

Routeviews Update + What Is LISP?

Regional Interconnection Forum/NAPLA

LACNIC XII

May 2009

Panama City, Panama

David Meyer

dmm@1-4-5.net

Agenda

- A Quick Bit About Routeviews
 - History, Current Events & Futures
- What is LISP?
 - Architecture and Implementation

Routeviews Update



Agenda

- (Ancient) History
- Current Utilization Profile
- Issues in Operating Routeviews
- What's on the Horizon

History

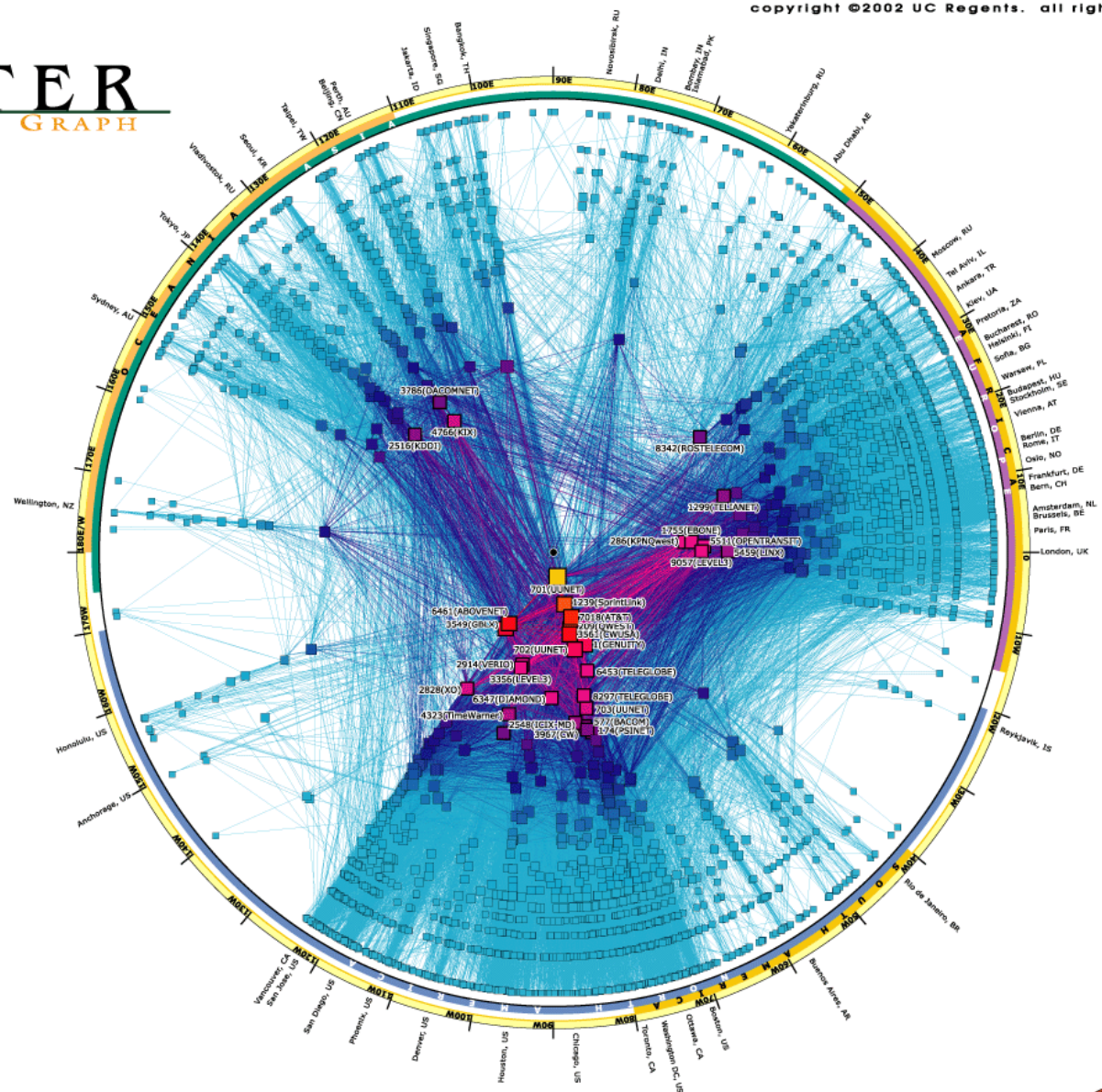
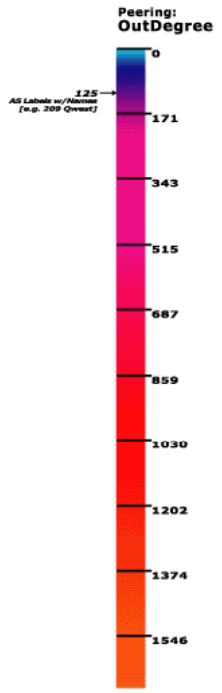
- Circa 1996: route-views started life as a purely operational tool
 - We needed a way to look at how other providers saw our prefix (this was pre-looking glass)
 - Randy Bush was kind enough to give me a eBGP multi-hop connection from MAE-WEST
- People started using route-views and contributing views

History

- It took a while for trust to build up
 - Would I leak one provider's routes to another?
- But this is what made route-views really take off (and thank you to the CAIDA folks...)

SKITTER

AS INTERNET GRAPH



Summary -- History

- Operators need(ed) real-time access to how other networks see their customers' routes
- Route-views was the first publicly available service fulfilling this need
 - "looking glasses" had not evolved
 - And people like 'sh ip bgp regex" on IOS
- Route-views has the richest set of peers
 - Currently 300+ peers

Today: Operational Use Profile

- route-views.routeviews.org receives 1000's of CLI connections daily from almost every service provider and from every part of the planet
 - And its holding millions of paths
- So its under some stress...
- Recent upgrade to NPE-G2 seems to have helped

And lots of research use too

- 100s of data downloads/day
 - 20+ GB/day transferred
- Both MRT and 'sh ip bgp' format data
 - Now bgpmon XML
- See google scholar with keywords "route views"
- And check out
 - <http://bgplay.routeviews.org/bgplay>

Operational Issues

- route-views is like a control-plane only ISP
 - All the same problems, only worse...
 - e.g., we're not actually anyone's customer, even though we take routes from you...

- As always, sys admin resources
 - New peer setup, dead peer detection and resolution
 - Contact maintenance
 - ...

Operational Issues

- Physical resources
 - Processor/memory/network scaling
 - route-views.routeviews.org has $O(5M)$ paths
 - Memory/CPU an issue
 - 'sh ip bgp' screen scraping stopped
 - more on that later
- Aging infrastructure
 - Old collectors, RAID arrays, etc

Operational Issues, cont.

- multiple, topologically diverse collection points
 - multi-hop eBGP connections hurt
 - Any failure along the long multi-hop BGP peering path makes data unavailable
 - single data collection point cannot capture the full routing dynamics of richly connected Internet topology
- presence at international locations critical

Operational Issues, cont.

- Database front end lacking
 - Data sets are getting too big to download
- A set of toolkit for data mining
 - Data sets are getting too big to look at

New Collectors (deployment in progress)

- Kathmandu, Nepal (NPIX)
- Johannesburg, SA (JINX)
- Brisbane, AU (PIPE and others)
- Sydney, AU (EQIX, NZ Peering Exchange)
- Jakarta, Indonesia (IIX)
- Curitiba, Brazil (Federal Univ. of Parana)
- San Paulo, Brazil (PTT Metro, etc)

Horizons

- New data format/collector software
 - bgpmon → xml format UPDATES/RIBs
 - Converters for MRT and 'sh ip bgp'
 - See <http://bgpmon.netsec.colostate.edu>
- Would like to hire staff to develop a set of toolkit for data mining
 - analysis and visualization tools
- More analysis of the properties of data sets
 - For example, we'd like a better understanding of sample bias
- Adding more peers that are further down the hierarchy (e.g., small ISPs, etc)

Horizons, finally...

- We're working closely with our friends at the NSRC (www.nsrc.org) to site collectors at sites around the developing world....
- If you'd like to host a collector or contribute a view, please contact us at help@routeviews.org
- Next: LISP

LISP: An Architectural Solution to Multi-homing, Traffic Engineering, and Internet Route Scaling

*Vince Fuller, Darrel Lewis, John Zwiebel,
Andrew Partan, Noel Chiappa, Dino Farinacci & David Meyer*

Agenda

- Problem Statement
- Architectural Concepts
- Data Plane Operation
- Control Plane Operation
 - Mapping Database Mechanism
- Interworking LISP Sites and Legacy Sites
- LISP Internet Groper -- lig
- Other LISP Use Cases
- Implementation & Deployment Status
- Spec References
- Q & A

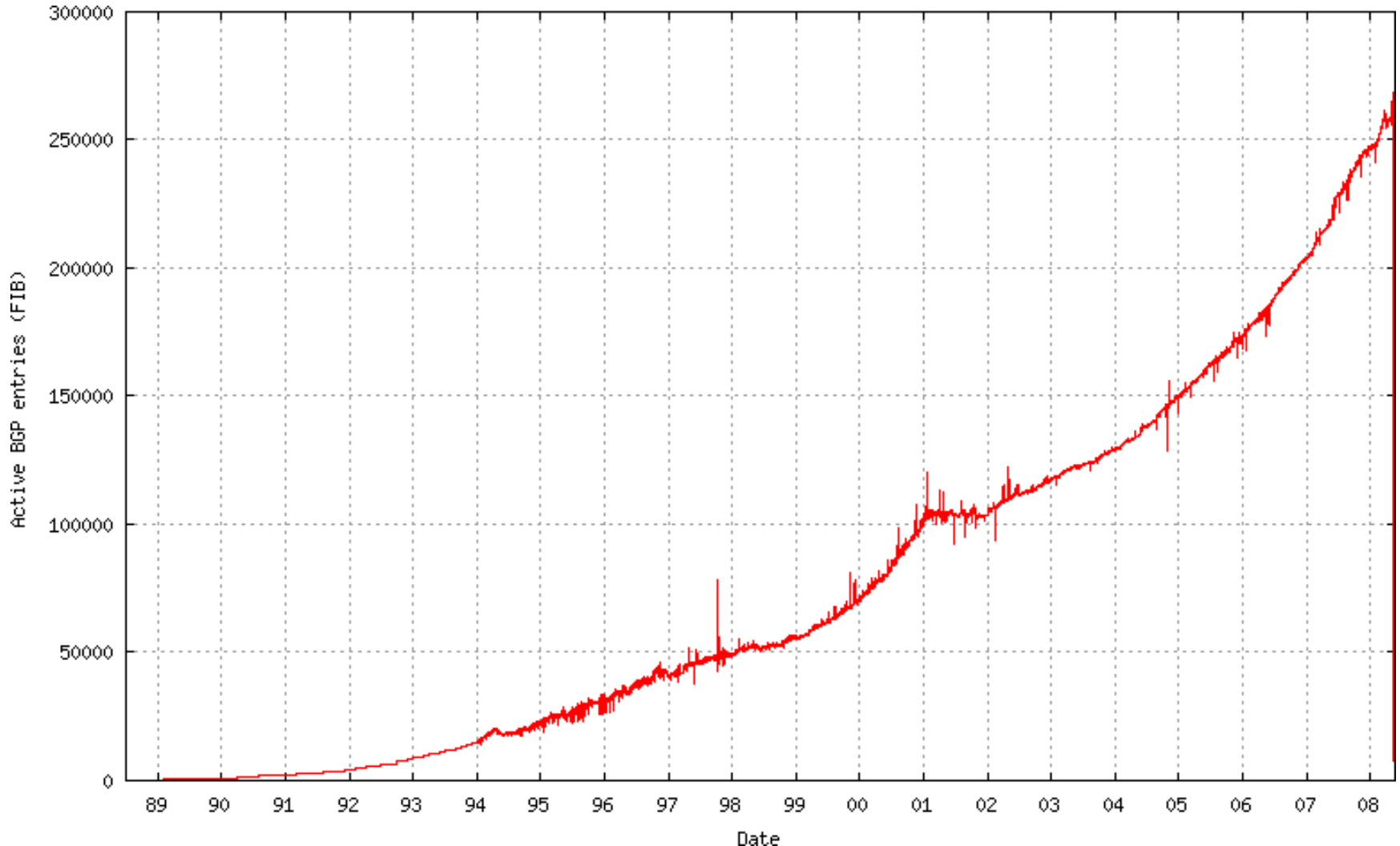
Problem Statement

- 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
- However, we've found all kinds of other applications for LISP
 - More on that later

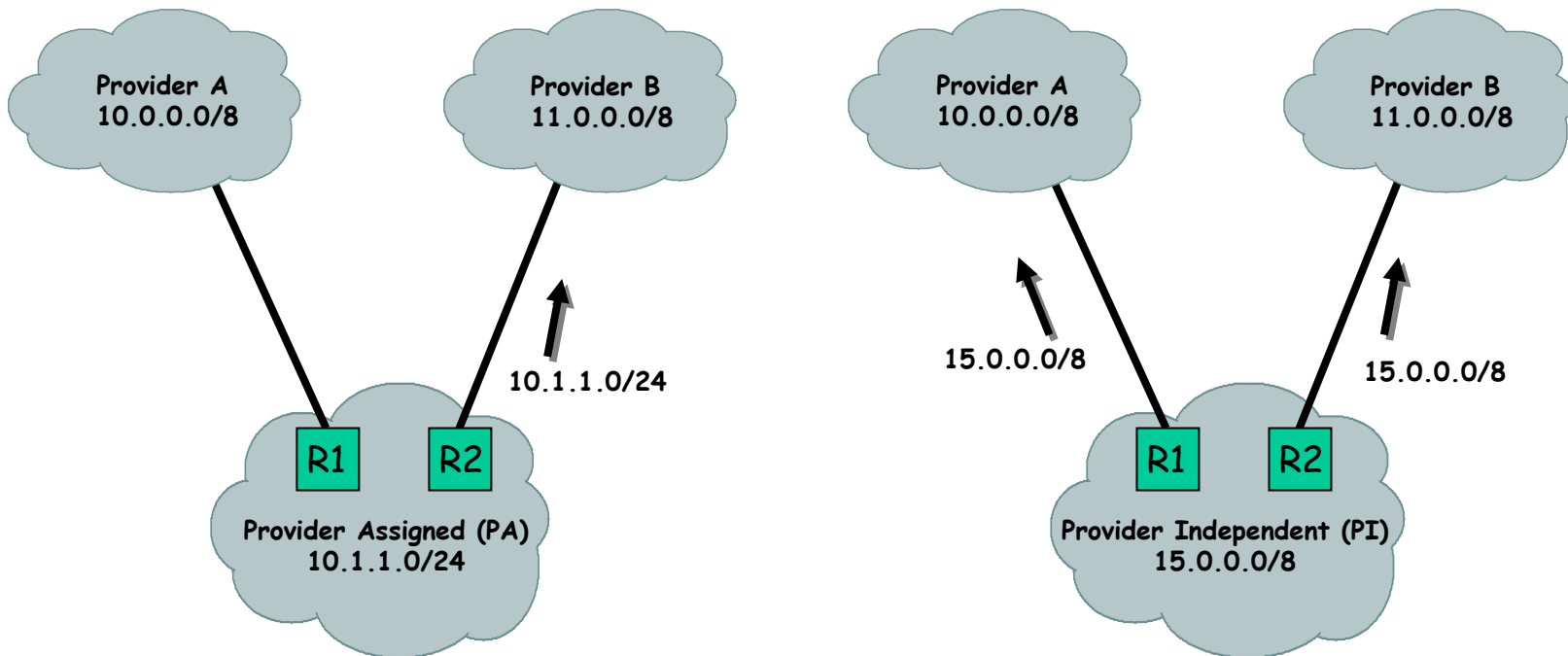
Open Policy for LISP

- It's been >2 years since the IAB Routing & Addressing Workshop
- This is not a Cisco only effort
 - 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 spring 2009

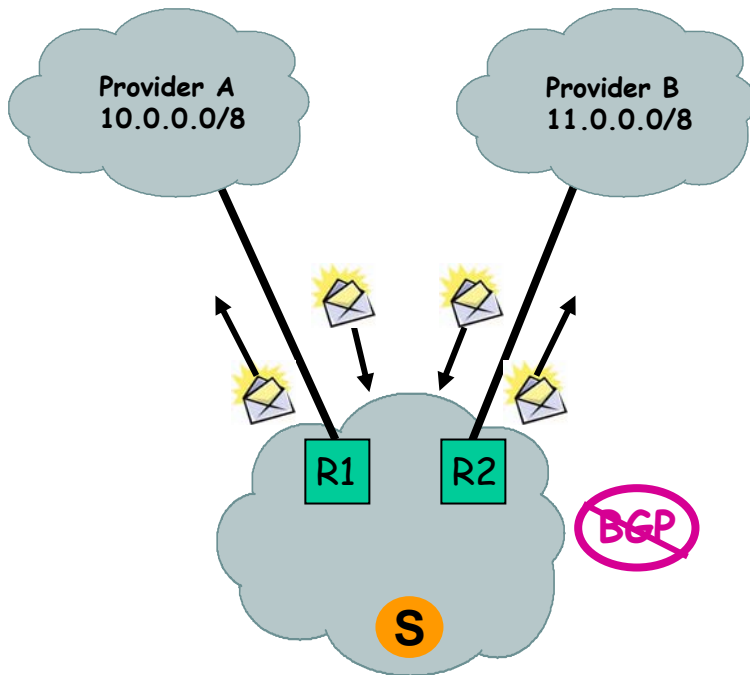
So...Scaling Internet Routing State



Why The Scaling Problem? (Hint: Multihoming/TE)



LISP Design Goals



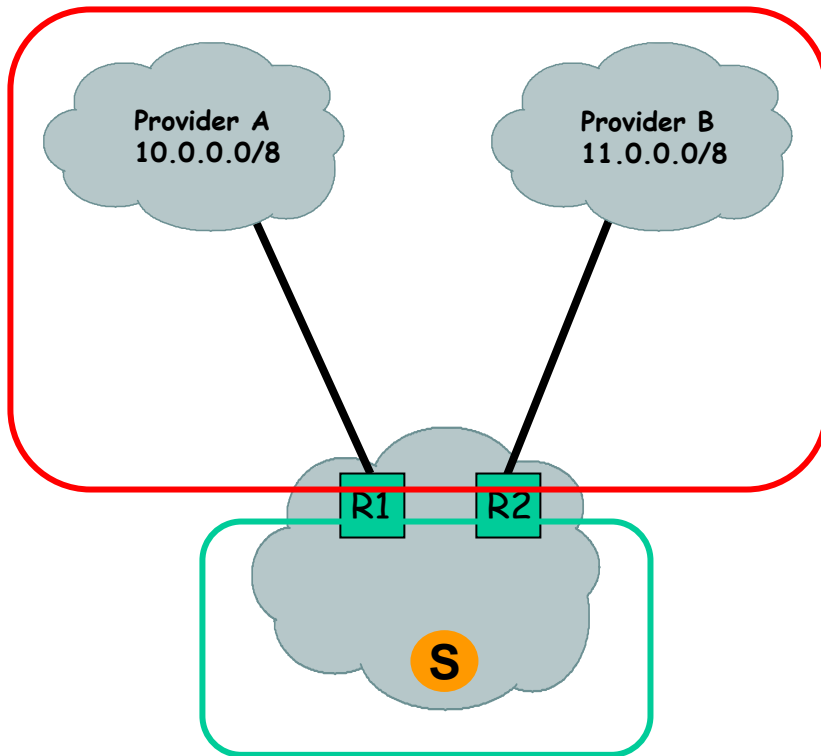
(1) Improve site multi-homing

- a) Can control egress with IGP routing
- b) Hard to control ingress without imore 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 Design Goals



(3) Decouple site addressing from provider

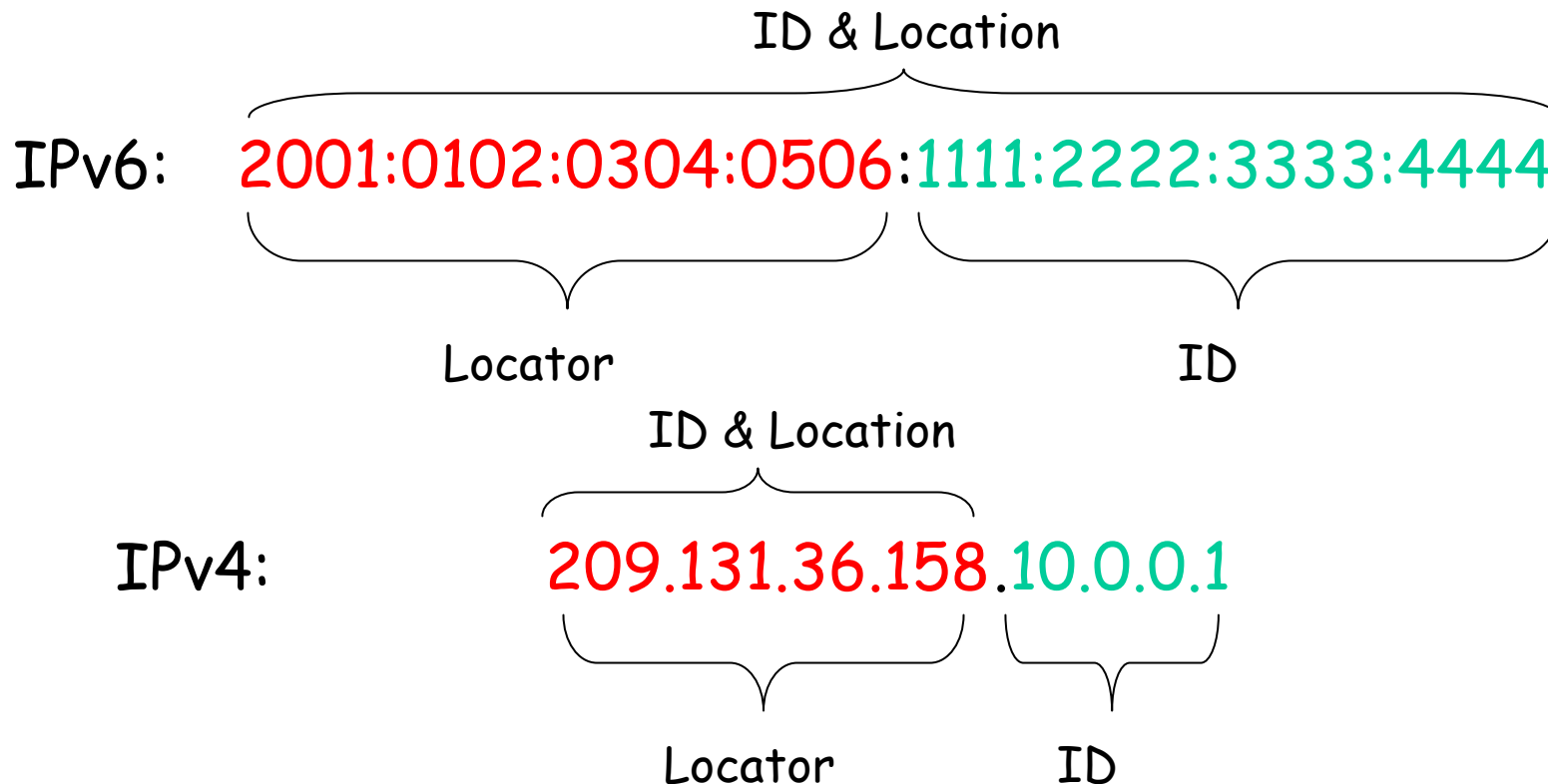
- a) Avoid renumbering when site changes providers
- b) Site host and router addressing decoupled from core topology

(4) Add new addressing domains

- a) From possibly separate allocation entities

(5) Do 1) through 4) and reduce the size of the core routing tables

What does this separation look like in practice?



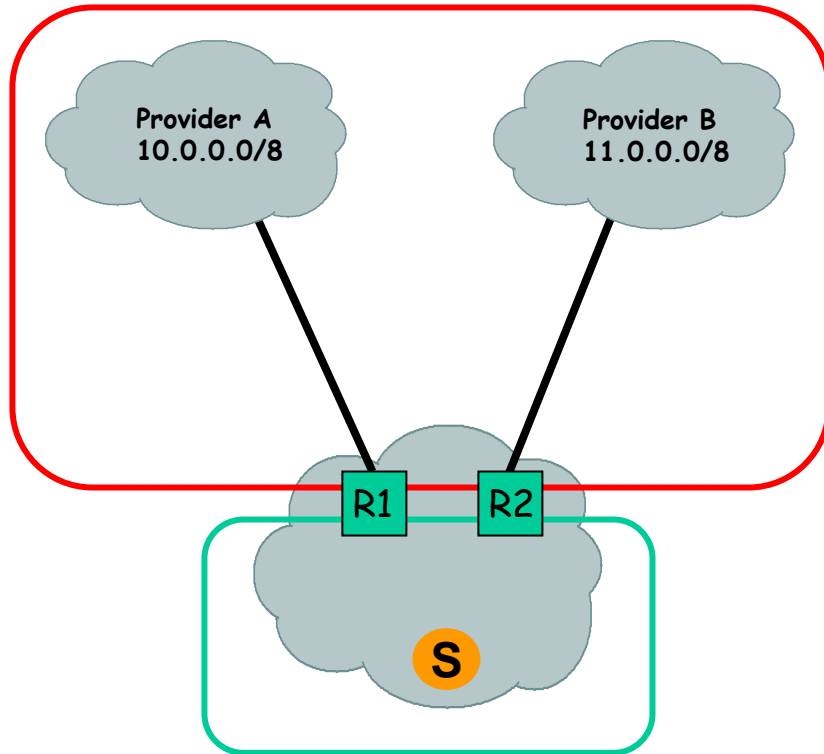
Why the Separation?

- *Level of Indirection* allows us to:
 - Keep either ID or Location fixed while changing the other
 - Create separate namespaces which can have different allocation properties
- By keeping IDs fixed
 - Assign fixed addresses that never change to hosts and routers at a site
- You can change Locators
 - Now the sites can change providers w/o renumbering
 - Now the hosts can move (limited mobility)

Some Brief Definitions

- **IDs or EIDs**
 - End-site addresses for hosts and routers at the site
 - Core doesn't generally know about them
 - Not globally routable
 - What you find in A/AAAA records
 - New namespace, aggregated along allocation hierarchy
- **RLOCs or Locators**
 - Infrastructure addresses for LISP routers and ISP routers
 - Hosts do not know about them
 - Globally routable
 - Aggregated along the Internet connectivity topology
 - Existing namespace (i.e., what is routed in the DFZ)

Multi-Level Addressing



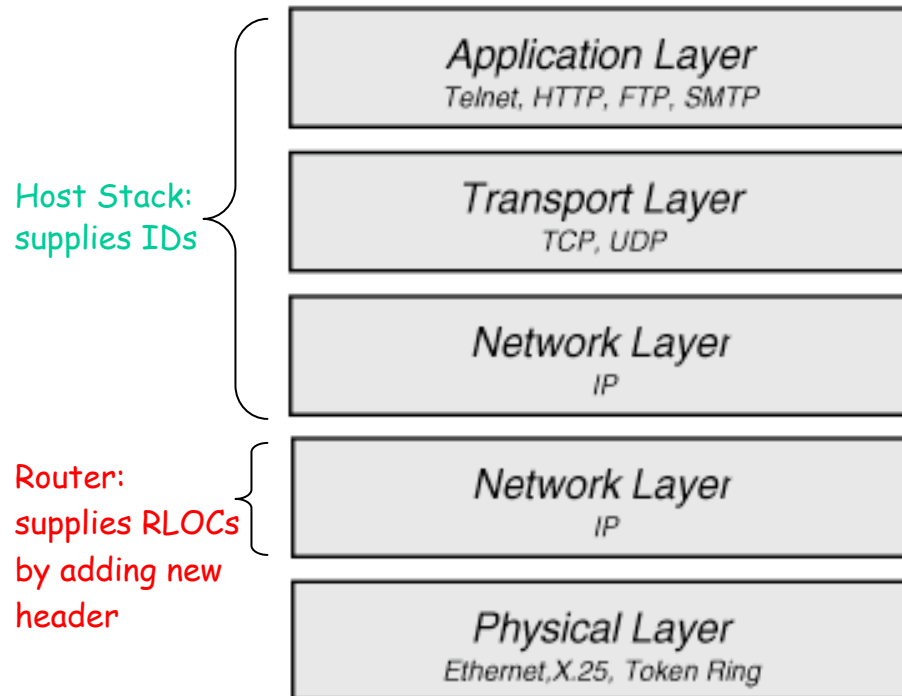
RLOCs used in the core

EIDs are inside of sites

What is LISP?

- 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
 - Address family agnostic

What is LISP?



“Jack-Up” or “Map-n-Encap”

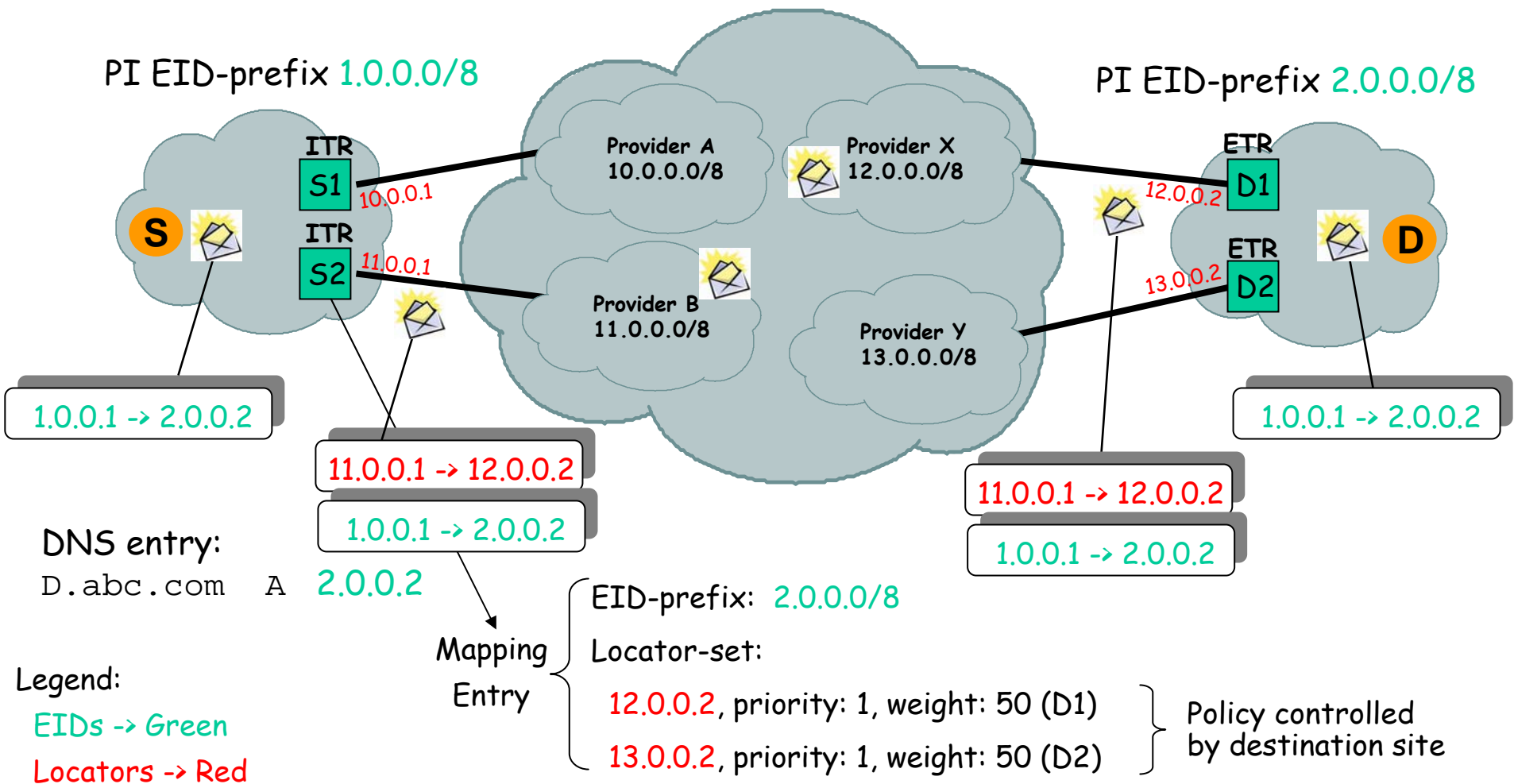
What is LISP?

- Data plane
 - Design for encapsulation and router placement
 - Design for locator reachability
 - Data-triggered mapping service
- Control plane
 - Design for a scalable mapping service (ALT)
 - Map-Servers and Map-Resolvers

LISP Network Elements

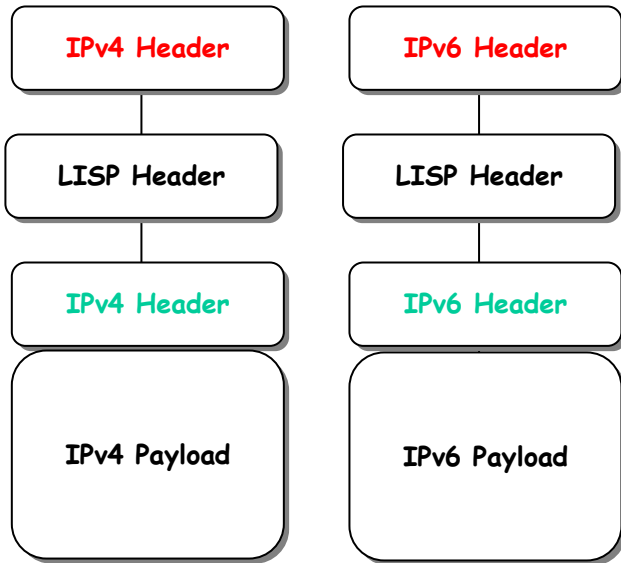
- Ingress Tunnel Router (ITR)
 - Finds EID to RLOC mapping
 - Encapsulates at source site
- Egress Tunnel Router (ETR)
 - Owns EID to RLOC mapping
 - Decapsulates at destination site
- xTR
 - Term used when not referring to directionality
 - Basically a LISP router

Data Plane: Unicast Packet Forwarding

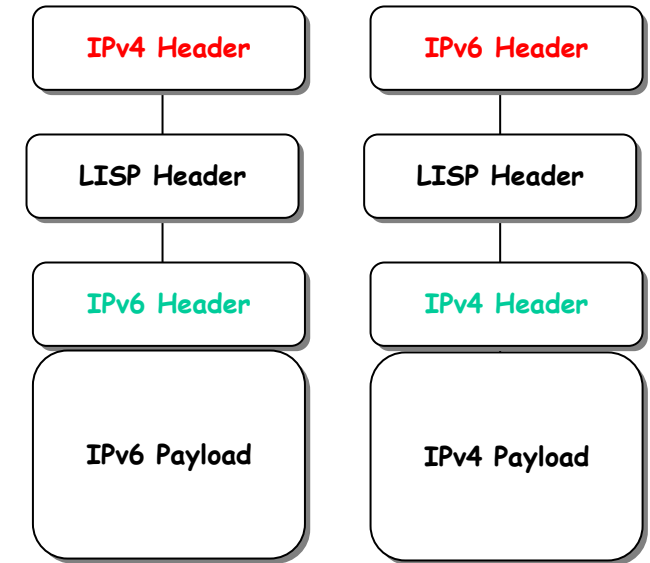


BTW, You Can Use LISP for IPv6 Transition

Uniform Locators



Mixed Locators



When IPv4 addresses runs out

When IPv6-only or dual-stack core exists

Legend: EIDs -> Green, Locators -> Red

When the xTR has no Mapping -- LISP Control Plane

- Need a scalable EID -> RLOC mapping service
- 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

Mapping Database Designs

- You need a “map” before you can “encap”
 - Design tradeoff: push versus pull benefit/cost
 - Needs to be scalable to 10^{10} entries
- We use a Map-Server/Resolver technology
 - Tied together with an Alternate Topology
 - “the ALT” (Hybrid push/pull)

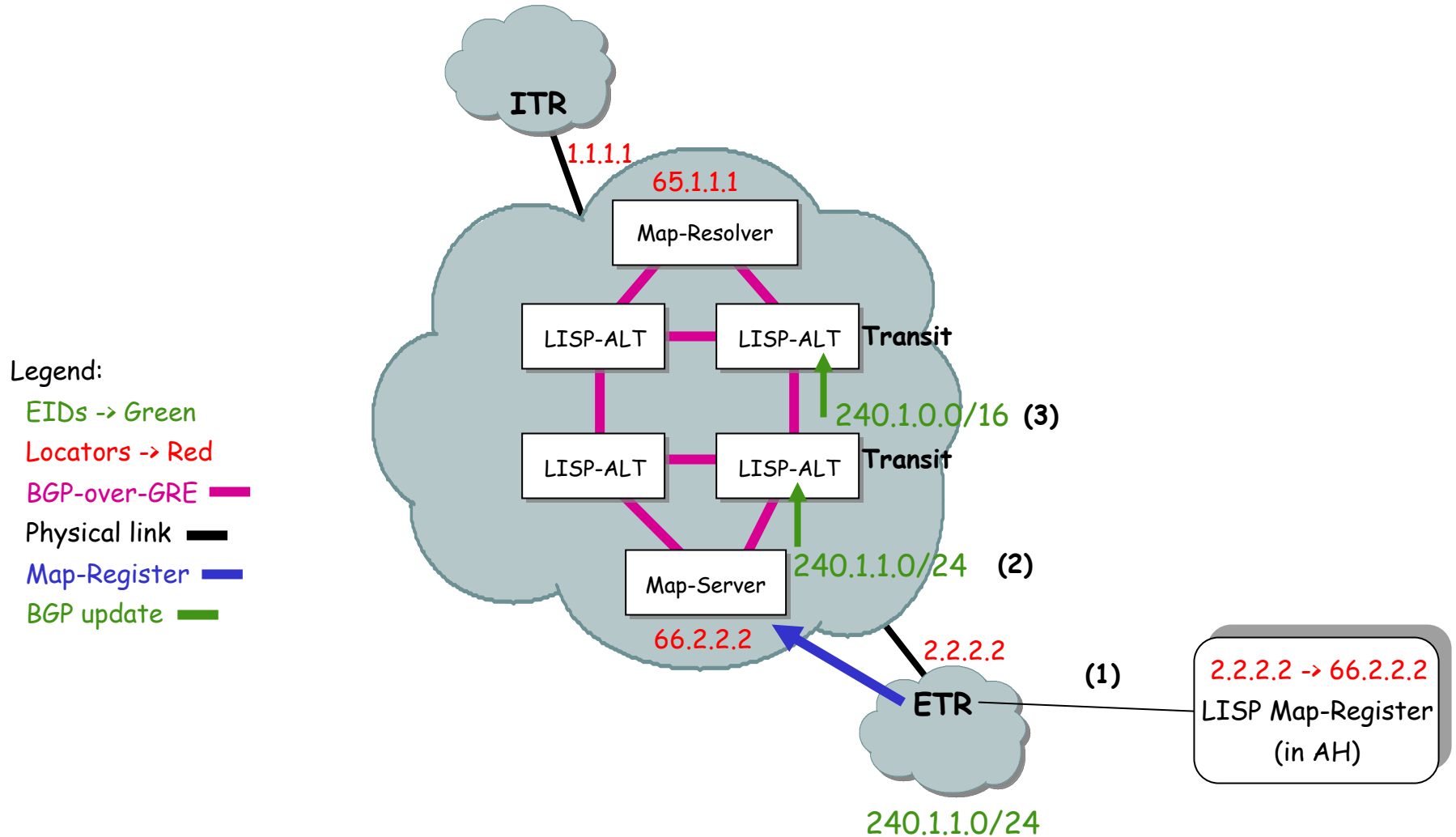
Some Brief Definitions

- LISP Mapping Database
 - Stored in all ETRs, not centralized
 - Authoritative Map-Replies sent from ETRs
- LISP Map Cache
 - Acquired and stored in ITRs for the set of sites actively sending packets to
 - ITRs must respect policy of Map-Reply data
 - TTLs, RLOC up/down status, RLOC priorities/weights

Mapping Service Interface

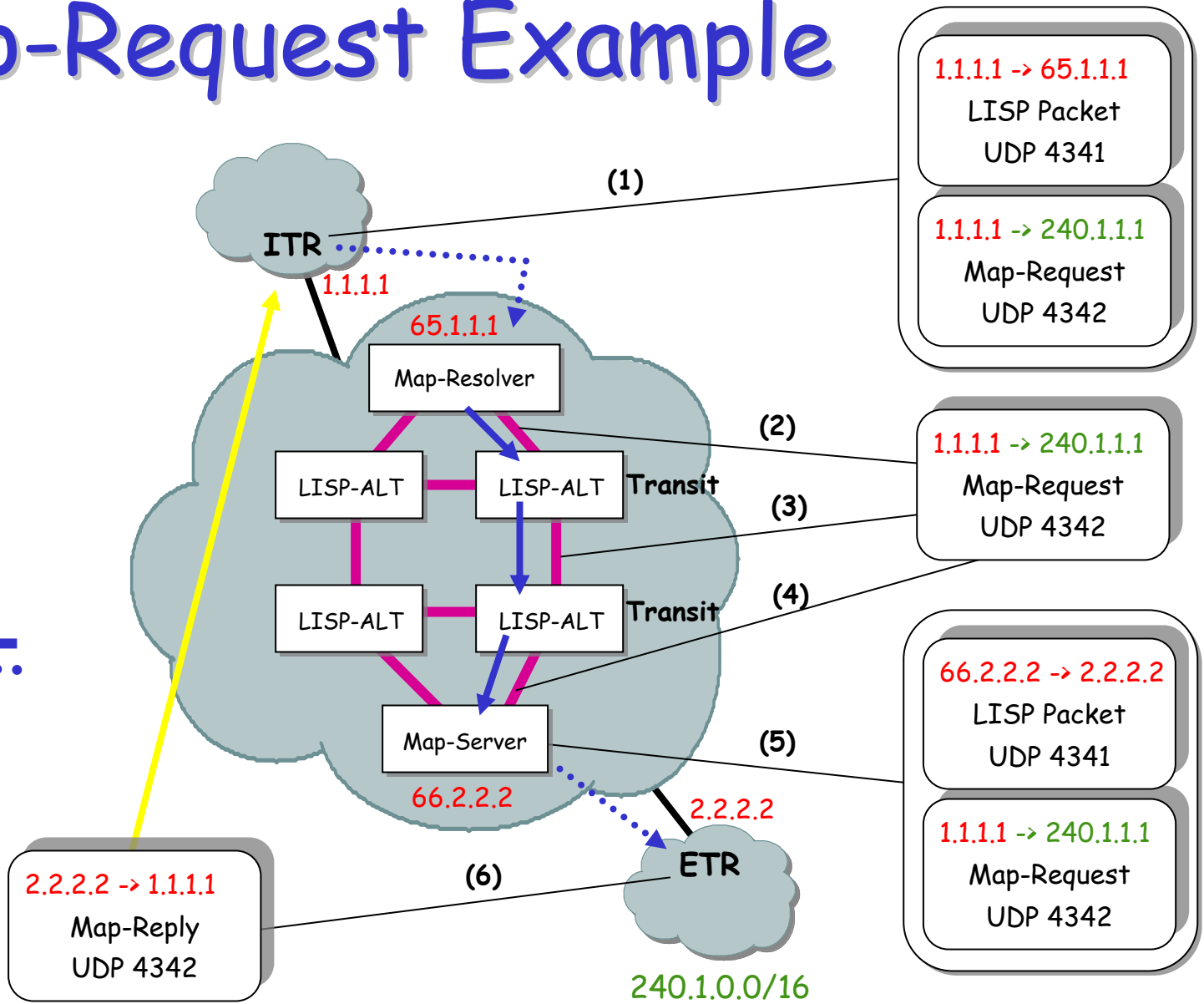
- Map-Servers and Map-Resolvers
 - DNS servers/resolvers tied together with "static routes"
 - The ALT is dynamic
 - Uses a different instance of BGP (alternate VRF)
- ETRs register site EID-prefixes with Map-Servers
 - Securely with pair-wise trust model (no PKI needed)
 - Policy can be applied on Map-Servers before EID-prefix accepted into mapping service
- ETRs (the site) are authoritative for their own database mappings
 - This provides the ability to do ingress TE

How Map-Server Registration Works



Map-Request Example

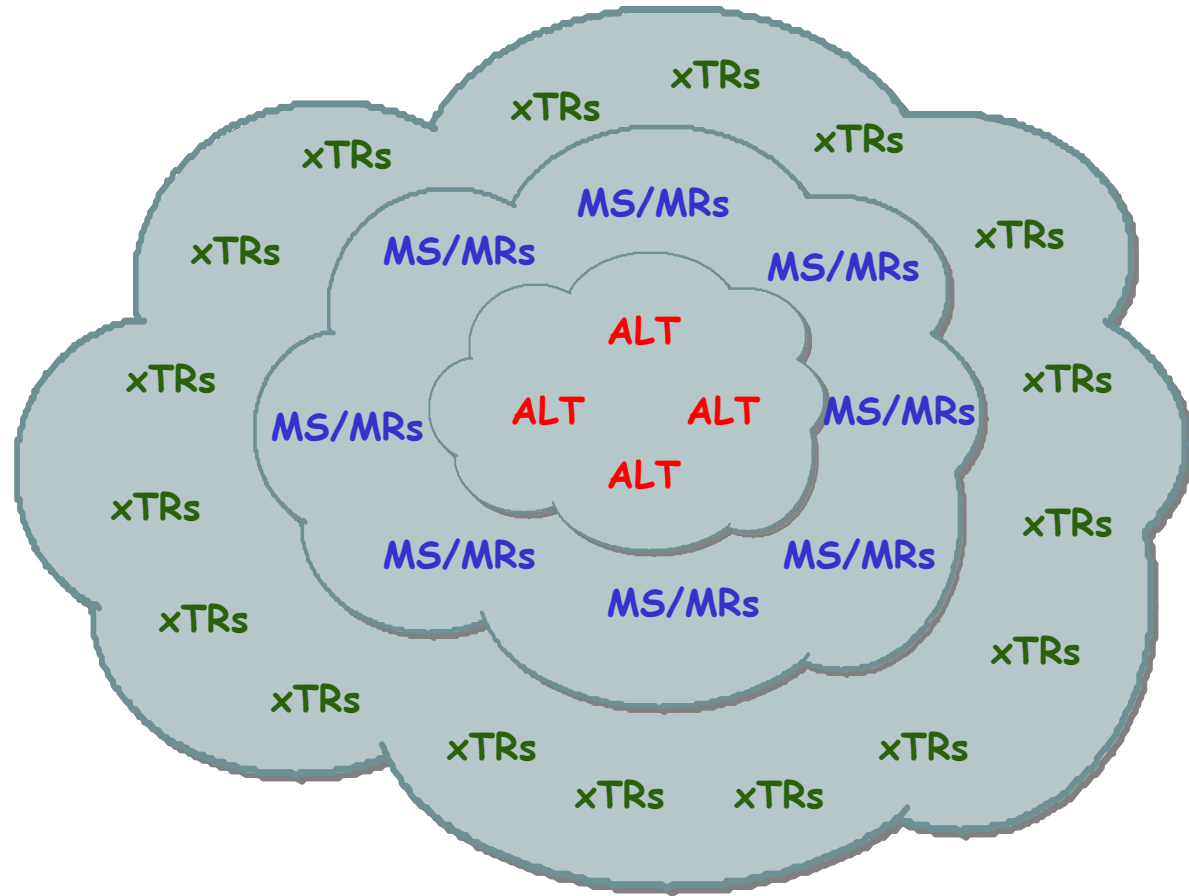
Legend:
 EIDs -> Green
 Locators -> Red
 BGP-over-GRE —
 Physical link —
 Map-Request path ...



Mapping Database Infrastructure

Legend:

- LISP Sites -> green
- 1st layer access infrastructure -> blue
- 2nd layer core infrastructure -> 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)
- <http://www.lisp4.net> uses a v4 PTR
- <http://www.lisp6.net> uses a v6 PTR
- See me later for details

LISP Internet Groper (lig)

- Fetches a database mapping entry
 - Both router and linux 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
```

LISP Internet Groper (lig)

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

```
rutile# lig self
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
```

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

A Few LISP Use Cases

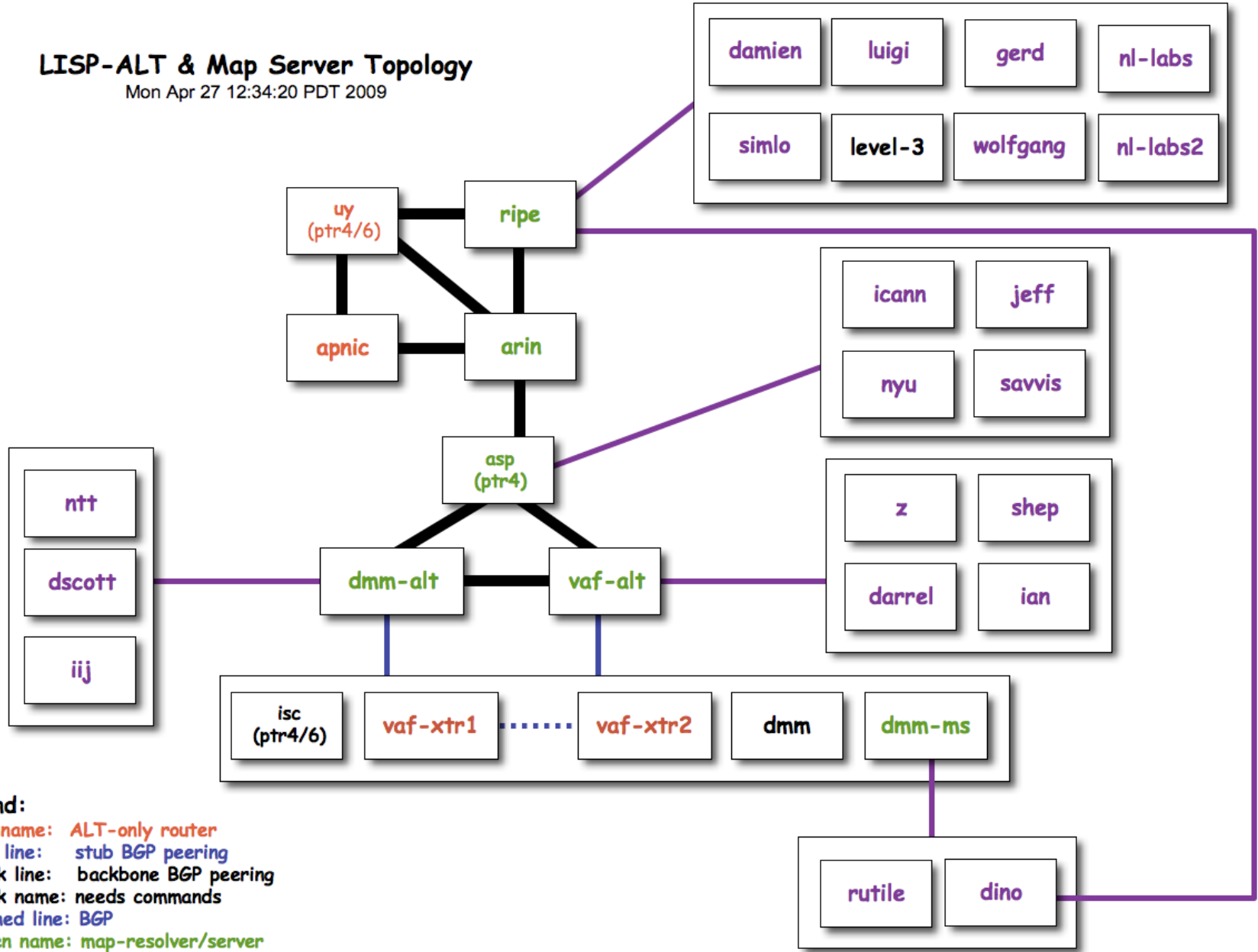
- 1) Scales routing tables in Internet core
- 2) Supports low-opex site active-active multi-homing
- 3) Supports low-opex ISP active-active multi-homing
- 4) Avoids site renumbering with provider independence
- 5) Data Center mobility of Virtual Machines (VMs)
- 6) Data Center Server Load Balancing (SLBs)
- 7) A/V Truck Roll
- 8) L2 or L3 VPNs with or without parallelism
- 9) Slow hand-set mobility in localized regions
- 10) Better residential multi-homing
- 11) IPv6-only site connectivity over existing Internet
- 12) Movement/reallocation of Cloud Computing Resources

Prototype Implementation

- Cisco has a LISP prototype implementation
- Supports:
 - `draft-farinacci-lisp-12.txt`
 - `draft-fuller-lisp-alt-05.txt`
 - `draft-lewis-lisp-interworking-02.txt`
 - `draft-fuller-lisp-ms-00.txt`
 - `draft-farinacci-lisp-lig-01.txt`
- Software switching only
- Supports LISP for both IPv4 and IPv6
 - ITR, ETR, and PTR
 - LISP-NAT for IPv4 only

LISP-ALT & Map Server Topology

Mon Apr 27 12:34:20 PDT 2009



Legend:

- Red name: ALT-only router
- Blue line: stub BGP peering
- Black line: backbone BGP peering
- Black name: needs commands
- Dashed line: BGP
- Green name: map-resolver/server
- Violet name: client of mr/ms

Internet Drafts

`draft-farinacci-lisp-12.txt`
`draft-farinacci-lisp-multicast-01.txt`
`draft-fuller-lisp-alt-05.txt`
`draft-fuller-lisp-ms-00.txt`
`draft-lewis-lisp-interworking-02.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`

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`

Q & A

Thanks!

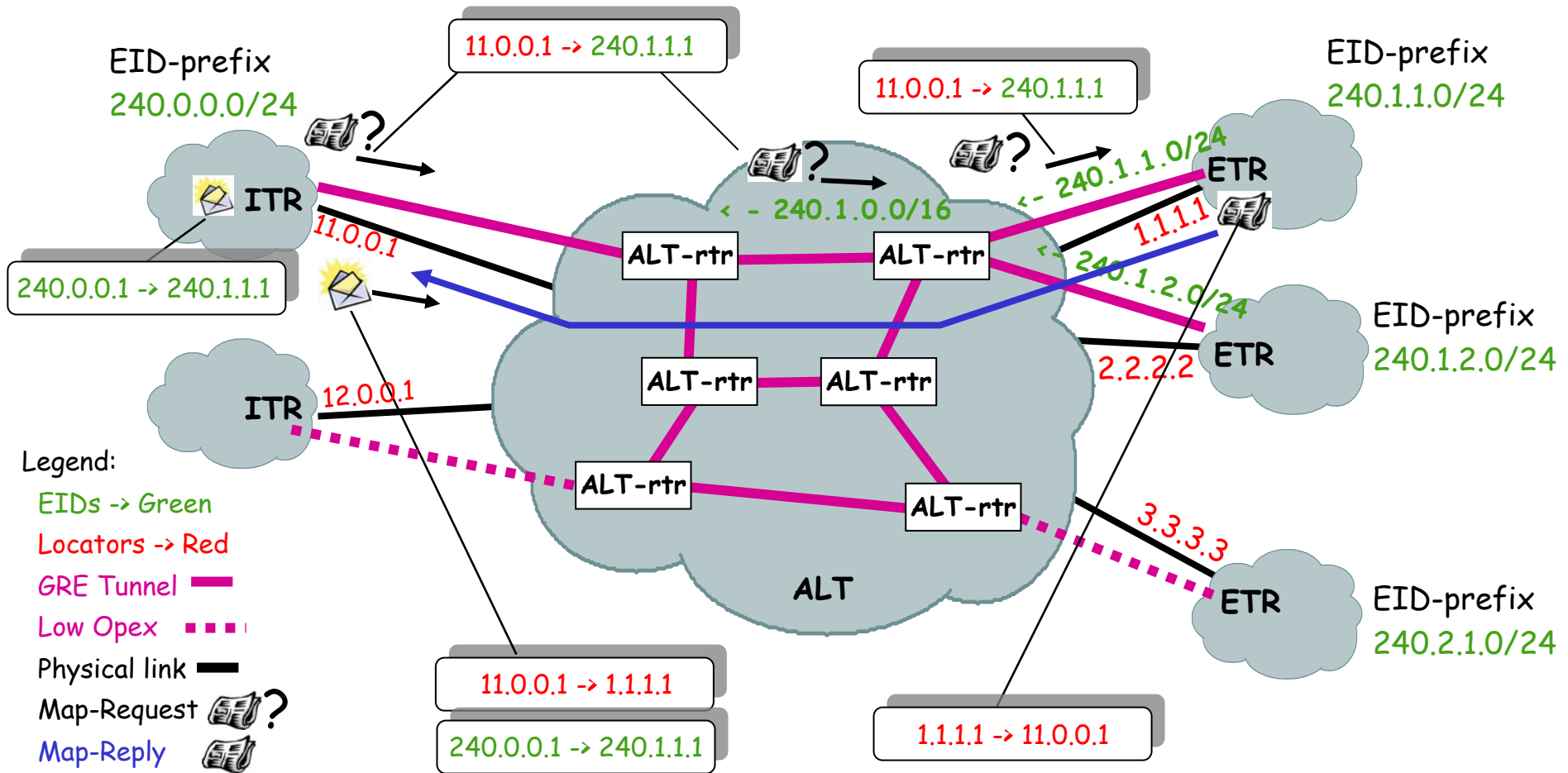
Backup Slides

What is LISP+ALT?

- Advertise EID-prefixes in BGP on an alternate topology of GRE tunnels
- Used only to forward Map-Requests to the authoritative ETR
- An ALT Device is:
 - An xTR configured with GRE tunnels
 - A pure ALT-only router for aggregating other ALT peering connections
- An ALT-only device can be off-the-shelf gear:
 - Router hardware
 - Linux host (or whatever)
 - Just needs to run BGP and GRE

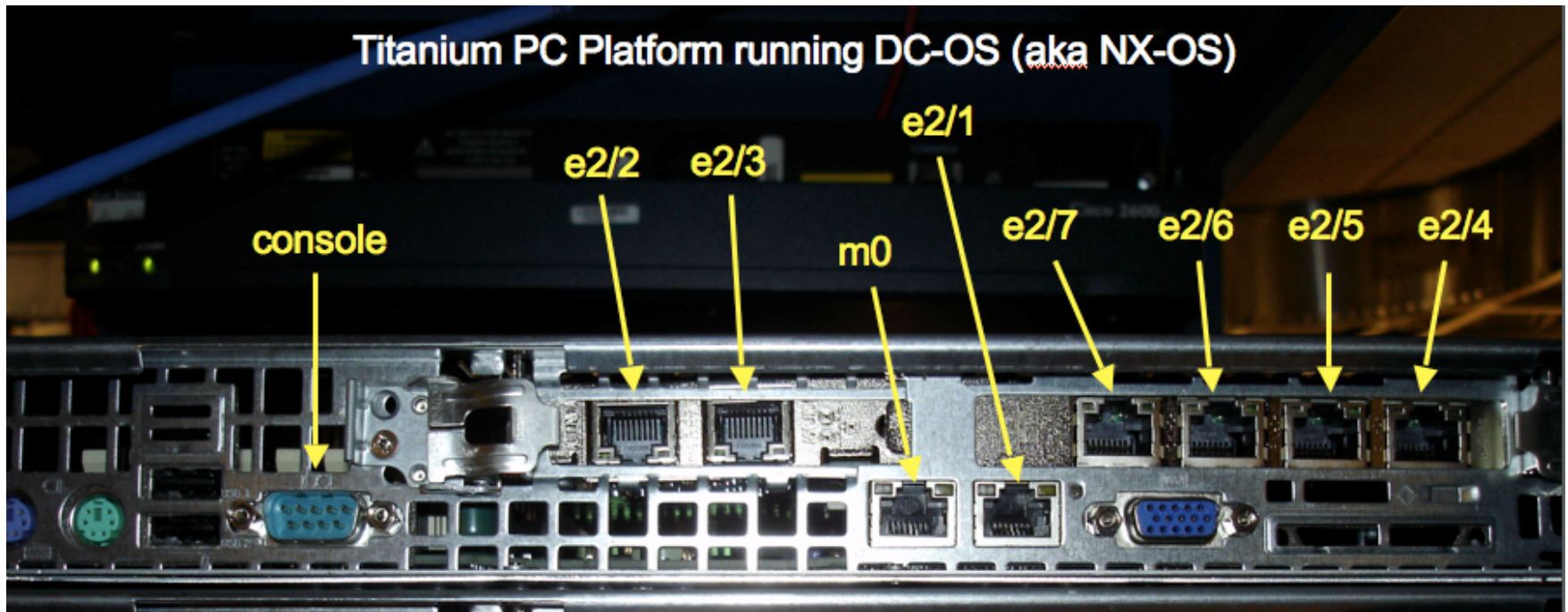
How LISP+ALT Works

When sites are attached to the ALT with GRE tunnels



Platform and OS

- NX-OS runs on Linux
- Used as a research platform
- Can run on Nexus 7K/5K & Service Blades



Implementation Schedule

- Started implementation at Prague IETF
 - March 2007
- Implementation put on pilot network
 - July 2007
- Since then released over 125 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/John/Andrew
 - Beta test done on pilot network by volunteers

Other Coding Efforts

- IOS implementation under-way
 - Loc/ID split functionality
- Considering IOS-XR implementation
 - TE-ITR/TE-ETR functionality
- OpenLISP implementation been available for FreeBSD a while and being updated
 - For testing the specs
- New native Linux implementation
 - Port of the Luigi's FreeBSD implementation (?)
- Any other efforts?

LISP Deployment

- LISP Pilot Network Operational
 - Deployed for nearly 2 years
 - ~32 sites across 7 countries
 - US, UK, BE, JP, UY, AU, DE
 - Uses the NX-OS Titanium Platform
 - IOS and OpenLISP platforms to be added
 - EID-Prefixes used
 - 153.16.0.0/16 and 2610:00d0::/32
 - RLOCs used
 - Current site attachment points to the Internet

LISP Deployment

- Map-Server/Map-Resolver is first layer of infrastructure
 - MS and MRs are colocated
 - xTRs register to 2 Map-Servers and use one of them for map resolution services
- We have ALT connected sites as well

LISP Deployment

- ALT is second layer of infrastructure
- LISP-ALT Infrastructure Built
 - GRE tunnels numbered out of 240.0.0.0/4
 - LISP-ALT uses 32-bit AS numbers
 - EID-prefixes BGP advertised from 'lisp' VRF
 - RLOCs in default VRF

Naming & Addressing

- Domain name `lisp4.net`
 - Contains hosts and routers from IPv4 EID space
- Domain name `lisp6.net`
 - Contains hosts and routers from IPv6 EID space

Naming & 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 & 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

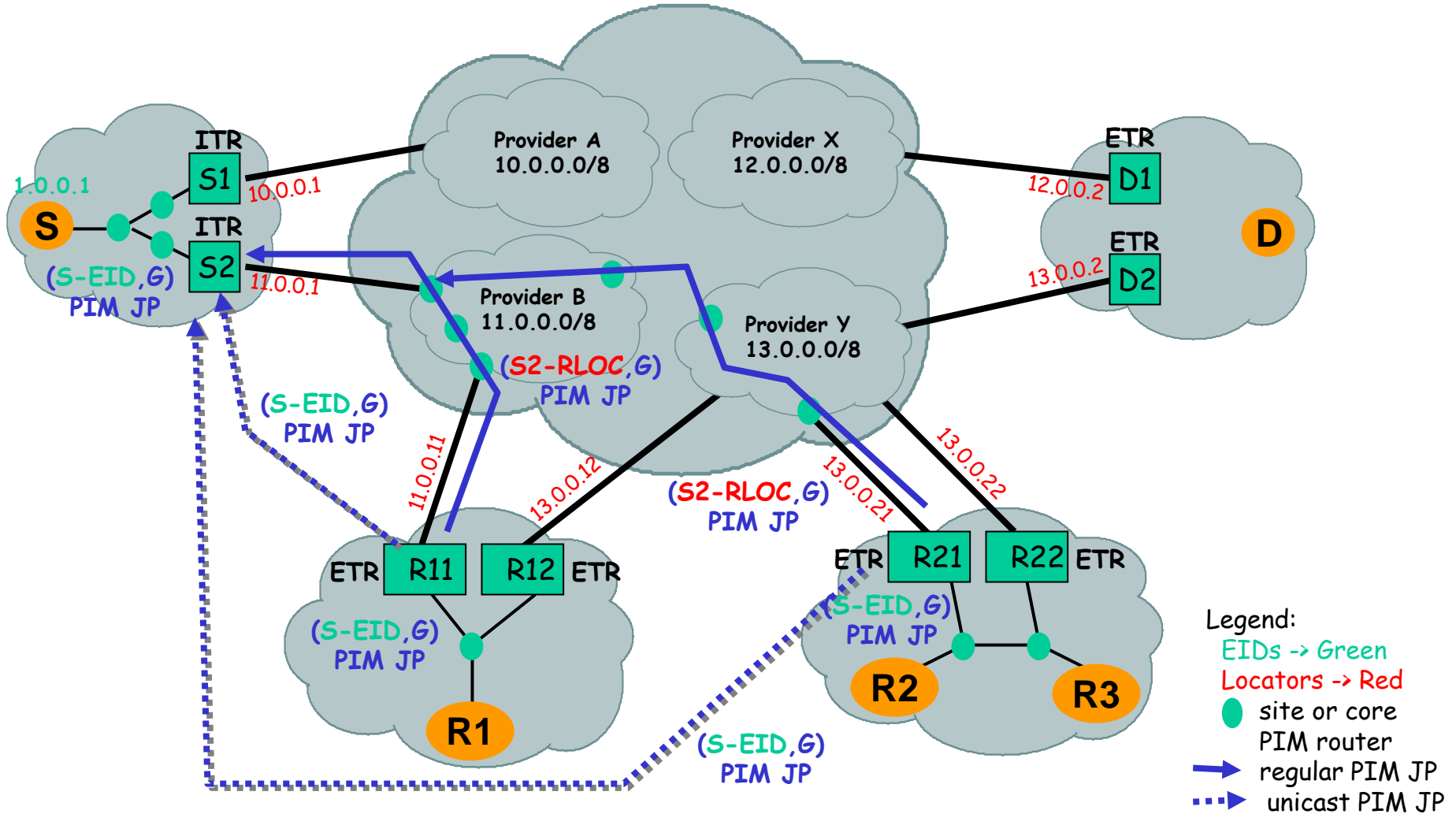
LISP Deployment

- 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:
 - UofO, ISC, and UY
 - <http://www.lisp6.net> reachable through IPv6 PTR
 - <http://www.lisp4.net> reachable through IPv4 PTR

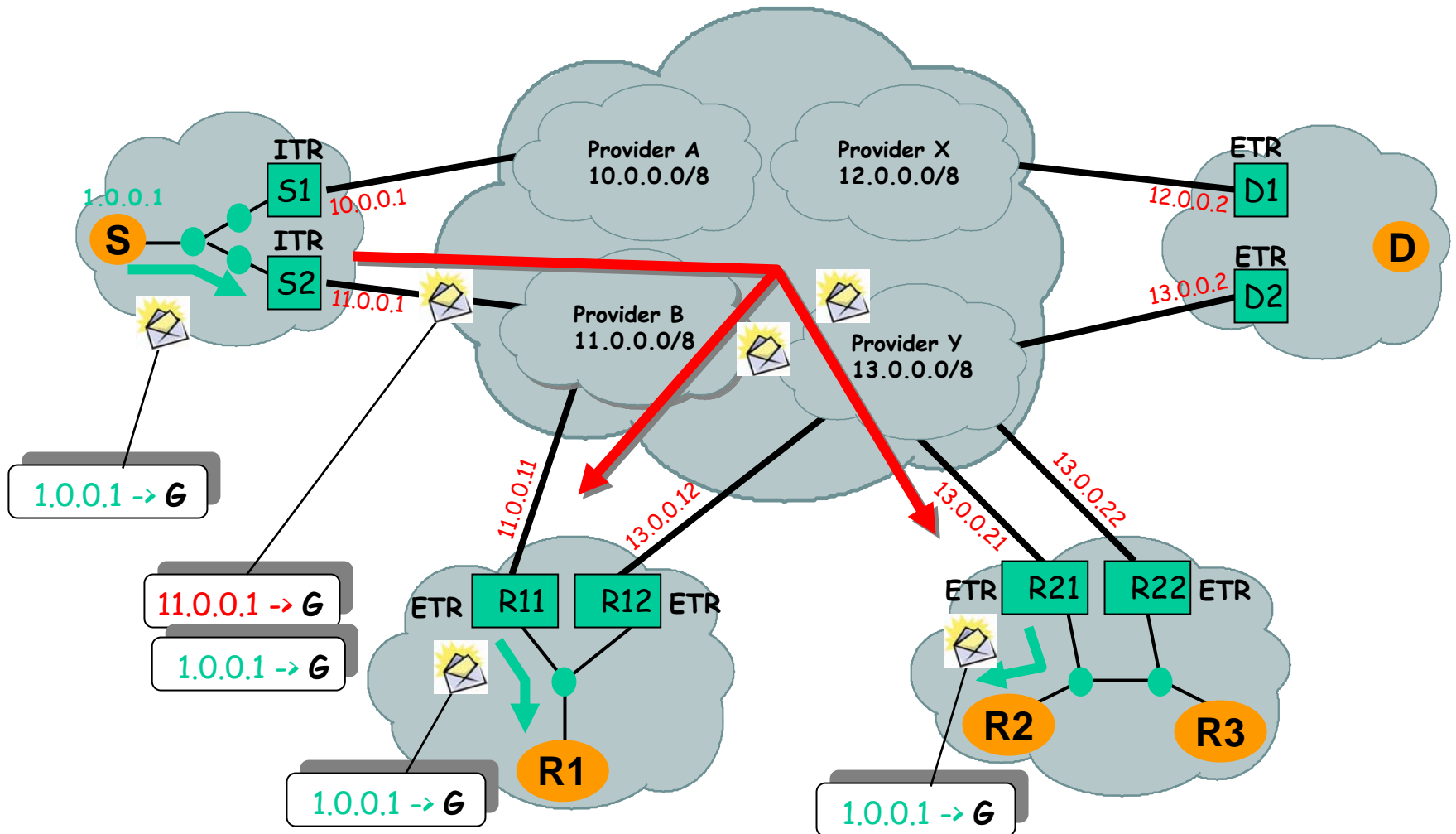
Multicast Packet Forwarding

- Group addresses have neither ID or Location semantics
 - G is topologically opaque - can be used everywhere
- $(S\text{-EID}, G)$
 - $S\text{-EID}$ is source host
 - G is group address receivers join to
 - State resides in source and receiver sites
- $(S\text{-RLOC}, G)$
 - $S\text{-RLOC}$ is ITR on multicast tree
 - G is group address receivers join to
 - State resides in core

Multicast Packet Forwarding



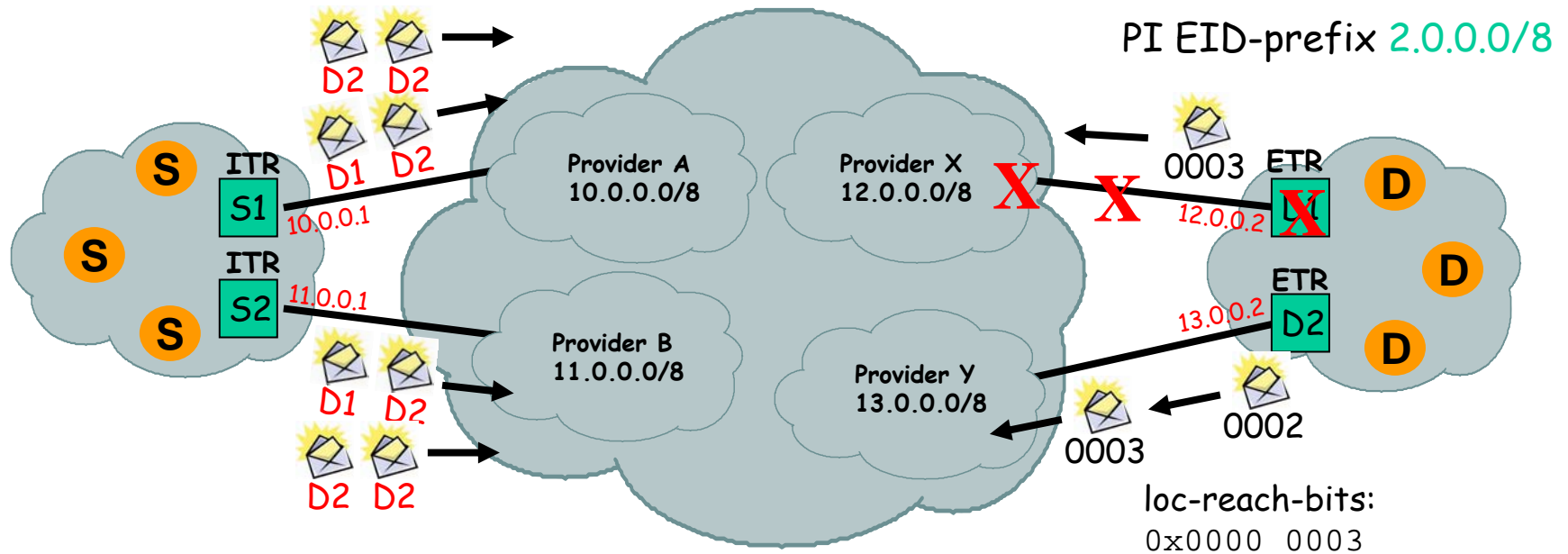
Multicast Packet Forwarding



Locator Reachability

- When RLOCs go up and down
 - Don't want this reflected in mapping database -- keep the `rate` factor small
- 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

How "loc-reach-bits" Work



Legend:
 EIDs -> Green
 Locators -> Red

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'

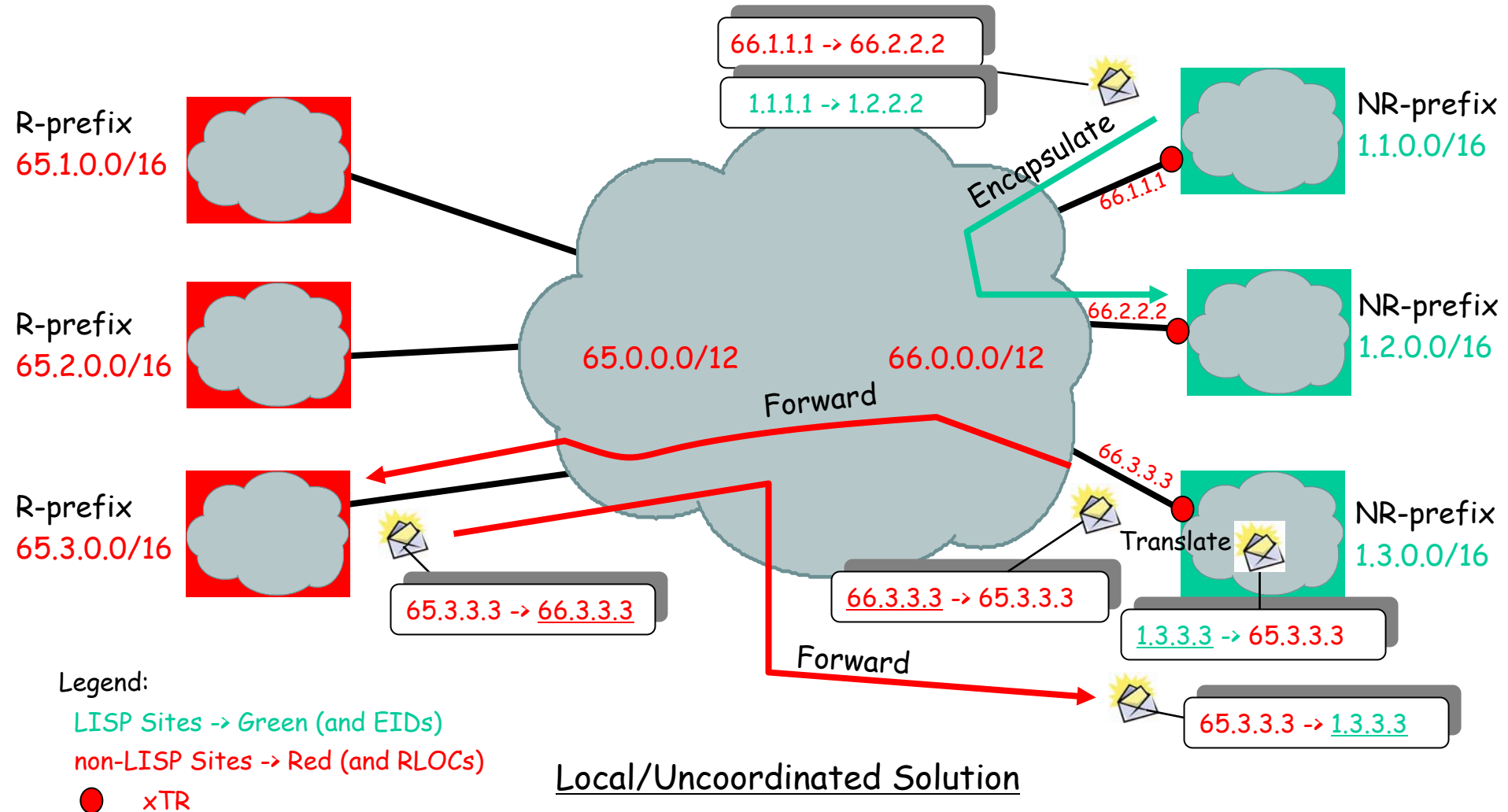
LISP Interworking

- These combinations must be supported
 - Non-LISP site to non-LISP site
 - Today's Internet
 - LISP site to LISP site
 - Encapsulation over IPv4 makes this work
 - IPv4-over-IPv4 or IPv6-over-IPv4
 - LISP-R site to non-LISP site
 - When LISP site has PI or PA routable addresses
 - LISP-NR site to non-LISP site
 - When LISP site has PI or PA non-routable addresses

LISP Interworking

- Is the destination an EID or not?
 - Asks the ITR which receives a packet from a source in it's site
- ITR attached to ALT
 - If the destination matches the ALT routing table, it's an EID
- ITR using a Map-Resolver
 - The map-cache will tell you to encapsulate or natively forward
 - Map-Resolvers send "Negative Map-Replies"

Interworking using LISP-NAT



Legend:

LISP Sites -> Green (and EIDs)

non-LISP Sites -> Red (and RLOCs)

● xTR

Interworking using PTRs

