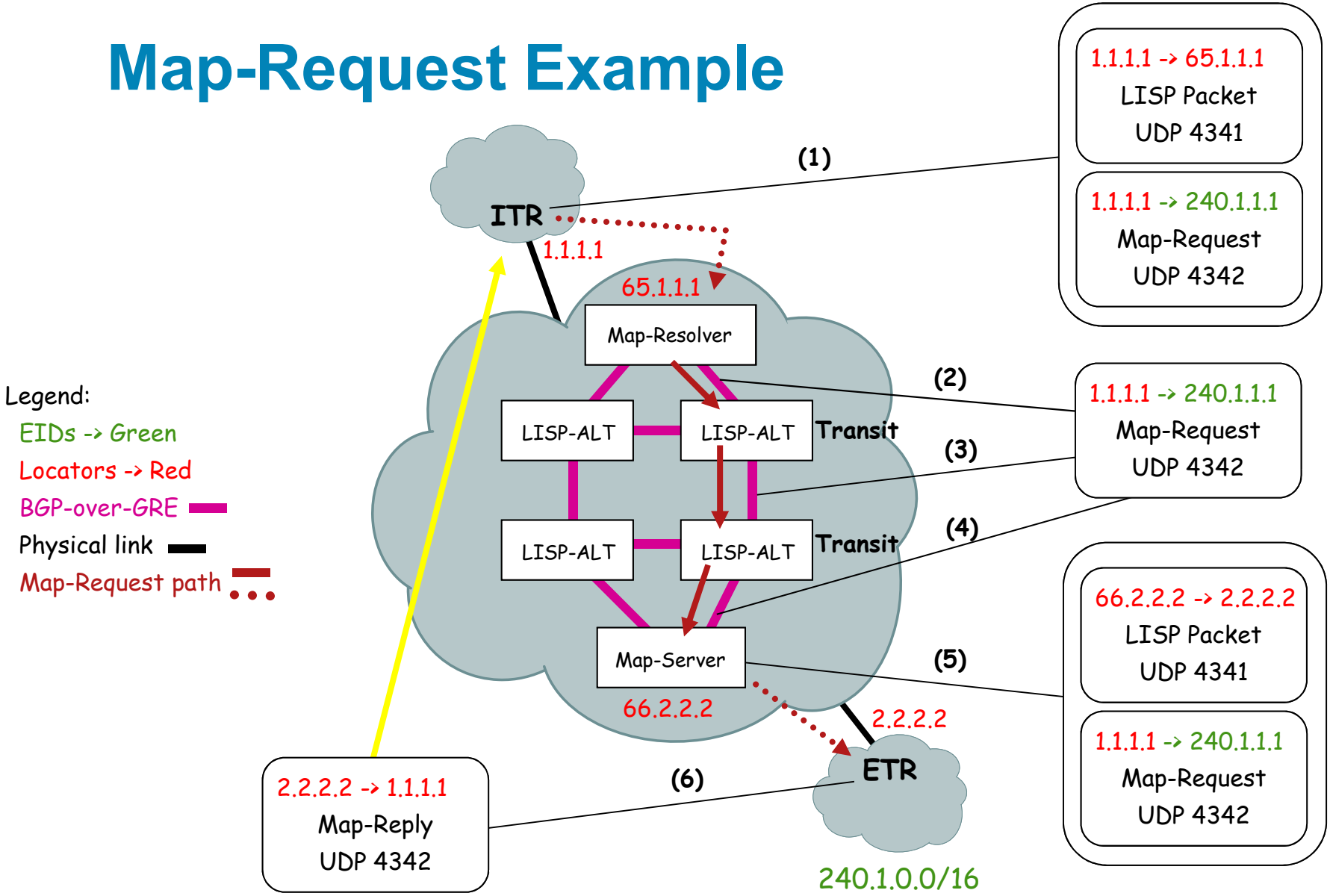


Want the ability to swap Mapping Database Infrastructure without changing sites

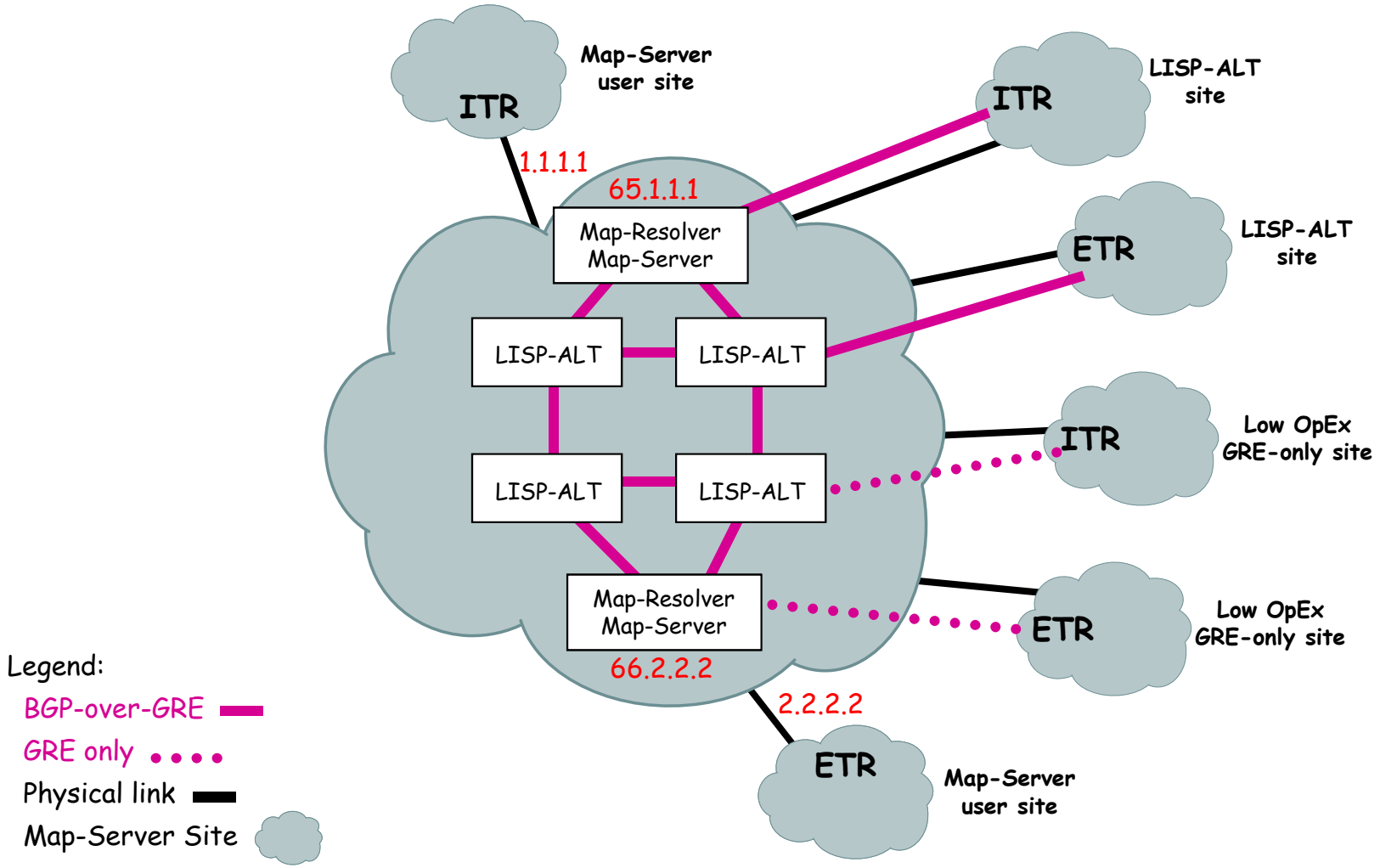
How Map-Server Registration Works



Map-Request Example



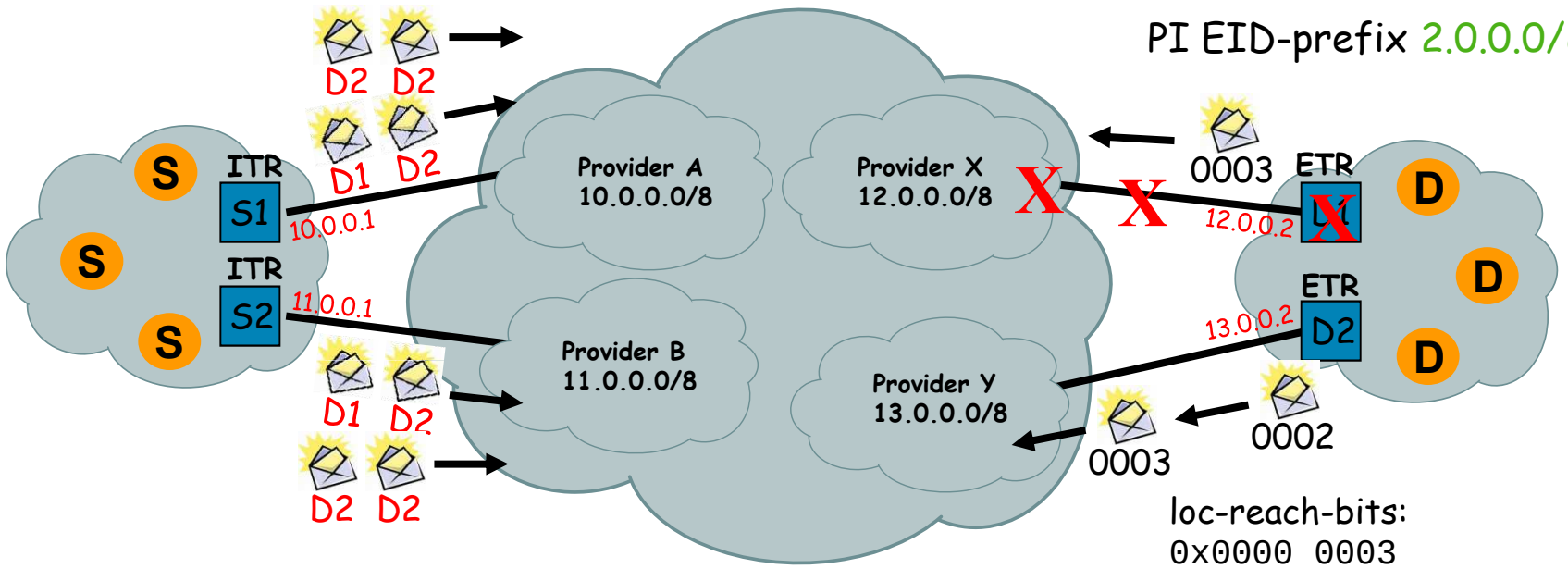
Interoperates with LISP-ALT Sites



Locator Reachability

- When RLOCs go up and down
  - Don’t want this reflected in mapping database -- keep the rate of database change very low
- Use following mechanisms:
  - Underlying BGP where available
  - ICMP Unreachables, when sent and accepted
  - Use data reception heuristics
  - Use loc-reach-bits in data packets and mapping data
- Don’t use poll probing
  - Won’t scale for the pair-wise number of sites and RLOC sets that will exist
- Use DPI heuristics?
- Use data-plane keepalives?
- Data-plane locator reachability bits for certain classes of failures

How “loc-reach-bits” Work



Mapping Entry

EID-prefix: 2.0.0.0/8

Locator-set:

12.0.0.2, priority: 1, weight: 50 (D1) -> ordinal 0

13.0.0.2, priority: 1, weight: 50 (D2) -> ordinal 1

```
  7654 3210
b'xxxx xxxx'
```

Legend:

EIDs -> Green

Locators -> Red

LISP Interworking

- LISP will not be widely deployed day-1
- Need a way for LISP-capable sites to communicate with rest of Internet
- Two basic Techniques
  - LISP Network Address Translators (LISP-NAT)
  - Proxy Tunnel Routers (PTRs)
- PTRs have the most promise
  - Creates a monetized service for infrastructure players

Security in LISP

- EID-prefixes are injected into the mapping system securely
  - Uses shared-key IPsec-AH
  - Using access control on map-server
- ITRs do not accept unsolicitd Map-Replies
- ITRs accept Map-Replies only with nonces inserted in Map-Requests
- ALT can be secured with sBGP
- Map-Replies could carry public keys
  - So ITR can encrypt encapsulated data with ESP headers

Management of LISP

- LISP Internet Groper (lig)

  Fetches a database mapping entry

  Both router and host lig available

```
titanium-dino# lig titanium-dmm.lisp4.net
Send map-request to 128.223.156.139 for 153.16.10.254 ...
Received map-reply from 128.223.156.134 with rtt 0.042518 secs

Map-cache entry for titanium-dmm.lisp4.net EID 153.16.10.254:
153.16.10.0/24, uptime: 00:00:01, expires: 23:59:58, via map-reply, auth
  Locator           Uptime    State  Priority/Weight Packets In/Out
  128.223.156.134   00:00:01  up     1/100            0/0
```

Management of LISP

- LISP Internet Groper (lig)

  Verifies you have registered your own EID-prefix to the mapping system

```
rutile# lig sel
Send loopback map-request to 128.223.156.139 for 153.16.12.0 ...
Received map-reply from 207.98.65.94 with rtt 0.002839 secs

Map-cache entry for EID 153.16.12.0:
153.16.12.0/24, uptime: 00:11:12, expires: 23:59:57, via map-reply, self
  Locator       Uptime    State  Priority/Weight  Packets In/Out
  207.98.65.94  00:11:12  up     1/100            0/0
```

Management of LISP

- LISP Internet Groper (lig)

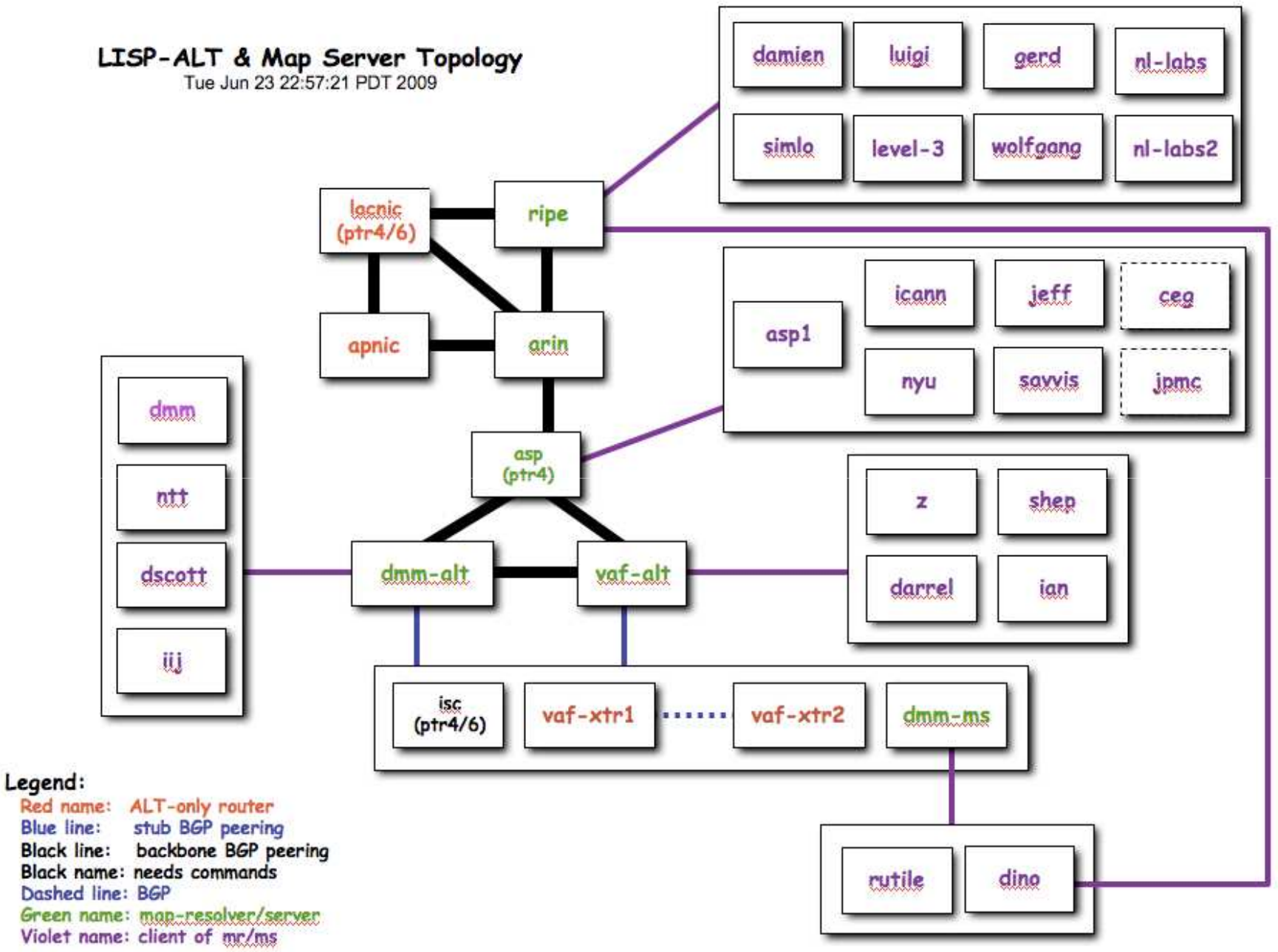
  Supports cross address-family

```
titanium-dino# lig self6
Send loopback map-request to 193.0.0.170 for 2610:d0:2105:: ...
Received map-reply from 173.8.188.25 with rtt 0.231016 secs

Map-cache entry for EID 2610:d0:2105::
2610:d0:2105::/48, uptime: 00:00:01, expires: 23:59:58, via map-reply, self
  Locator             Uptime    State  Priority/Weight  Packets In/Out
  173.8.188.25        00:00:01  up     1/33             0/0
  173.8.188.26        00:00:01  up     1/33             0/0
  173.8.188.27        00:00:01  up     1/33             0/0
  2002:ad08:bc19::1   00:00:01  up     2/0              0/0
```

Implementation Schedule

- Started implementation at Prague IETF
  - March 2007
- Implementation put on pilot network
  - July 2007
- Since then released over 130 release builds
  - Releases occur on demand with new features and bug fixes concurrently
- We have phased testing
  - Unit/System Test done in development
  - Alpha test done on pilot network by Dave/Darrel/Vince/JohnZ/Andrew
  - Beta test done on pilot network by volunteers



Naming and Addressing

- IPv4 EID Assignments from `153.16.0.0/16`

North America	`153.16.0.0/20`
/22 for regions in the US	
Europe	`153.16.32.0/20`
Asia	`153.16.64.0/20`
/21 for regions in Asia	
Africa	`153.16.96.0/20`
Latin America	`153.16.128.0/20`
Reserved	`153.16.{160,192,224}.0/20`

Naming and Addressing

- IPv6 EID Assignments from `2610:00d0::/32`

  `2610:00d0:x000::/36`

  x is continent

  `2610:00d0:xy00::/40`

  y is region in continent x

  `2610:00d0:xy00::/48`

  Sites allocate out of /48

LISP Deployment

- LISP Interworking Deployed
  - Have LISP 1-to-1 address translation working
    - `http://www.translate.lisp4.net`
  - Proxy Tunnel Router (PTR)
    - IPv4 PTRs: Andrew, ISC, and UY
    - IPv6 PTRs: Dave (UofO), ISC, and UY
    - `http://www.lisp6.net` reachable through IPv6 PTR
    - `http://www.ptr.lisp4.net` reachable through IPv4 PTR
- Go type now into your browser: **`http://www.lisp4.net`**
  - Web server in LISP site at University of Oregon
  - Demonstrates “LISP-Interworking” in action - you at non-LISP site talking to a LISP site
  - It’s in green because it’s an EID!

Open Policy for LISP

- It's been >2 years since the IAB Routing and Addressing Workshop
- This **is not** a Cisco only effort
  - We have approached and recruited others
  - There are no patents (cisco has **no IPR** on this)
  - All documents are Internet Drafts
- We need and seek designers, implementors, testers, and researchers
- 2 years in IRTF (Routing Research Group (RRG))
- 2 IETF BOFs
  - Dublin and San Francisco IETFs
- IETF LISP Working Group formed summer 2009

LISP Peer Review

- We have been encouraged by the following peer reviewers:

  Vint Cerf

  Father of the Internet and Google Chief Scientist

  Dave Clark

  Luminary Internet Researcher from MIT

  Noel Chiappa

  Locator/ID Separation Visionary and creator of NIMROD

  Paul Mockapetris

  Inventor of DNS

  Len Bosack

  Founder of cisco

Internet Drafts

draft-ietf-lisp-01.txt

draft-ietf-lisp-multicast-01.txt

draft-ietf-lisp-alt-01.txt

draft-ietf-lisp-ms-01.txt

draft-ietf-lisp-interworking-00.txt

draft-meyer-lisp-eid-block-01.txt

draft-meyer-loc-id-implications-01.txt

draft-farinacci-lisp-lig-00.txt

draft-mathy-lisp-dht-00.txt

draft-iannone-openlisp-implementation-02.txt

draft-brim-lisp-analysis-00.txt

draft-meyer-lisp-cons-04.txt

draft-lear-lisp-nerd-04.txt

draft-curran-lisp-emacs-00.txt

Cisco Leading the LISP Standardization Effort

Draft	Current Status	Next Steps/Target
LISP base protocol (draft-ietf-lisp-08)	WG Document Submitted: 08/13/2010	Experimental RFC by 12/31/2010
LISP+ALT (draft-ietf-lisp-alt-04)	WG Document Submitted: 4/25/2010	Experimental RFC by 12/31/2010
LISP Interworking (draft-ietf-lisp-interworking-01)	WG Document Submitted: 02/12/2010	Submission of -02 draft to refresh the document
LISP Map Server (draft-ietf-lisp-ms-05)	WG Document Submitted: 4/26/2010	Experimental RFC by 12/31/2010
LISP Multicast (draft-ietf-lisp-multicast-03)	WG Document Submitted: 4/16/2010	Experimental RFC by 12/31/2010
LISP Internet Groper (draft-ietf-lisp-lig-00)	WG Document Submitted: 4/10/2010	Several implementations (incl. open source) available
LISP Mobile Node (draft-meyer-lisp-mn-03)	Not WG Document Submitted: 08/02/2010	Three prototype implementations underway
LISP Canonical Address Format (draft-farinacci-lisp-lcaf-00t)	Proposed for WG adoption Submitted: 4/13/2010	Planning -01 update
LISP MIB (draft-schudel-lisp-mib-00)	Not WG Document Submitted: 8/16/2010	Submit for Beijing IETF

References

- Public mailing list:

  lisp@ietf.org

- Core LISP team:

  lisp-dddvaz@external.cisco.com

- More info at:

  http://www.lisp4.net

  http://www.lisp6.net