IXPs

Internet Exchange Points (IXPs) + Public Peering

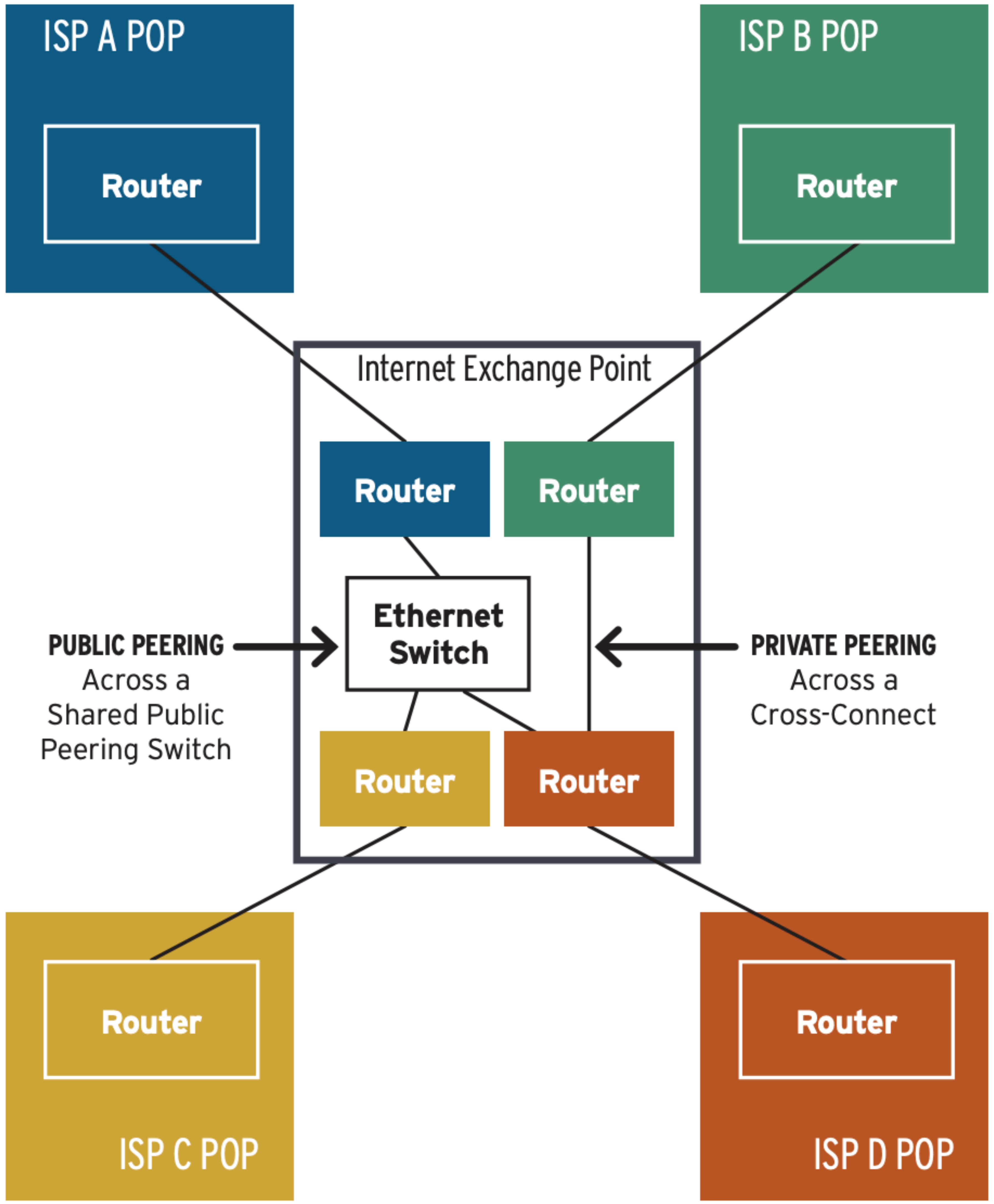
Carrier Hotels provide shared interconnect (switching fabric) between ISPs

Allow ISPs to BGP peer with a large number of organizations through a single link

Peerings that use shared fabric ("public interconnect") known as a "**Public Peering**"

You still have to negotiate the BGP peering with others on the exchange

In the U.S., private peering more common.
In Europe, public peering more common.



Many ISPs Will be Your Friend at an IXP

Most content providers will peer with you over public exchanges

So will cloud providers

Little downside for them not to if they're already on the exchange

Most ISPs (even Tier 1s) will sell you transit at an IXP (via private peering)

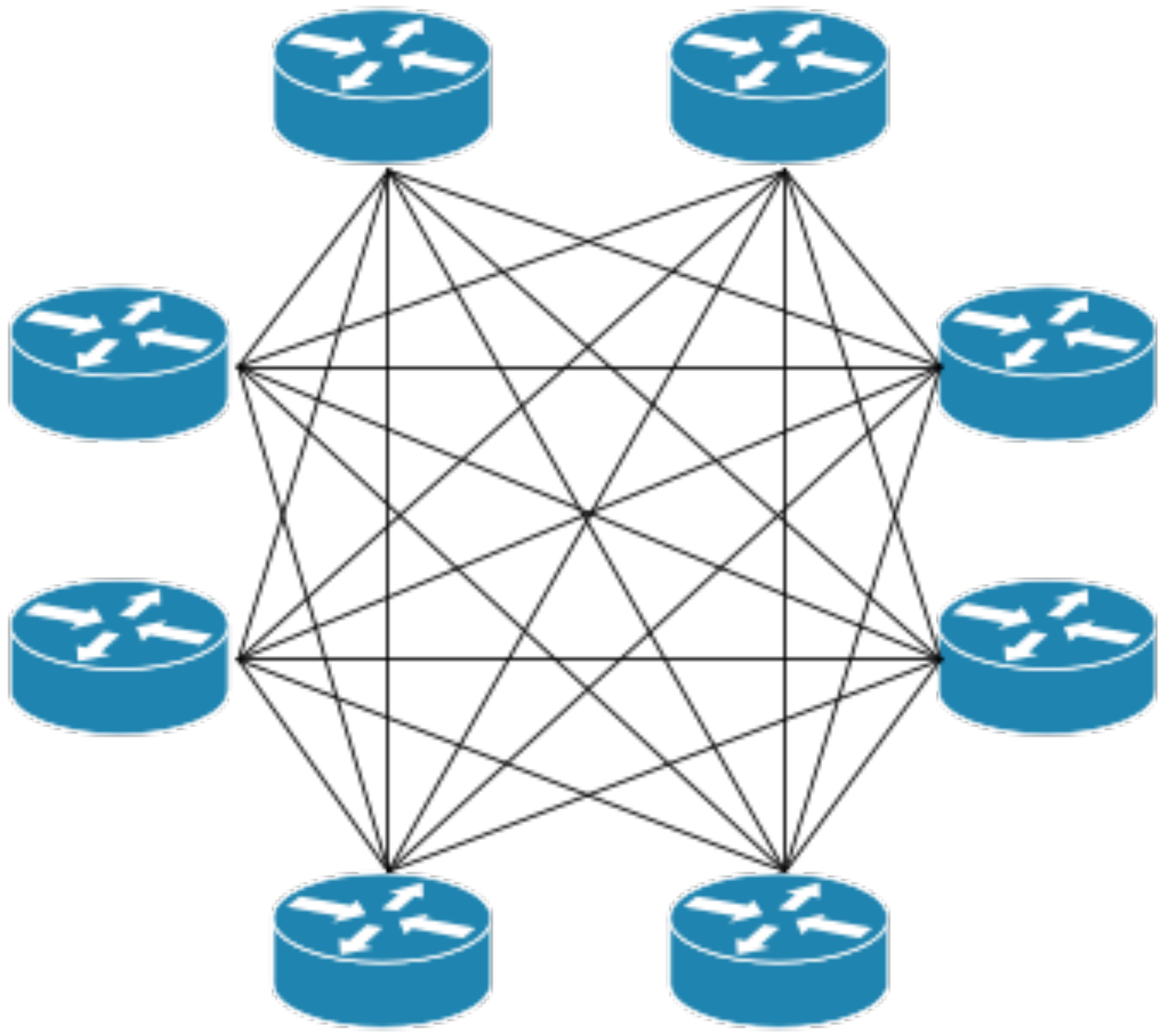
Multilateral Peering Exchanges

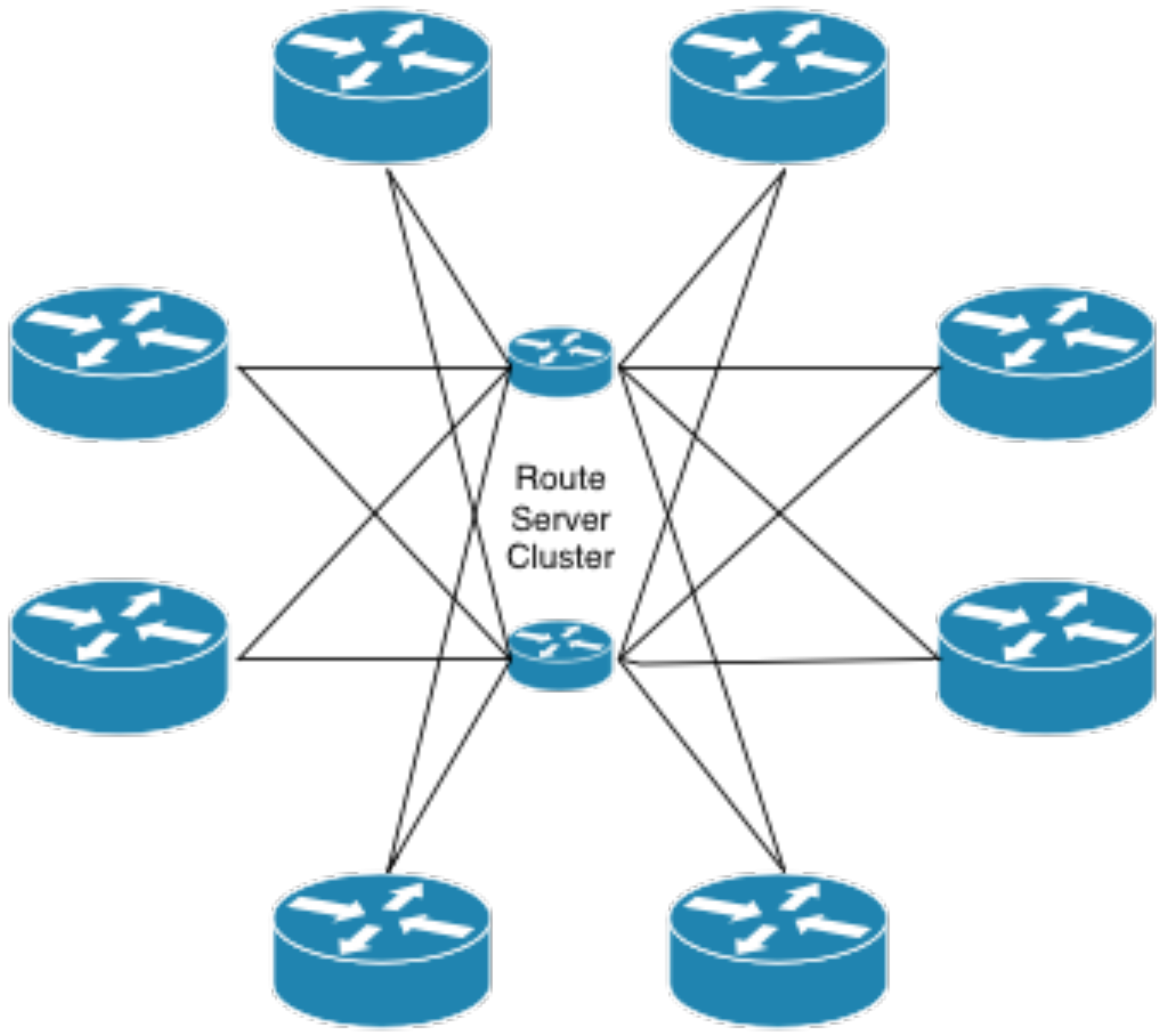
Typically, you'd establish peering relationships with others at an exchange

This can become cumbersome if there are a large number of players (could be 100s or 1000s of organizations at a single IXP!)

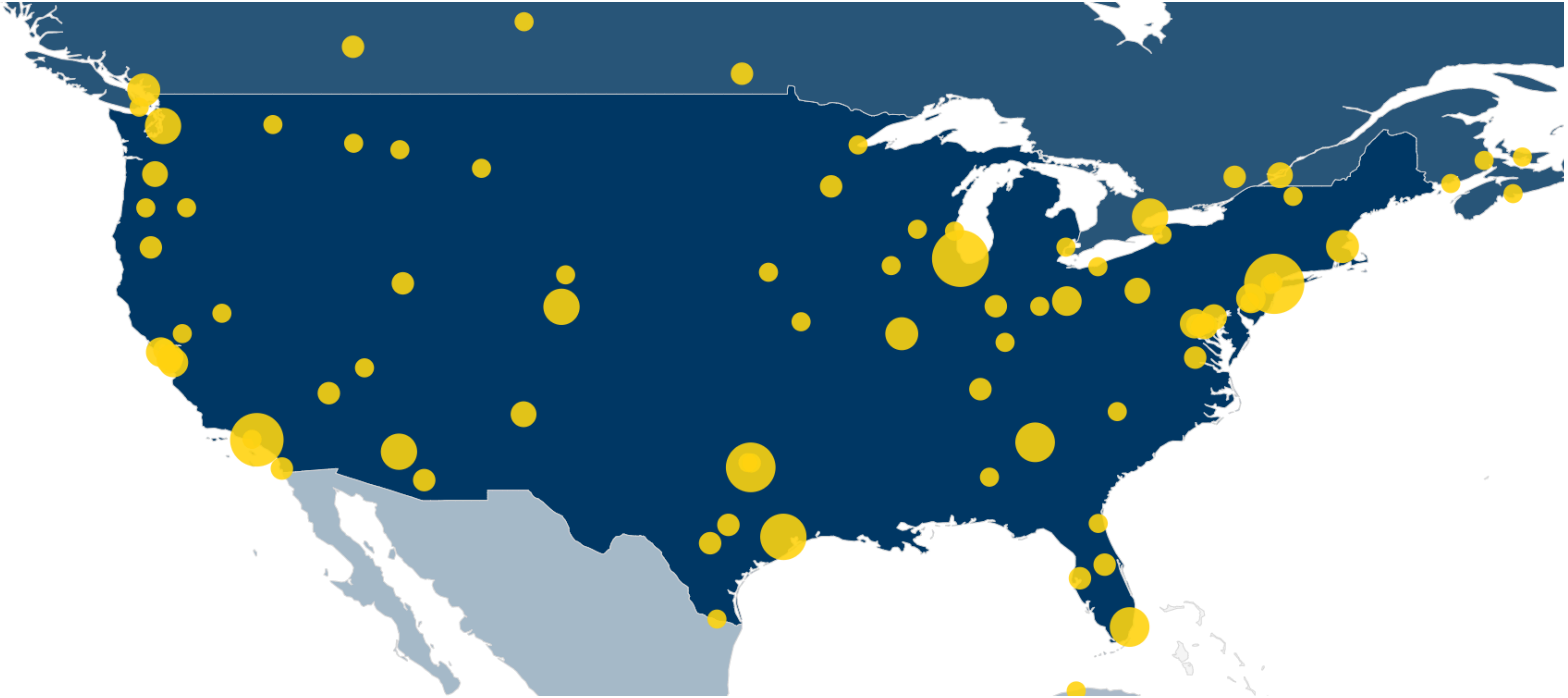
Route server allows those with open policies to all BGP peer with only a single entity to both advertise its routes and collect routes from others on the exchange

Known as **Multilateral Peering**

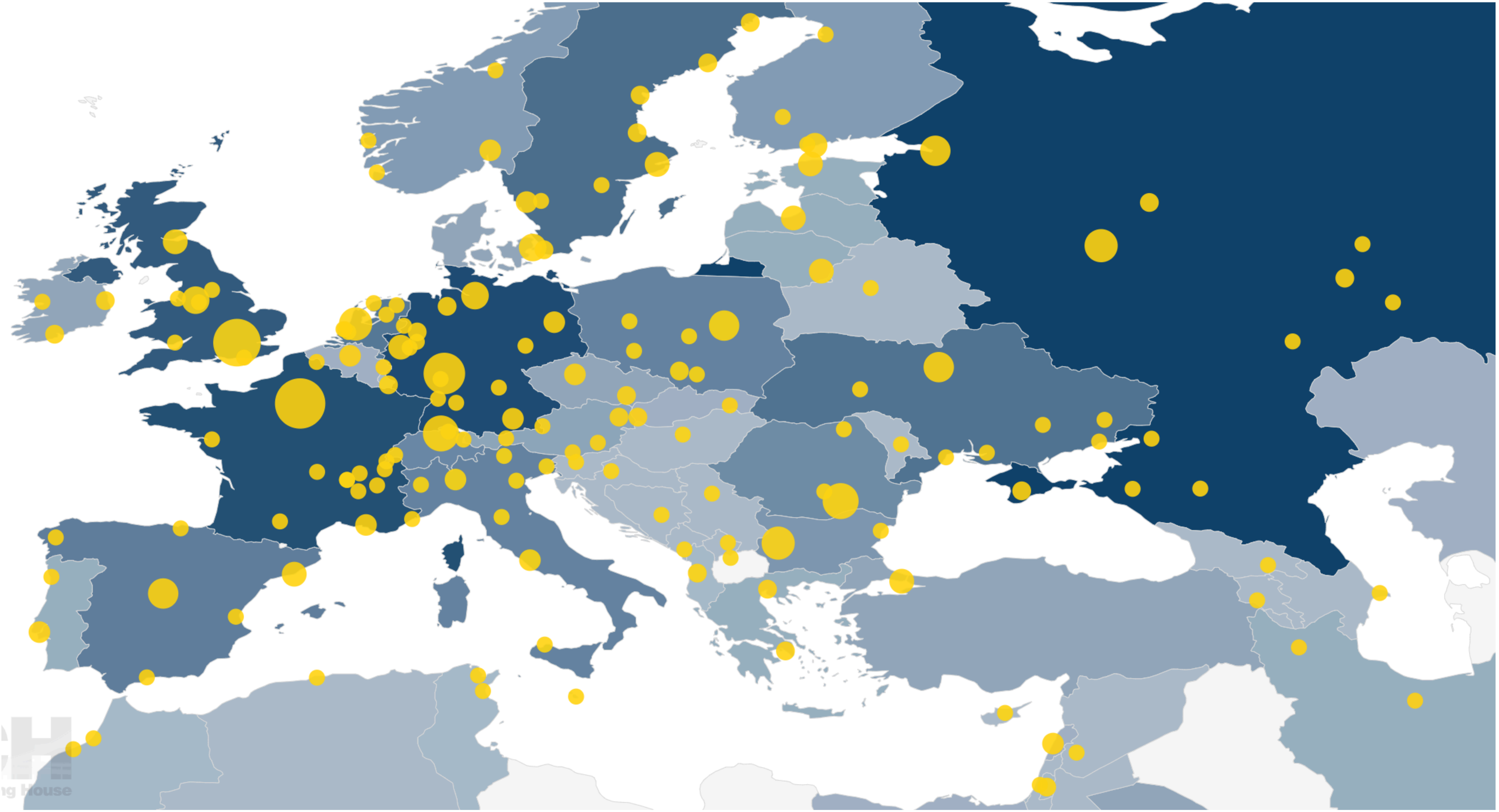




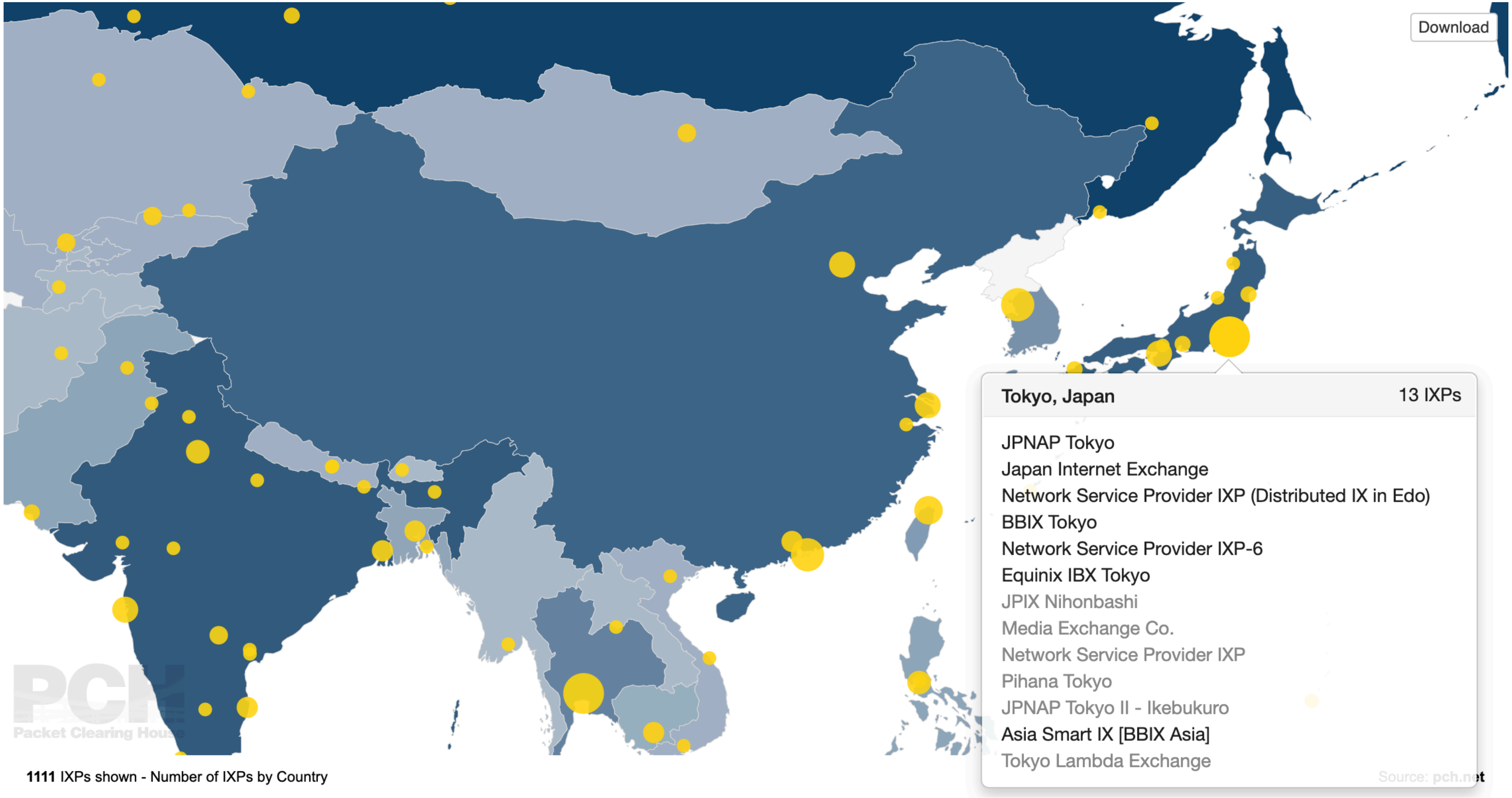
Packet Clearing House — Where are U.S. IXPs?



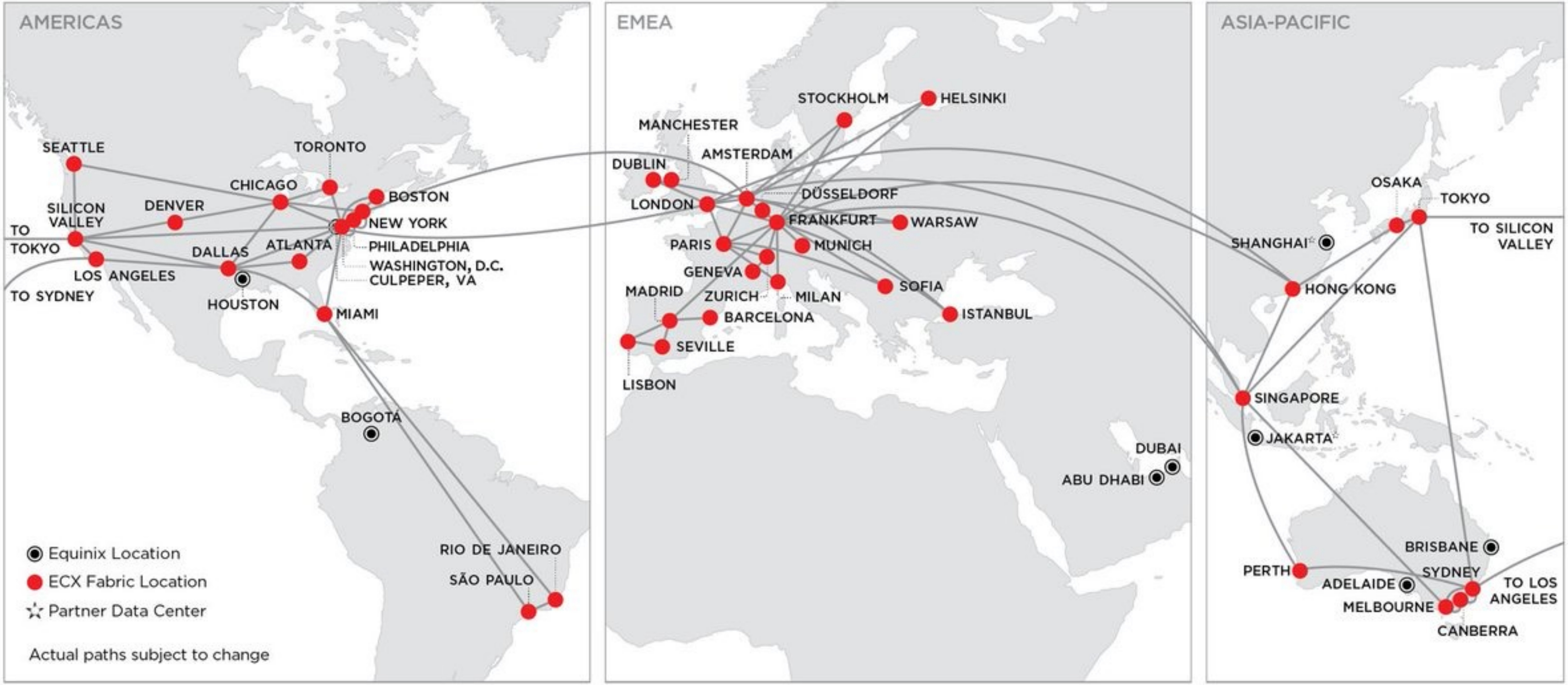
Packet Clearing House — Europe



Packet Clearing House — Asia



Equinix — Largest Commercial IXP Provider



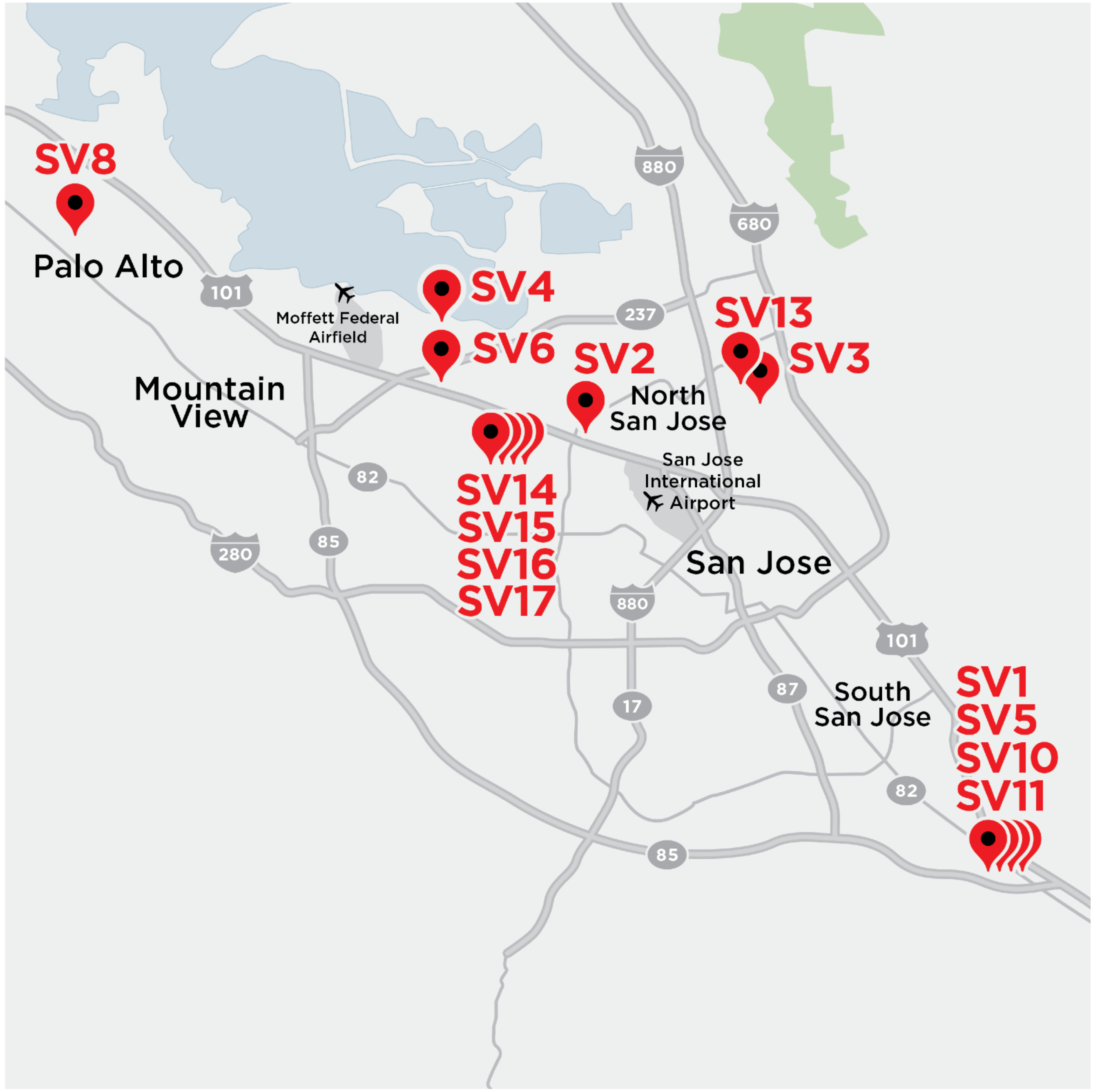
Equinix Internet Exchange San Jose (Bay Area)

Facilities:

- Equinix SV1/SV5/SV10 - San Jose
- Equinix SV2 - Santa Clara
- Equinix SV3 - San Jose
- Equinix SV4 - Sunnyvale
- Equinix SV8 - Palo Alto

~210 (publicly listed) participants on San Jose Exchange. ~101 on Palo Alto Exchange.

~500 organizations listed between those 5 hotels (many more than on public exchange)



European Example: AMS-IX

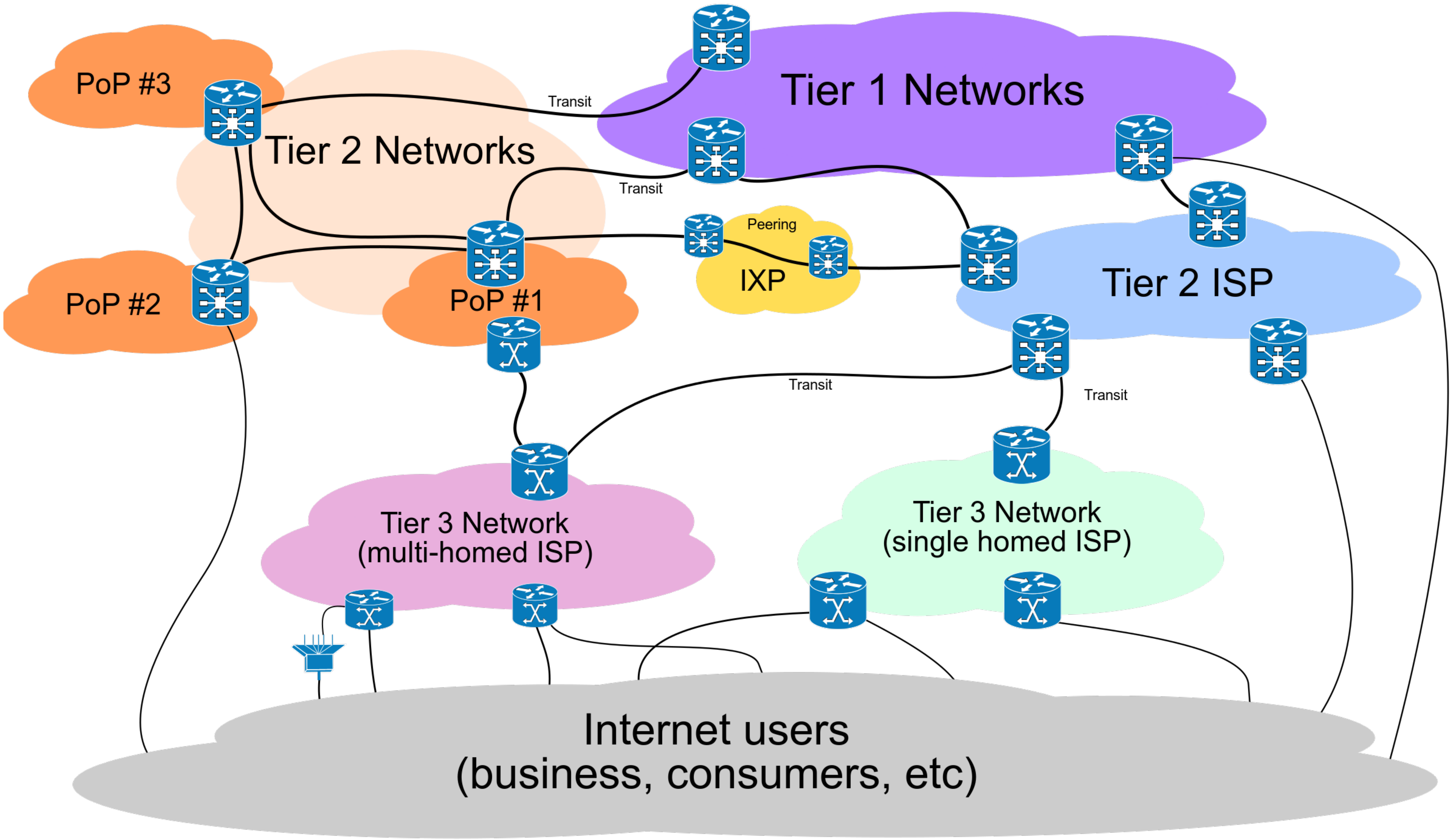
AMS-IX Points of Presence

	Data Centre	Phone	Address
1	NorthC Amsterdam (AMS01)	+31 (0)20 486 9773	Kabelweg 48a, Amsterdam
2	Digital Realty AMS17	-	Science Park 120, Amsterdam
3	Digital Realty AMS04	+31 (0)20 480 4415	H.J.E. Wenckebachweg 127, Amsterdam
4	Equinix AM1/2	+31 (0)53 434 0570	Luttenbergweg 4, Amsterdam
5	Equinix AM3	+31 (0)20 808 0015	Science Park 610, Amsterdam
6	Equinix AM5	+31 (0)20 592 8263	Schepenbergweg 42, Amsterdam
7	Equinix AM6	+31 (0)53 436 2666	Duivendrechtsekade 80A, Amsterdam
8	Equinix AM7	+31 (0)53 434 0570	Kuiperbergweg 13, Amsterdam
9	EuNetworks	+31 (0)20 354 8098	Paul van Vlissingenstraat 16, Amsterdam
10	Iron Mountain	+31 (0)20 316 5170	J.W. Lucasweg 35, Haarlem
11	Global Switch	+31 (0)20 666 6300	Henk Sneevlietweg 2-6, Amsterdam
12	Interxion	+31 (0)20 880 7700	Tupolevlaan 101, Schiphol-Rijk
13	Interxion	+31 (0)20 560 6600	Science Park 121, Amsterdam
14	Nikhef	+31 (0)20 592 2037	Science Park 105, Amsterdam

Members	882[1]
Ports	1,438 [1]
Peers	1,316 [1]
Peak in	9.022 Tb/s[2]
Peak out	10.287 Tb/s[2]
Daily in (avg.)	6.42[8] Tb/s[2]
Daily out (avg.)	6.43[8] Tb/s[2]

European Example: AMS-IX

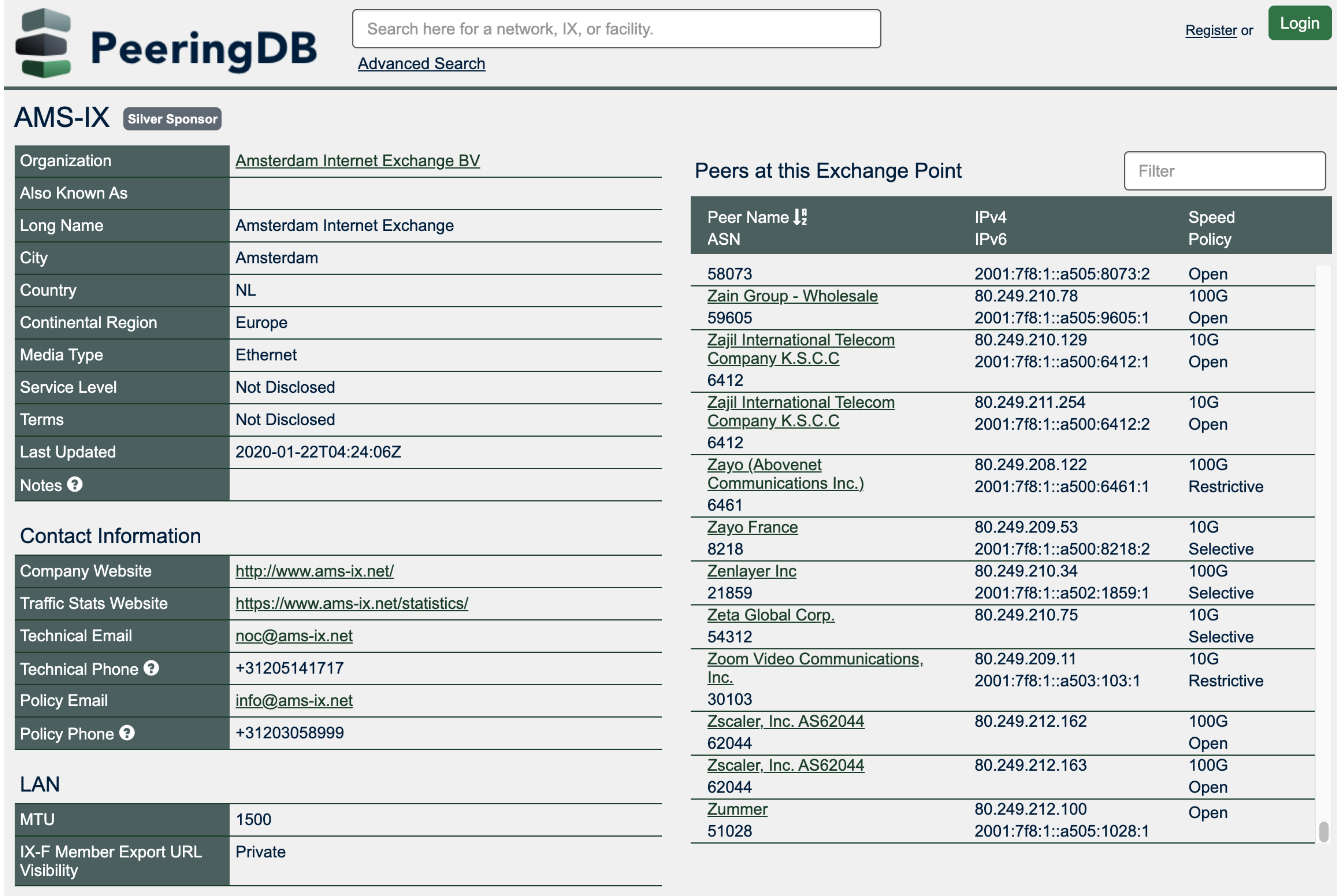
Most details about AMS-IX are public at https://www.ams-ix.net/


Tier 1 Networks
Transit
Tier 2 Networks
PoP #3
Transit
Peering
IXP
PoP #2
PoP #1
Tier 2 ISP
Transit
Transit
Tier 3 Network
(multi-homed ISP)
Tier 3 Network
(single homed ISP)
Internet users
(business, consumers, etc)

PeeringDB

PeeringDB is a free database of networks, IXPs, and facilities

https:/peeringdb.com/



PeeringDB

Search here for a network, IX, or facility.

Advanced Search

Register or Login

AMS-IX Silver Sponsor

Organization	Amsterdam Internet Exchange BV
Also Known As	
Long Name	Amsterdam Internet Exchange
City	Amsterdam
Country	NL
Continental Region	Europe
Media Type	Ethernet
Service Level	Not Disclosed
Terms	Not Disclosed
Last Updated	2020-01-22T04:24:06Z
Notes	

Contact Information

Company Website	http://www.ams-ix.net/
Traffic Stats Website	https://www.ams-ix.net/statistics/
Technical Email	noc@ams-ix.net
Technical Phone	+31205141717
Policy Email	info@ams-ix.net
Policy Phone	+31203058999

LAN

MTU	1500
IX-F Member Export URL Visibility	Private

Peers at this Exchange Point

Filter

Peer Name / ASN	IPv4 / IPv6	Speed / Policy
58073	2001:7f8:1::a505:8073:2	Open
Zain Group - Wholesale / 59605	80.249.210.78 / 2001:7f8:1::a505:9605:1	100G / Open
Zajil International Telecom Company K.S.C.C / 6412	80.249.210.129 / 2001:7f8:1::a500:6412:1	10G / Open
Zajil International Telecom Company K.S.C.C / 6412	80.249.211.254 / 2001:7f8:1::a500:6412:2	10G / Open
Zayo (Abovenet Communications Inc.) / 6461	80.249.208.122 / 2001:7f8:1::a500:6461:1	100G / Restrictive
Zayo France / 8218	80.249.209.53 / 2001:7f8:1::a500:8218:2	10G / Selective
Zenlayer Inc / 21859	80.249.210.34 / 2001:7f8:1::a502:1859:1	100G / Selective
Zeta Global Corp. / 54312	80.249.210.75	10G / Selective
Zoom Video Communications, Inc. / 30103	80.249.209.11 / 2001:7f8:1::a503:103:1	10G / Restrictive
Zscaler, Inc. AS62044 / 62044	80.249.212.162	100G / Open
Zscaler, Inc. AS62044 / 62044	80.249.212.163	100G / Open
Zummer / 51028	80.249.212.100 / 2001:7f8:1::a505:1028:1	Open

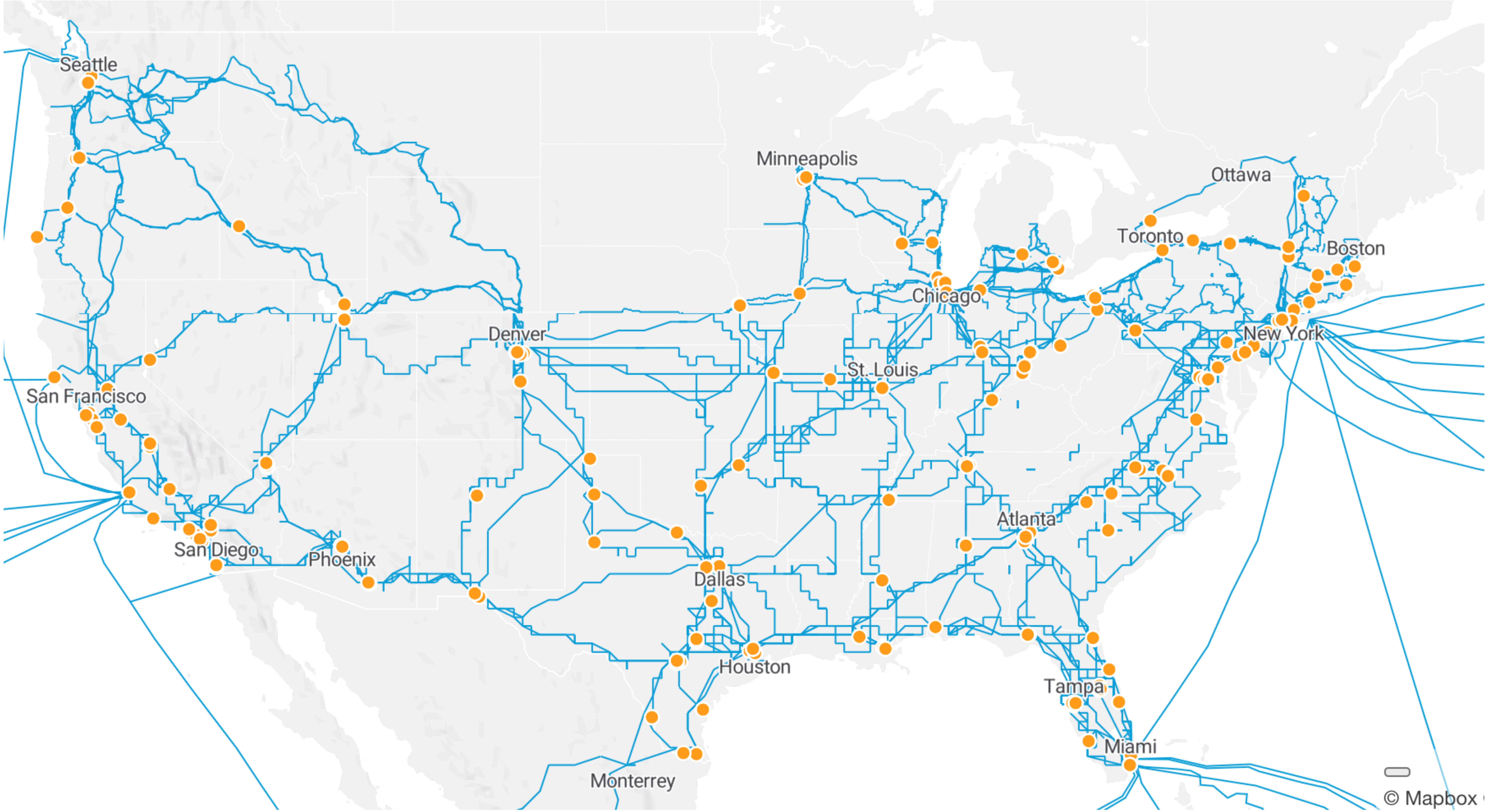
Tier 1 Network Backbone

AT&T

GTT


Vancouver
Seattle
Minneapolis
Montreal
Moncton
Toronto
Halifax
Albany
Osaka
Tokyo
Chicago
Detroit
Buffalo
Boston
Cleveland
New York
Pittsburgh
Salt Lake City
Columbus
Philadelphia
Kansas City
Ashburn
San Francisco
St. Louis
Denver
Washington, D.C.
San Jose
Las Vegas
Hong Kong
Charlotte
Dallas
Los Angeles
Phoenix
San Diego
Atlanta
Baton Rouge
Jacksonville
Houston
Miami
Singapore
Mexico City
to Sydney, to Melbourne
to Sydney
to Sydney
to São Paulo
to São Paulo
Dublin
Cork
Manchester
London
Reading
Amsterdam
Hamburg
Copenhagen
Berlin
Warsaw
Brussels
Frankfurt
Prague
Paris
Munich
Vienna
Zürich
Budapest
Geneva
Milan
Bordeaux
Bucharest
Sofia
Marseille
Rome
Istanbul
Barcelona
Madrid
to Johannesburg
to Johannesburg

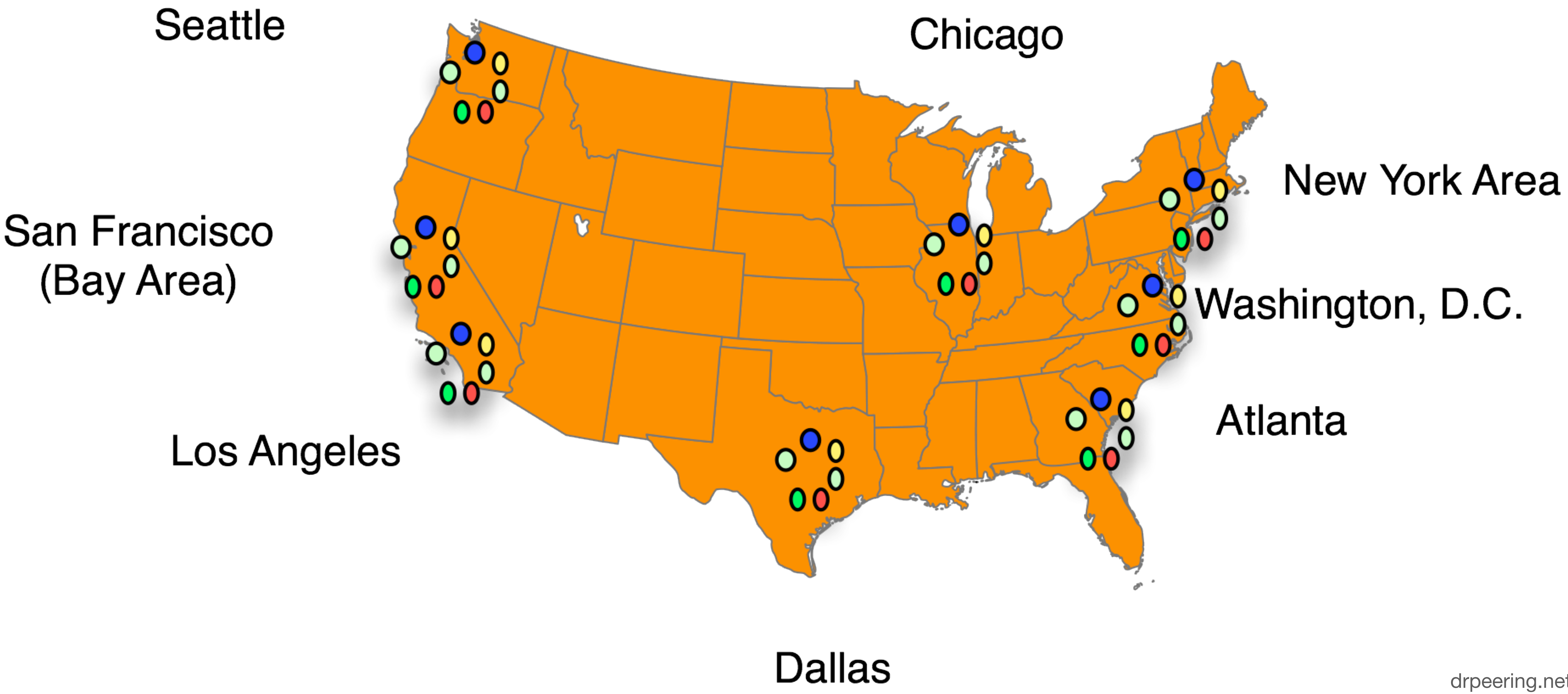
Lumen



Seattle
Minneapolis
Ottawa
Toronto
Boston
Chicago
Denver
New York
St. Louis
San Francisco
Atlanta
San Diego
Phoenix
Dallas
Houston
Tampa
Miami
Monterrey
© Mapbox


Hurricane Electric
Edmonton
Saskatoon
Calgary
Winnipeg
Vancouver
Seattle
Spokane
Missoula
Billings
Fargo
Minneapolis
Quebec City
Moncton
Ottawa
Montreal
Halifax
Hillsboro
Portland
Boise
Rochester
Madison
Detroit
Toronto
Portland
Cheyenne
Des Moines
Chicago
Cleveland
Boston
Salt Lake City
Omaha
Davenport
Pittsburgh
NYC
Garden City
Reno
Denver
Kansas City
Indianapolis
Columbus
Philadelphia
Ashburn
Sacramento
Bay Area
Saint Louis
Las Vegas
Albuquerque
Oklahoma City
Nashville
Virginia Beach
Charlotte
Phoenix
Little Rock
Los Angeles
Atlanta
Dallas
Tucson
Montgomery
Austin
Jacksonville
San Antonio
Houston
Tampa
Miami

The 8 U.S. Interconnection Regions

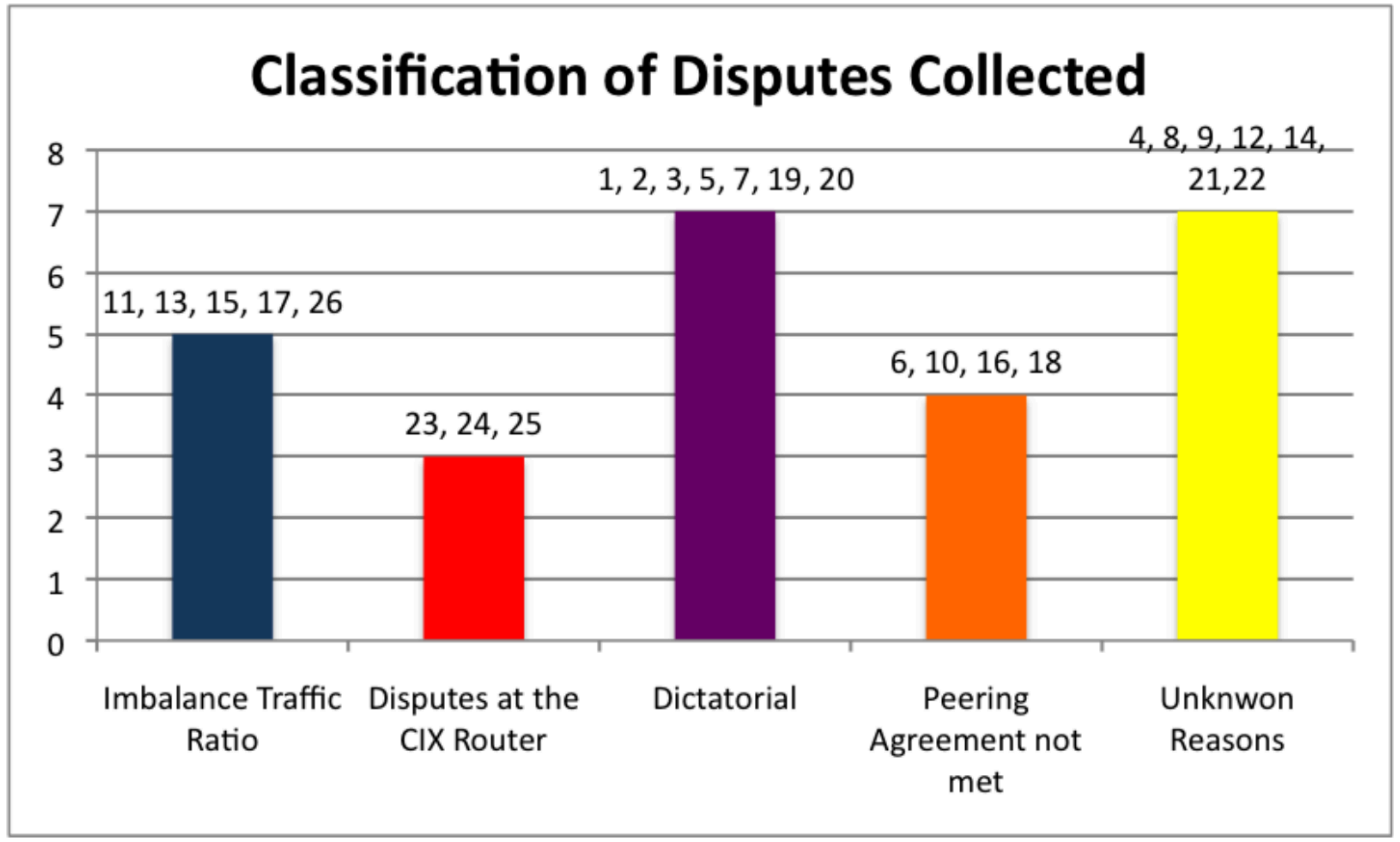


Peering Disputes

What happens if Tier-1s De-Peer?

"Cogent has decided not to exchange traffic directly with TeliaSonera's AS 1299 or indirectly with AS 1299 through a third-party provider," Telia told its customers. "As a result, Cogent has partitioned the Internet and disrupted the flow of traffic between Cogent and TeliaSonera customers."

“Cogent has been controversial in the ISP market for low bandwidth pricing and its public disputes over peering with AOL (2003),Level 3 Communications (2005), France Telecom (2006), Limelight Networks (2007), Telia Carrier (March 2008), and Sprint Nextel (October 2008).”



Source: Anatomy of the Internet Peering Disputes

a_cute_epic_axis · 3 yr. ago

Packet Whisperer

Whenever there is a peering dispute between Cogent and anyone else, assume first that Cogent is in the wrong and willing to use underhanded tactics and cronyism to resolve it. Only deviate from this when comprehensive proof shows otherwise. They no longer get the benefit of the doubt due to their past bullshit.

33 Reply Share ···

Flattening

The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh

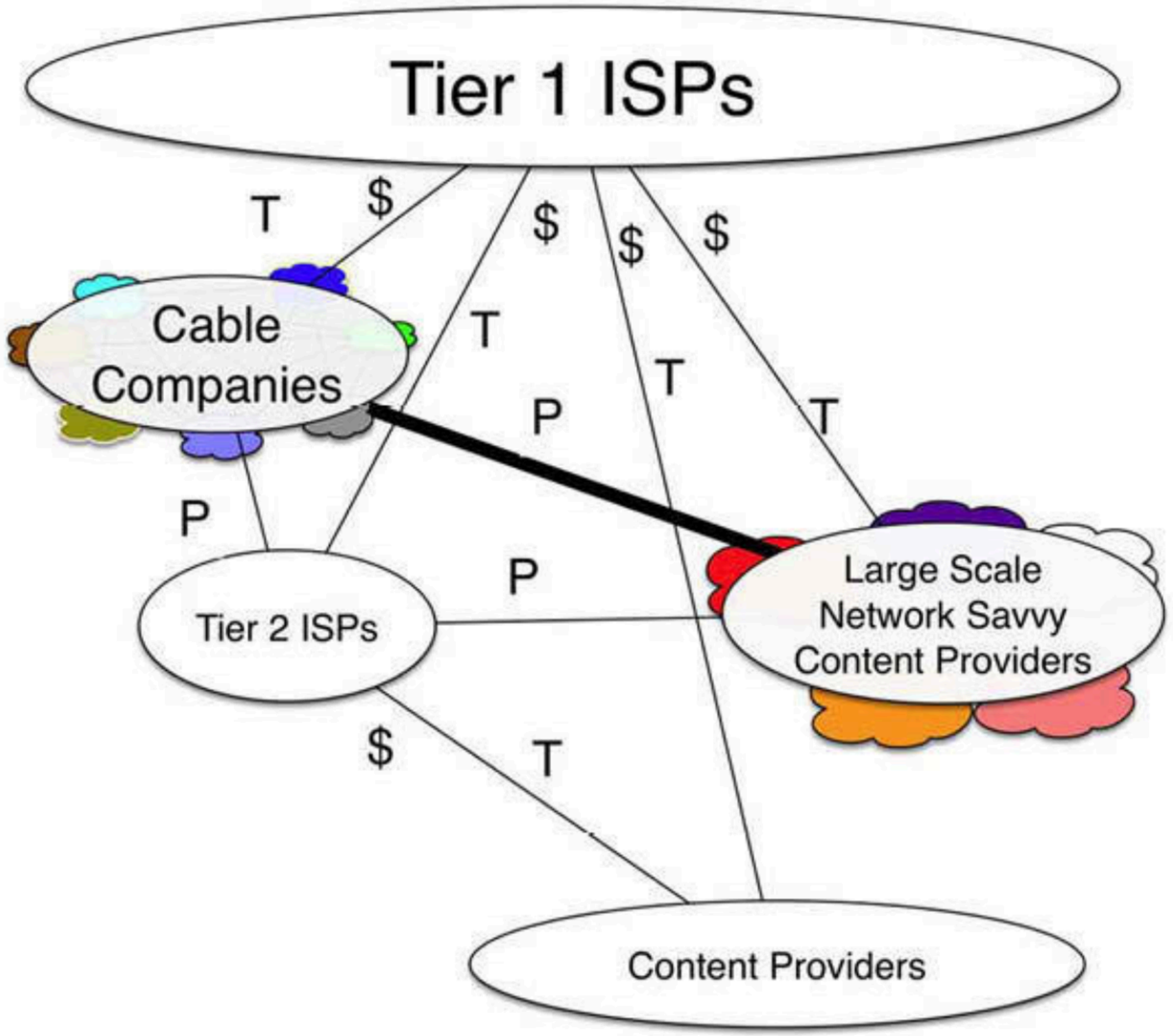
Three Changes in the World:

- An increasing fraction of Internet traffic originates from a few CPs or CDNs (e.g., Google, YouTube, Akamai, Cloudflare). This shift is due to the large penetration of video streaming.

- The major CDNs and CPs have expanded to almost every region of the developed world, so that they can be co-located with many ASes at Internet Exchange Points (IXPs).

- IXPs have increased rapidly in number, making it easy and cheap for an AS to establish peering links with other ASes co-located at the same IXP.

Tier 1 Peerings Mattered Less




ars TECHNICA
BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

POLICY —

How Comcast became a toll-collecting, nuke-wielding hydra

Comcast wants cash to deliver cached Netflix traffic to its subscribers. Has ...

NATE ANDERSON - 11/30/2010, 1:35 PM

A Small Transit Provider Case Study

AS19653 – Small Transit Provider in Climax, Michigan
Founded in 1911 as an Independent Telephone Company.
Started as a CLEC in 1996.
Independent ILEC-CLEC-ISP. CLLI = CLMXMIXI

2011 – Joined NANOG

Telephone Company (ILEC-CLEC)
Tier 3 ISP
100% transit (two OC-12s)

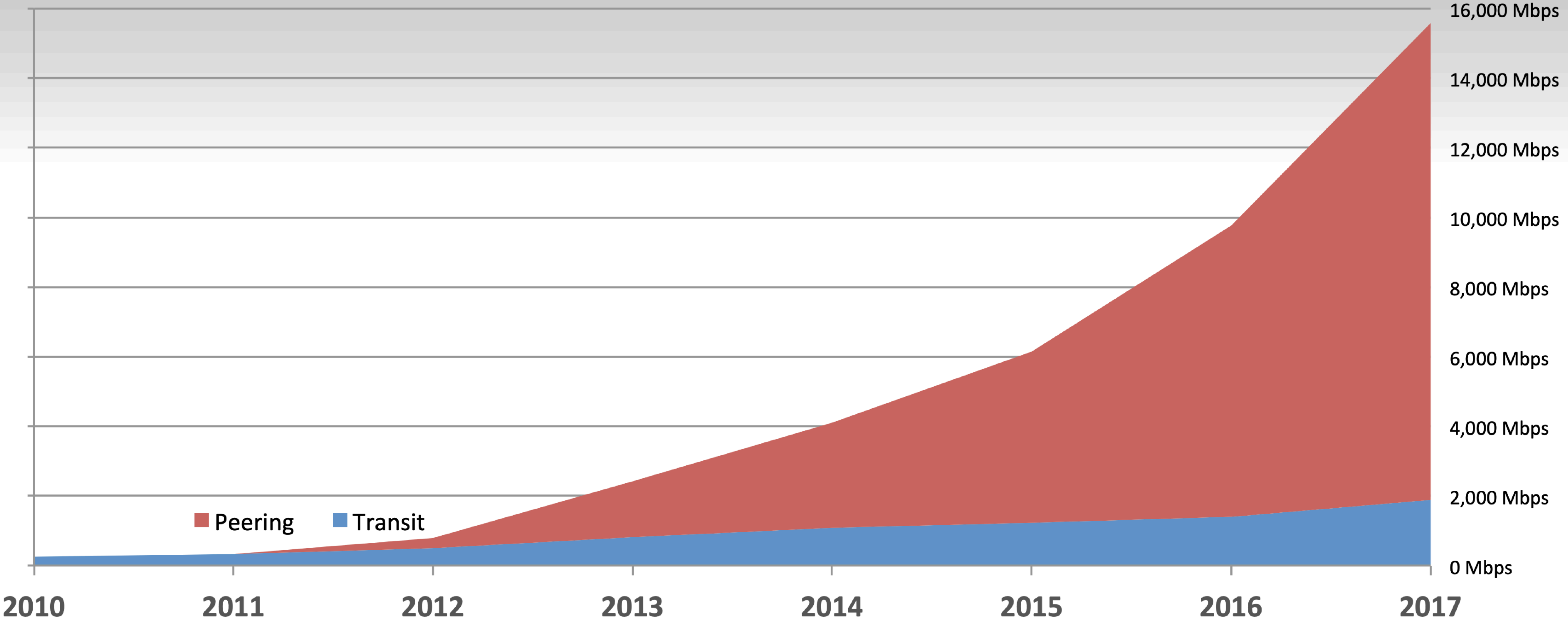
CLMX
AS19653

2017 – (after 18 NANOGs)

Packet Optical Service Provider
Tier 2 ISP
88% Peering
12% transit
More than 100G in upstream ports

Source: The Internet is Flat: Revisited

AS19653 Evolution of Transit to Peering

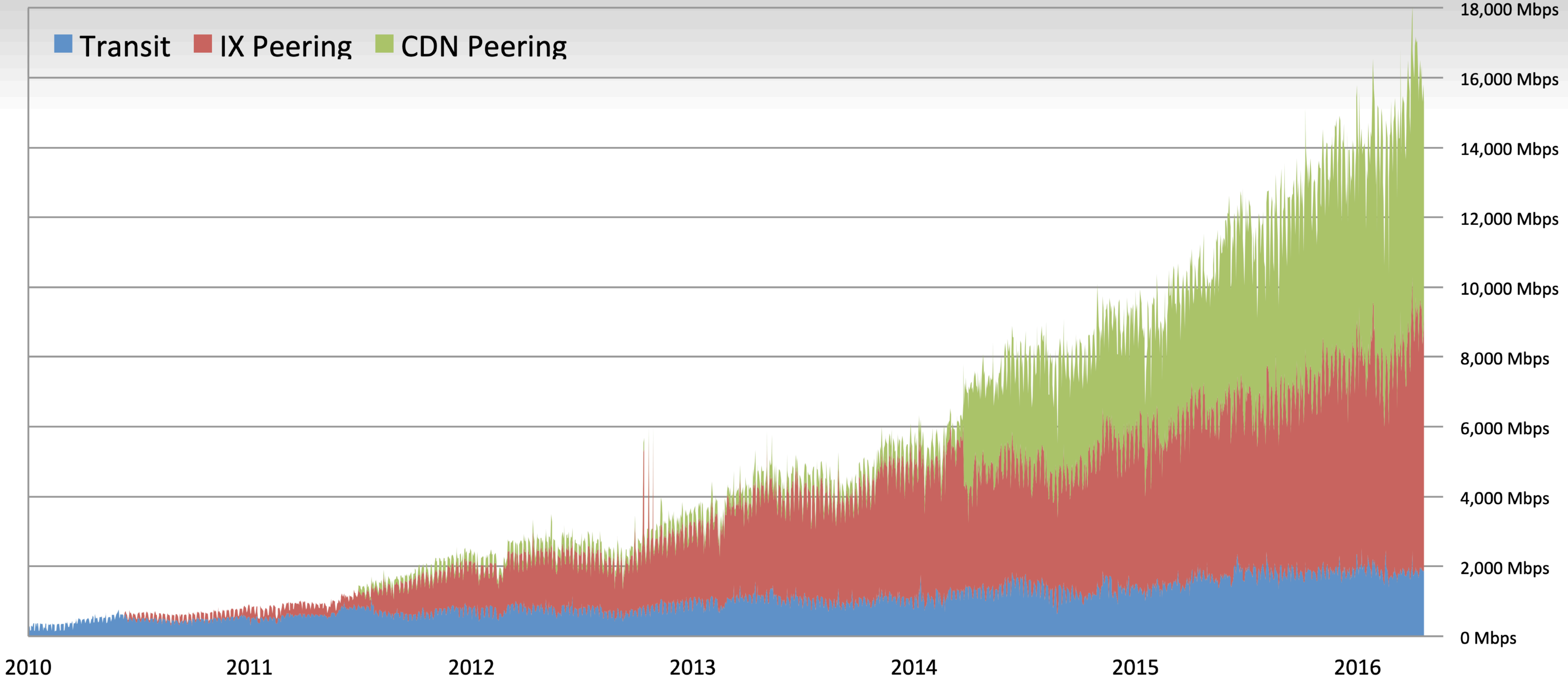


Peering	0%	0%	36%	68%	74%	80%	86%	88%
Transit	100%	100%	64%	32%	26%	20%	14%	12%

A Small Transit Provider Case Study

SOURCE: AS19653

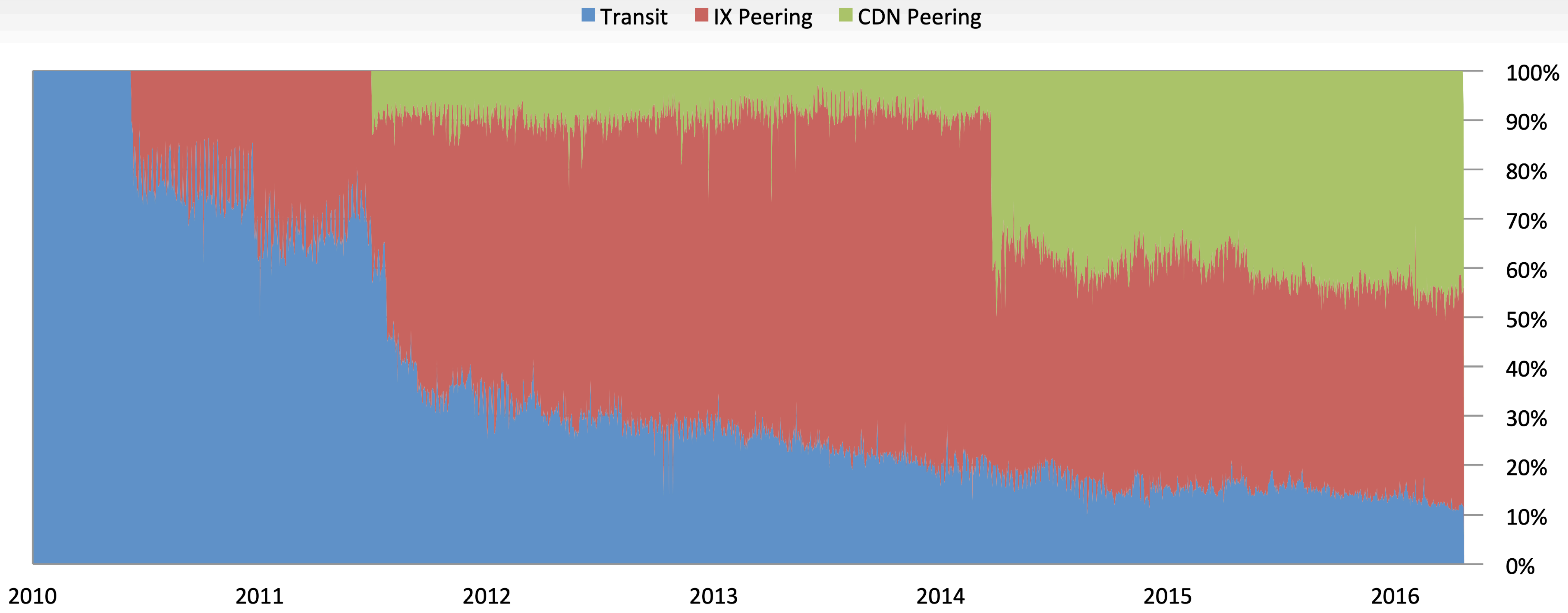
AS19653 From Transit to IXP Peering to CDN



A Small Transit Provider Case Study

SOURCE: AS19653

Percentage of Total Traffic AS19653 Transit/Peering/CDN



A Small Transit Provider Case Study

SOURCE: AS19653

Modern Routing Practices

CS249i: The Modern Internet

Stanford University

Project 1 Notes

A Few Notes on BGP

BGP — routers share *most efficient* (shortest path) with their neighbors

- You don't see *all* routes. Rather routes to all routed prefixes.

Your Routing Information Base (RIB) is organized by routed prefix

- A bunch of details, only a few really matter to you:
    - (Routed) Prefix
    - AS Path
    - Next Hop

Traffic is sent to the route that matches the most specific prefix

`rib_ipv4_unicast` Entry (in JSON)


```
{
  "sub_type":"rib_ipv4_unicast",
  "sequence_number":0,
  "prefix":"103.127.54.0/24"
  "entries":[
    {
      "peer_index": 1,
      "orginated_time": 1632281047,
      "path_identifier":0,
      "path_attributes":
        {  "type": 2,
          "as_paths": [
            {
              "segment_type":2, "num":8,
              "asns":[65400, 65105, 32, 46749, 46749, 6939, 137367, 17995]
            }
          ]
        },
        { "type":3, "nexthop": "171.67.69.32" },
        { "type":1, "value": 0 },
        { "type":4, "metric": 0 },
        { "type":8, "communities": [32, 454801053] }
      ]
    }
  ],
  "route_family": 65537
}
```

Traceroute + Router Interfaces

```
traceroute to google.com (142.250.72.174), 30 hops max
1   _gateway (171.67.69.32)  0.388 ms  0.369 ms  0.360 ms
2   * * *
3   10.214.4.249 (10.214.4.249)  1.043 ms
4   dc-sf-rtr-vl12.SUNet (171.66.0.207)  1.082 ms
5   dc-sfo-agg4--stanford-100g.cenic.net (137.164.23.178)  1.943 ms
6   dc-svl-agg8--sfo-agg4-100gbe.cenic.net (137.164.11.92)  2.532 ms
7   dc-svl-agg10--svl-agg8-300g.cenic.net (137.164.11.80)  1.860 ms
8   74.125.147.146 (74.125.147.146)  2.982 ms
9   108.170.242.254 (108.170.242.254)  3.95 ms
10  142.250.234.60 (142.250.234.60)  4.26 ms
11  142.250.211.208 (142.250.211.208)  10.564 ms
```

Looking Glass Servers

Useful to know BGP state at different routers — ISPs will often let you interrogate their public routing infrastructure — known as **Looking Glass** service

core1.ash1.he.net> show ip bgp routes detail 8.8.8.8

Matching Routes	28
Status Codes	**A** - Aggregate **B** - Best **b** - Not Install Best **C** - Confederation eBGP **D** - Damped **E** - eBGP **H** - History **I** - iBGP **L** - Local **M** - Multipath **m** - Not Installed Multipath **S** - Suppressed **F** - Filtered **s** - Stale **x** - Best-External

Status	Network	Next Hop	Learned	Metric	LocPrf	Weight	Path	Origin	ROA
BMEx	8.8.8.0/24	206.53.170.23	206.53.170.1 (64216)	0	100	0	15169	IGP	✅
ME	8.8.8.0/24	206.53.170.23	206.53.170.2 (64216)	0	100	0	15169	IGP	✅
ME	8.8.8.0/24	206.126.236.21	206.126.236.21 (15169)	0	100	0	15169	IGP	✅
ME	8.8.8.0/24	206.126.237.242	206.126.237.242 (15169)	0	100	0	15169	IGP	✅
I	8.8.8.0/24	206.83.10.13	216.218.252.230 (6939)	15	100	0	15169	IGP	✅
I	8.8.8.0/24	198.32.118.39	216.218.252.171 (6939)	79	100	0	15169	IGP	✅
I	8.8.8.0/24	198.32.161.20	216.218.252.99 (6939)	84	100	0	15169	IGP	✅
I	8.8.8.0/24	198.32.132.41	216.218.252.150 (6939)	120	100	0	15169	IGP	✅
I	8.8.8.0/24	198.32.132.41	216.218.252.254 (6939)	120	100	0	15169	IGP	✅
I	8.8.8.0/24	206.53.203.14	216.218.252.147 (6939)	165	100	0	15169	IGP	✅
I	8.8.8.0/24	208.115.136.21	216.218.252.226 (6939)	165	100	0	15169	IGP	✅
I	8.8.8.0/24	206.41.110.73	216.218.252.168 (6939)	170	100	0	15169	IGP	✅
I	8.8.8.0/24	198.179.18.72	216.218.252.28 (6939)	199	100	0	15169	IGP	✅
I	8.8.8.0/24	206.108.255.141	216.218.252.185 (6939)	245	100	0	15169	IGP	✅
I	8.8.8.0/24	198.32.242.133	216.218.252.177 (6939)	270	100	0	15169	IGP	✅
I	8.8.8.0/24	206.53.174.7	216.218.252.167 (6939)	310	100	0	15169	IGP	✅

University of Oregon Route Views

University of Oregon collects router's RIBs from globally distributed set of IXPs and routers

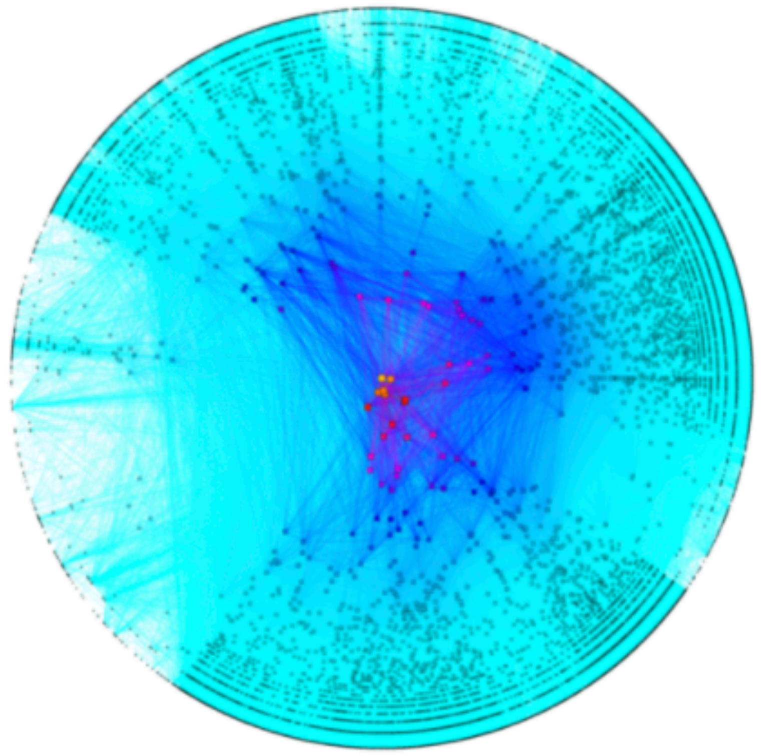
Publishes these on a regular basis at http://archive.routeviews.org/

- Data Archives
  - MRT format RIBs and UPDATEs (quagga bgpd, from route-views2.oregon-ix.net)
    MRT format RIBs and UPDATEs (quagga bgpd, from route-views3 as of Aug 13, 2013)
    MRT format RIBs and UPDATEs (quagga bgpd, from route-views4.routeviews.org)
    v6 MRT format RIBs and UPDATEs (quagga bgpd, from route-views6.oregon-ix.net)
    MRT format RIBs and UPDATEs from AMS-IX Collector (FRR bgpd, from route-views.amsix.routeviews.org)
    MRT format RIBs and UPDATEs from Chicago (FRR bgpd, from route-views.chicago.routeviews.org)
    MRT format RIBs and UPDATEs from NIC.cl Collector (FRR bgpd, from route-views.chile.routeviews.org)
    MRT format RIBs and UPDATEs from Equinix Ashburn (quagga bgpd, from route-views.eqix.routeviews.org)
    MRT format RIBs and UPDATEs from FL-IX (FRR bgpd, from route-views.flix.routeviews.org)
    MRT format RIBs and UPDATEs from GOREX (FRR bgpd, from route-views.gorex.routeviews.org)
    MRT format RIBs and UPDATEs from ISC (PAIX) (quagga bgpd, from route-views.isc.routeviews.org)
    MRT format RIBs and UPDATEs from KIXP (quagga bgpd, from route-views.kixp.routeviews.org)
    MRT format RIBs and UPDATEs from JINX (quagga bgpd, from route-views.jinx.routeviews.org)
    MRT format RIBs and UPDATEs from LINX (quagga bgpd, from route-views.linx.routeviews.org)
    MRT format RIBs and UPDATEs from NAPAfrica (FRR bgpd, from route-views.napafrica.routeviews.org)
    MRT format RIBs and UPDATEs from NWAX (quagga bgpd, from route-views.nwax.routeviews.org)

CAIDA ASRank — Inferring AS Relationships

CAIDA collects all routes from RouteViews. Attempt to infer relationships.

Read https://www.caida.org/catalog/datasets/as-relationships/ before starting Project 1 Part 3.



ASRank is CAIDA's ranking of Autonomous Systems (AS) (which approximately map to Internet Service Providers) and organizations (Orgs) (which are a collection of one or more ASes). This ranking is derived from topological data collected by CAIDA's Archipelago Measurement Infrastructure and Border Gateway Protocol (BGP) routing data collected by the Route Views Project and RIPE NCC.

ASes and Orgs are ranked by their customer cone size, which is the number of their direct and indirect customers. Note: We do *not* have data to rank ASes (ISPs) by traffic, revenue, users, or any other non-topological metric.

https://asrank.caida.org/

ASN name or number | search

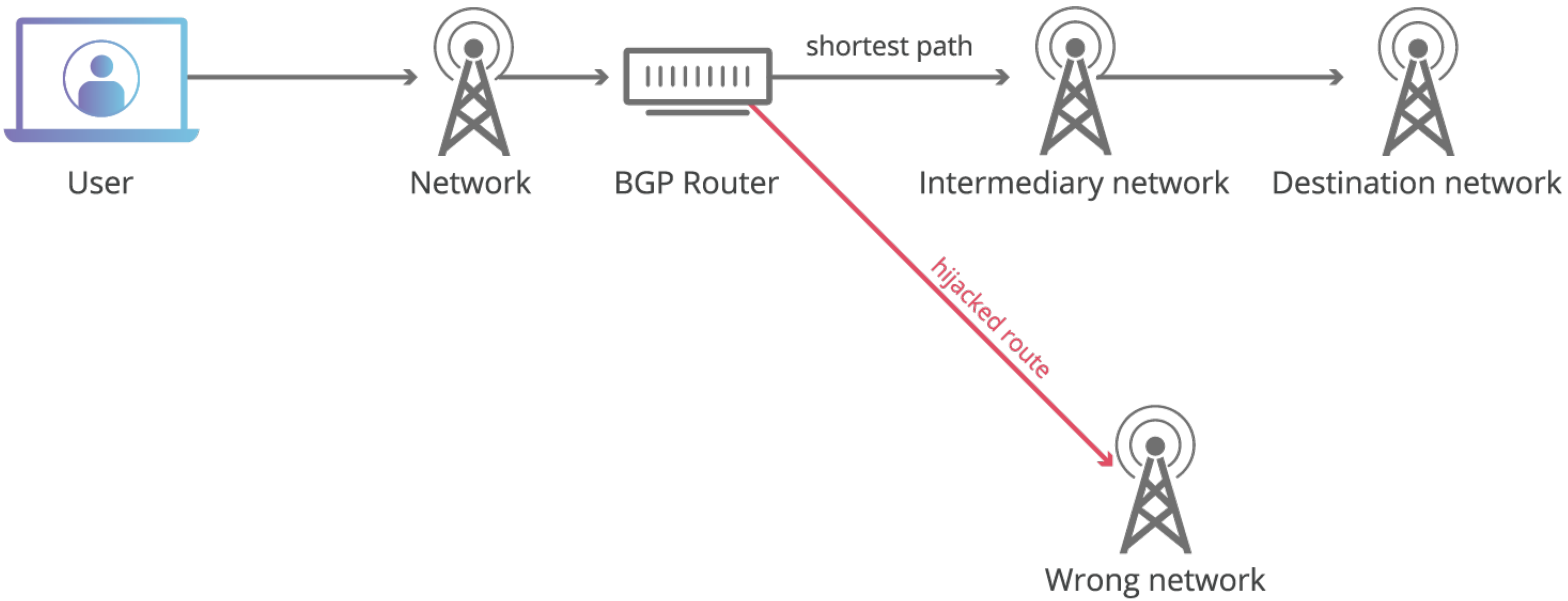
1 | 2 | 3 | 4 | .. | 1844

AS Rank ▲	AS Number ▿	Organization		cone size (ASes) ▿
1	3356	Level 3 Parent, LLC		46995
2	1299	Telia Company AB		36489

BGP Security

BGP Hijacking

BGP has no built in security! Any AS can advertise any prefix. Others will choose the shortest path — regardless of whether it's the correct path.



Real World Cases

In April 2018, a Russian provider announced IP prefixes that contained Route53 Amazon DNS servers.

They hijacked Amazon DNS queries so that DNS queries for **myetherwallet.com** went to attacker-controlled servers, which returned the wrong IP address, and directed HTTP requests to an imposter website

The hackers were thus able to steal approximately $152,000 in cryptocurrency.

🔒 *Would HTTPS have helped in this situation?*

ISP-Provided Protections

For an end customer, an ISP *should* only accept that end customer's IP address block. Any other prefix advertised from that customer should be dropped.

Easy for customers, but difficult for understanding what to filter from other ISPs

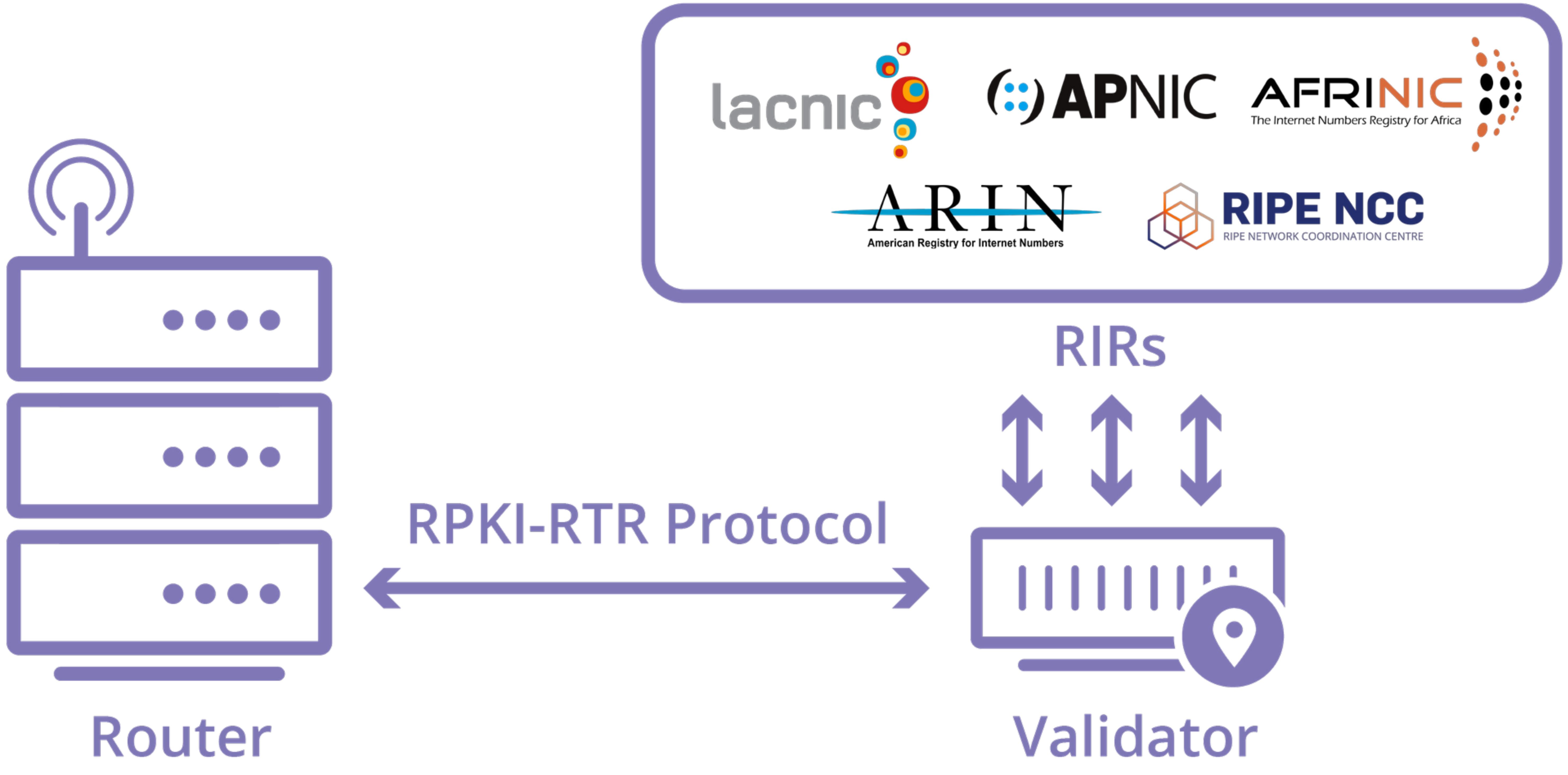
RPKI

Resource Public Key Infrastructure (RPKI)

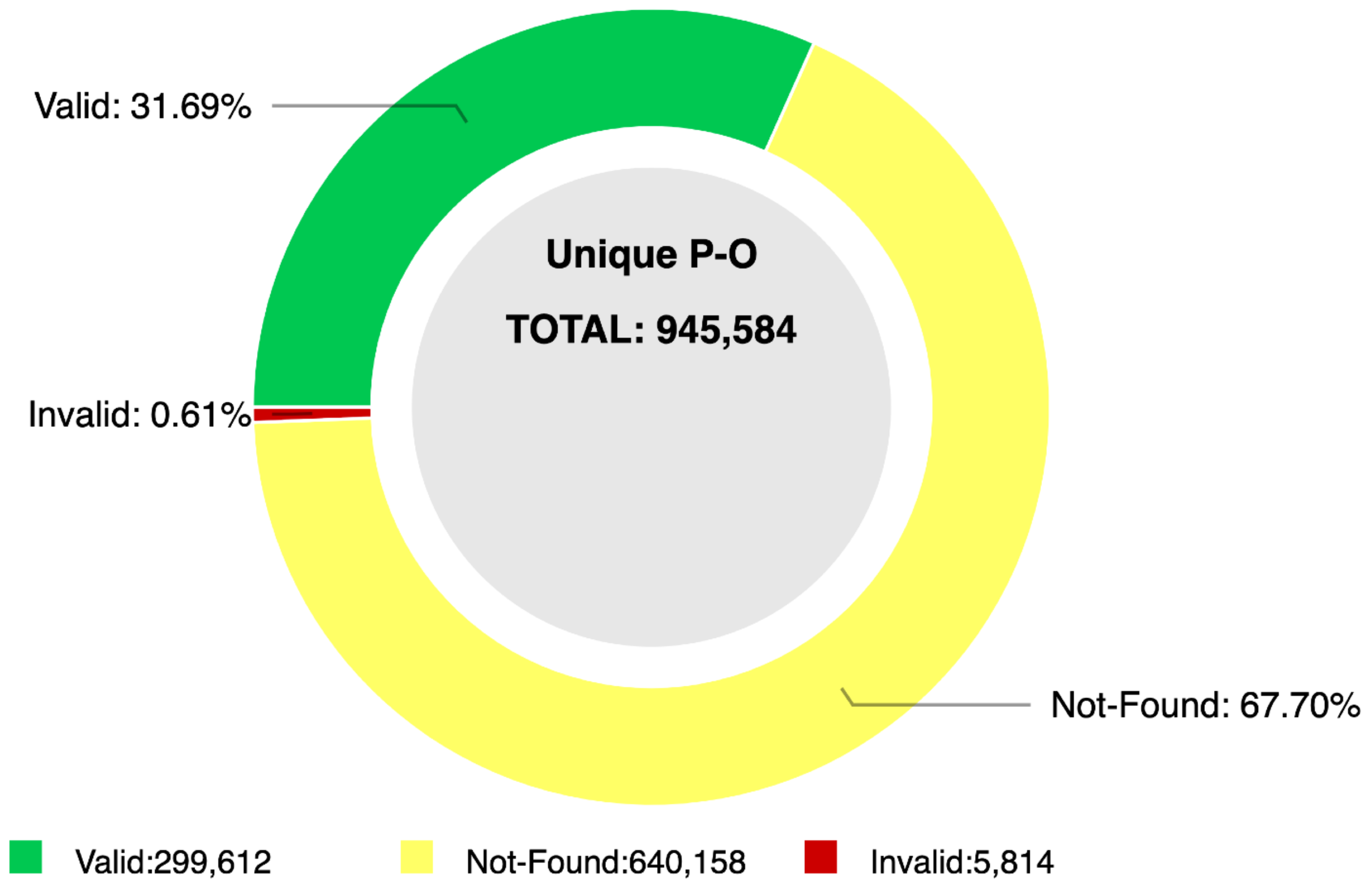
PKI that communicates who owns IP prefixes and the AS number that can originate — in an object known as a Route Origin Authorization (ROA).

RPKI uses X.509 certificates with extensions for IPs and ASNs (RFC 3779)

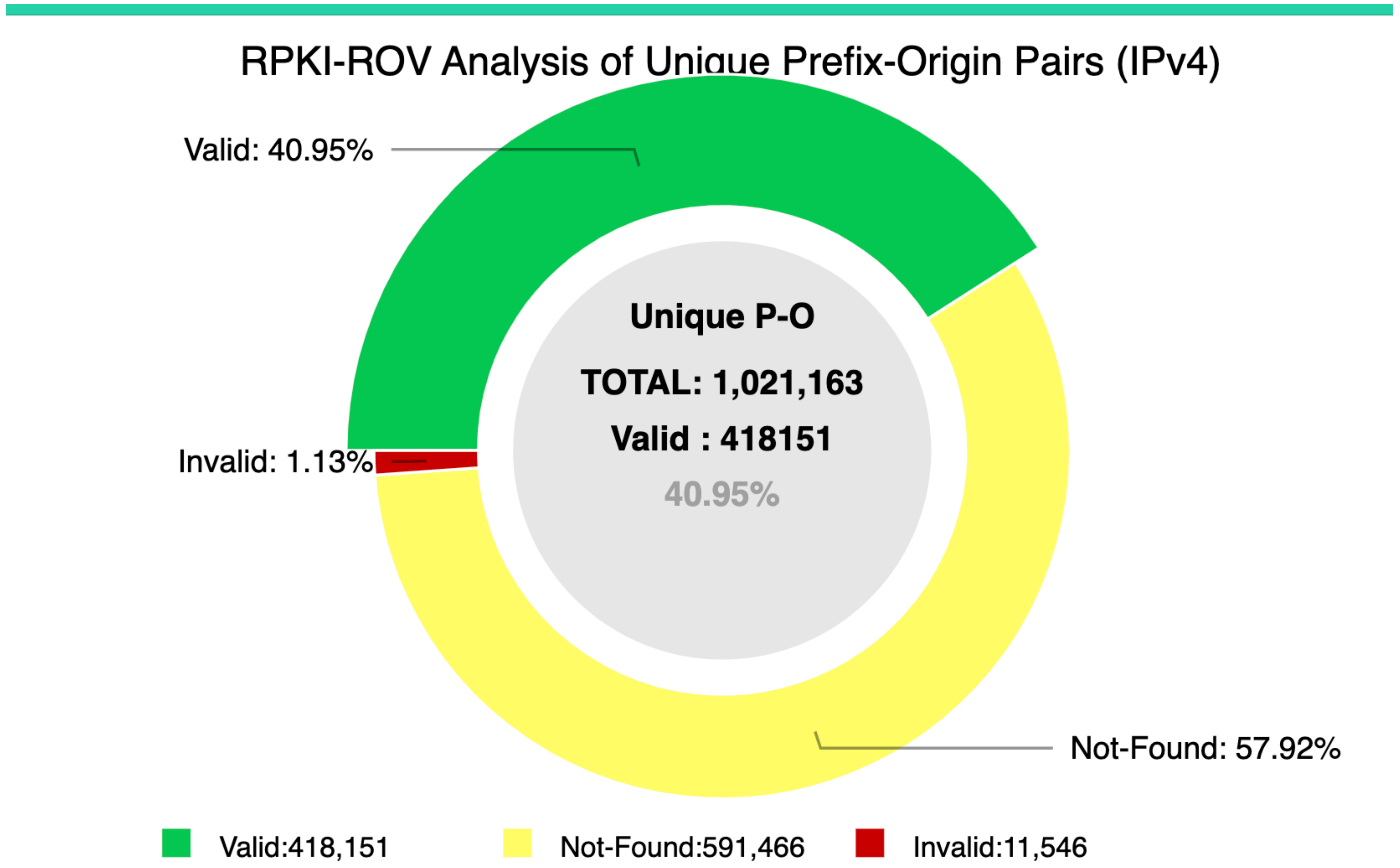
Each RIR (Internet Registry) posts their public keys — act as the trust anchors



RPKI Deployment

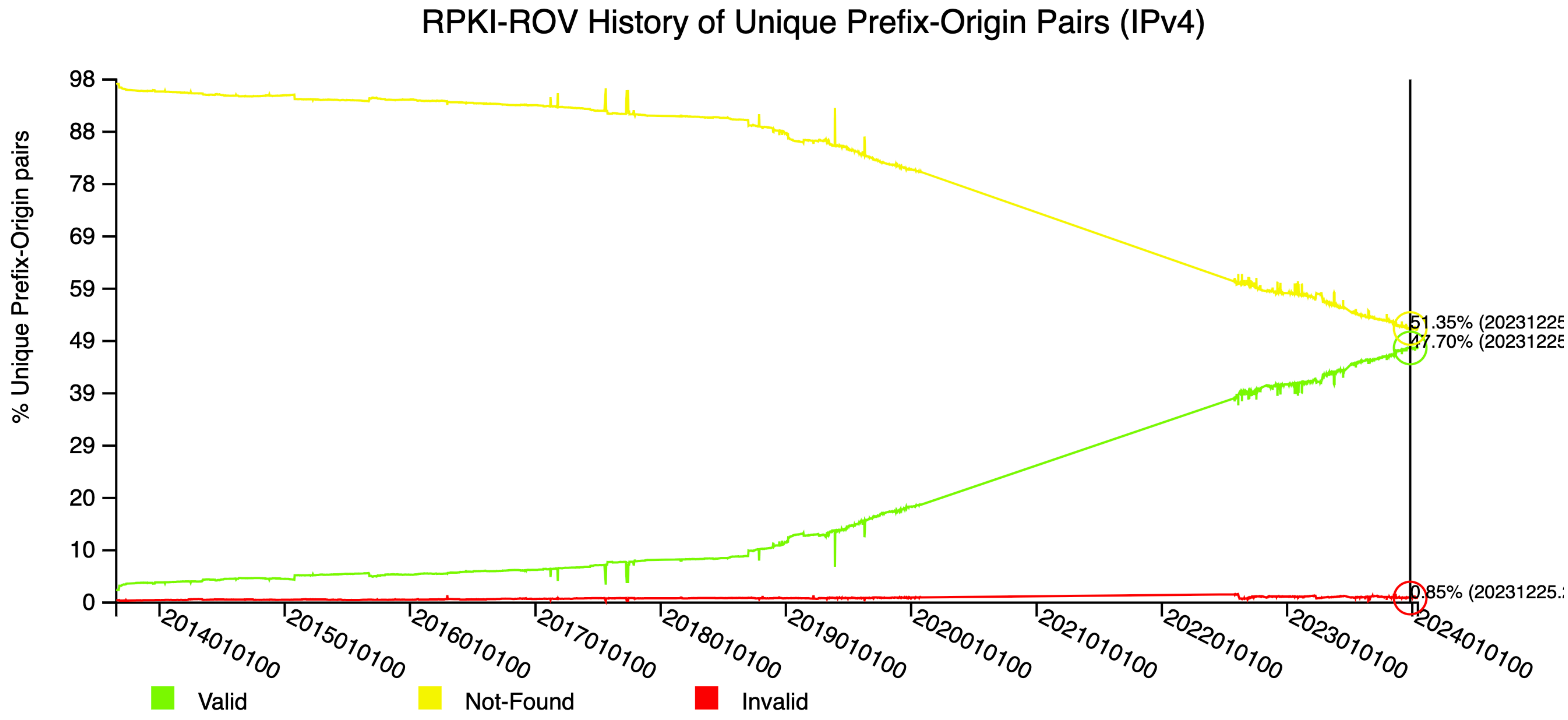


Last Year



Today

RPKI Deployment History



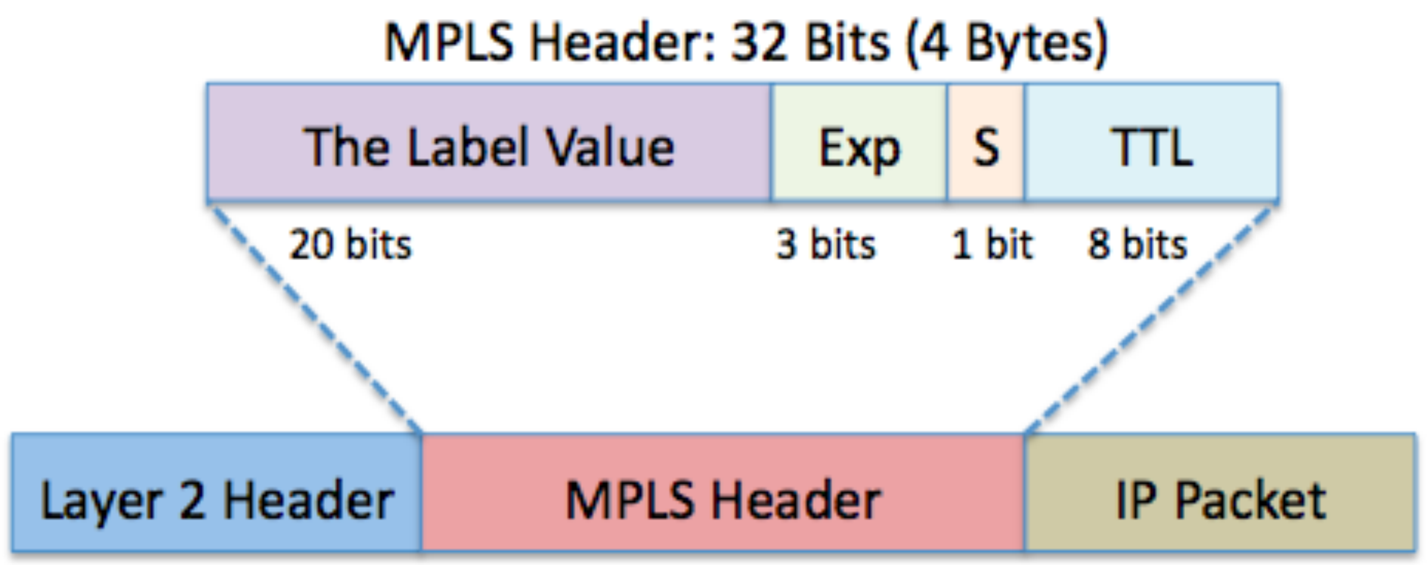
NIST RPKI Monitor: RPKI-ROV Analysis **Protocol:** IPv4 **RIR:** All

MPLS

MPLS — Multiprotocol Label Switching

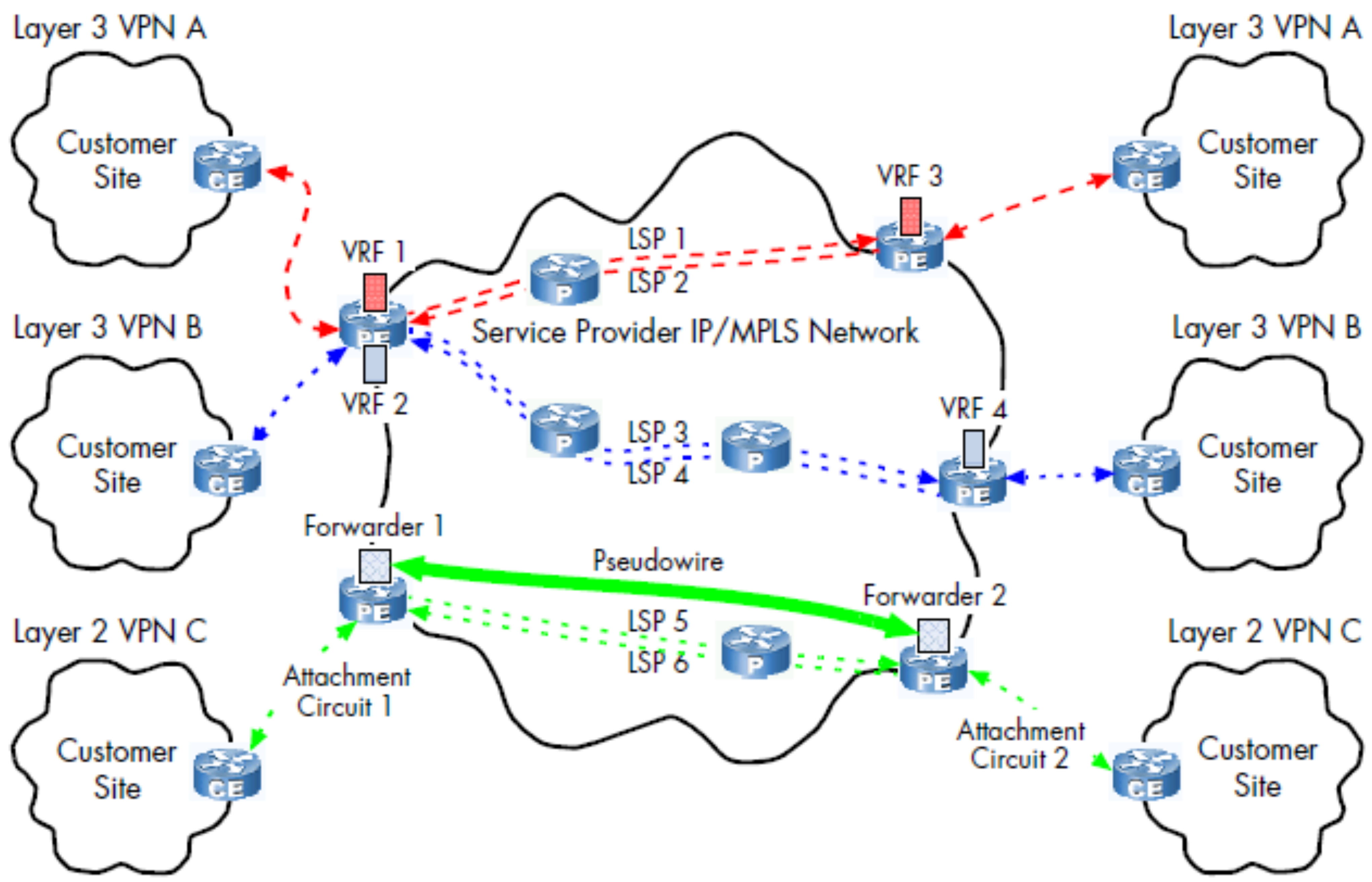
Routing technique where path through network is determined at ingress.

A short (Layer 2.5) label is tacked onto the front of the packet.



Routers use tag to *very quickly* forward to the next router. Egress strips label.

Effectively L2 Routing. Avoids expensive L3 IP longest prefix match at each hop.

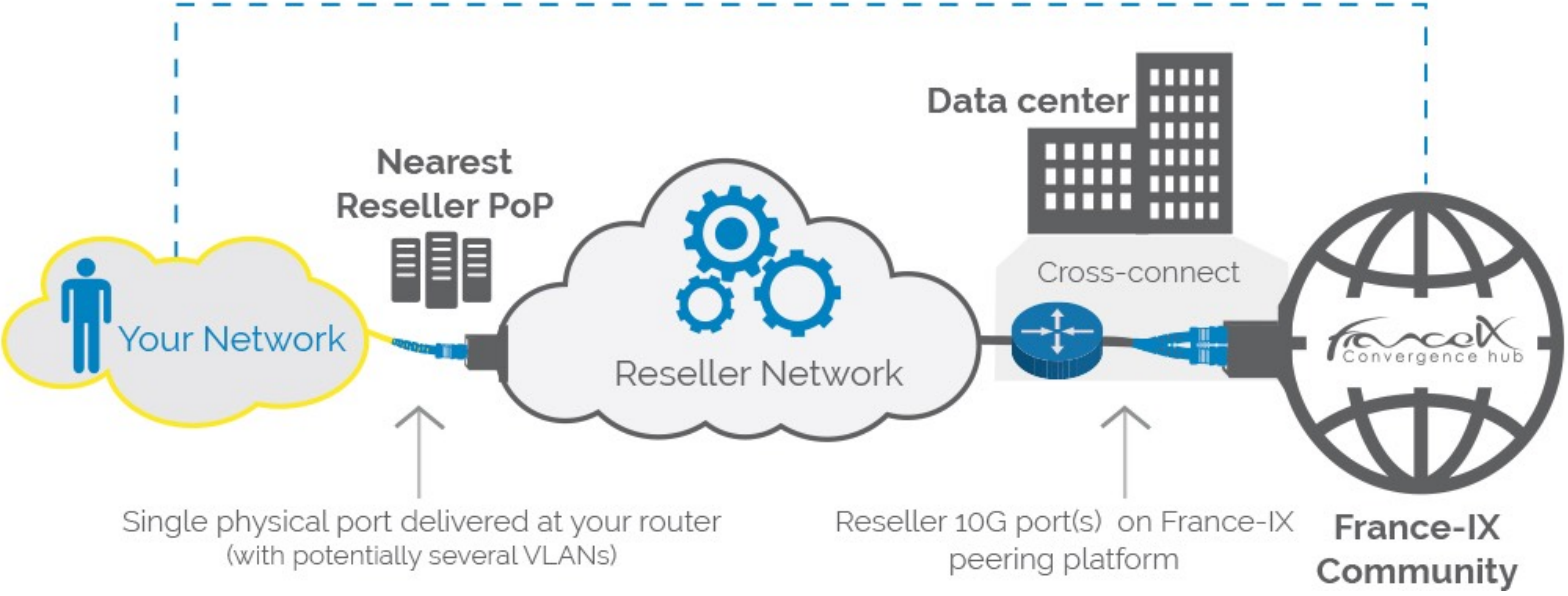


Tier 1s often use MPLS on their backbone

🚇 Remote Peering

Remote peering model

Your Peering VLAN from 100M to 2G or your dedicated nx10GE port



Remote Peering

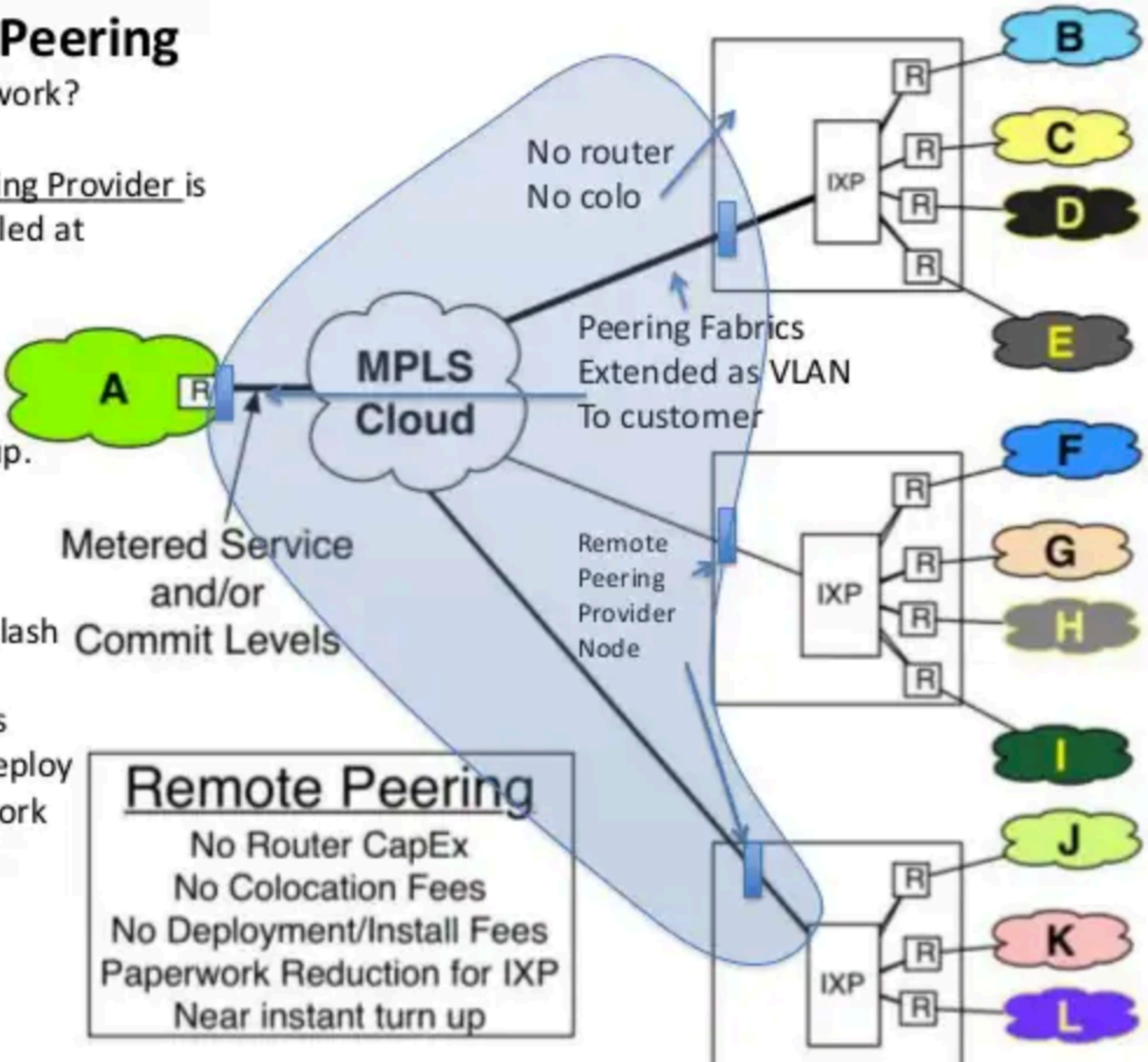
How does it work?

Remote Peering Provider is already installed at the IXPs.

Waves provisioned, instant turn up.

Neutral RPP no business clash

Peering Focus
Speeds IXP deploy
Little paperwork
One Contract



BGP Communities

BGP Communities 🏷️

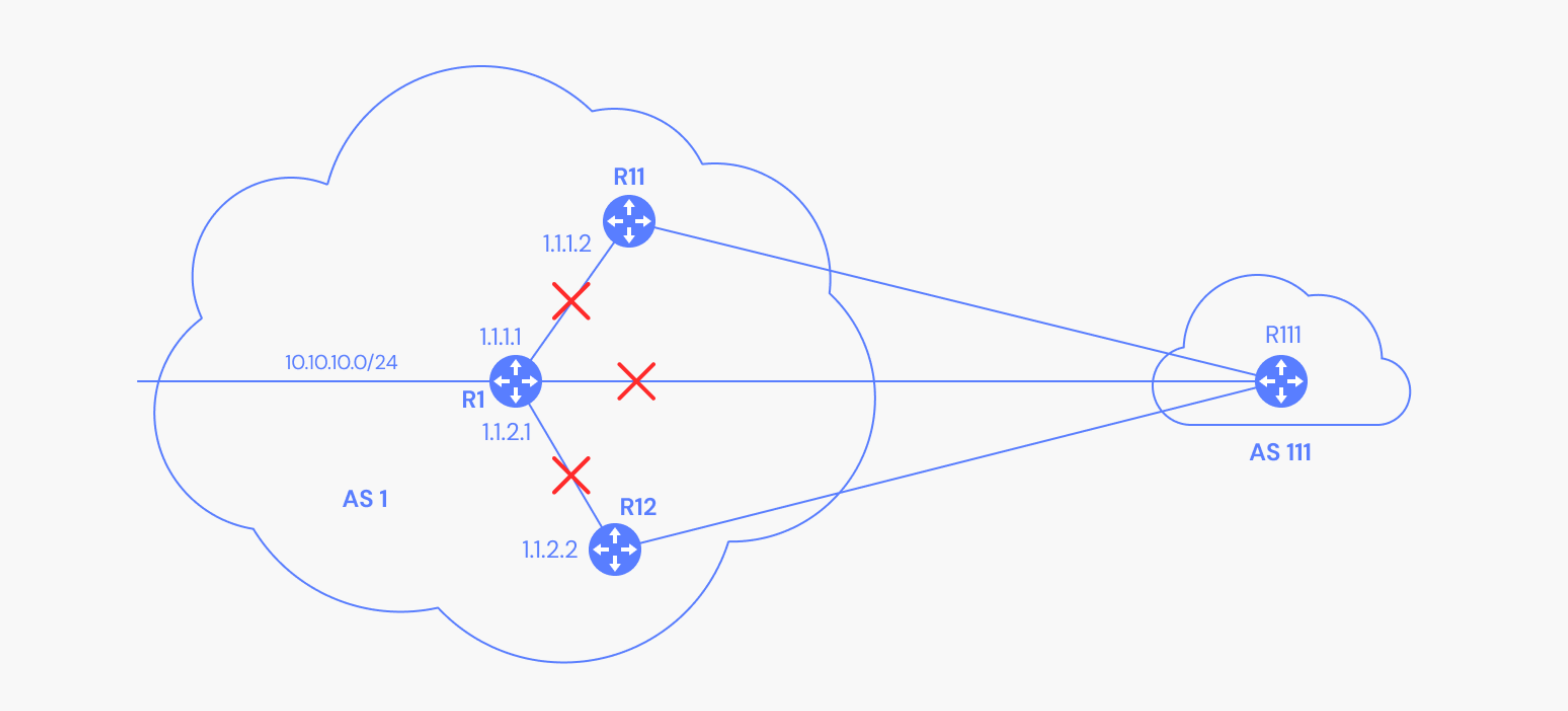
"BGP Communities" — BGP attribute that is parsed and passed to BGP peers

- Effectively tags that are attached to routes
- Communities are transitive! Passed along multiple routers.

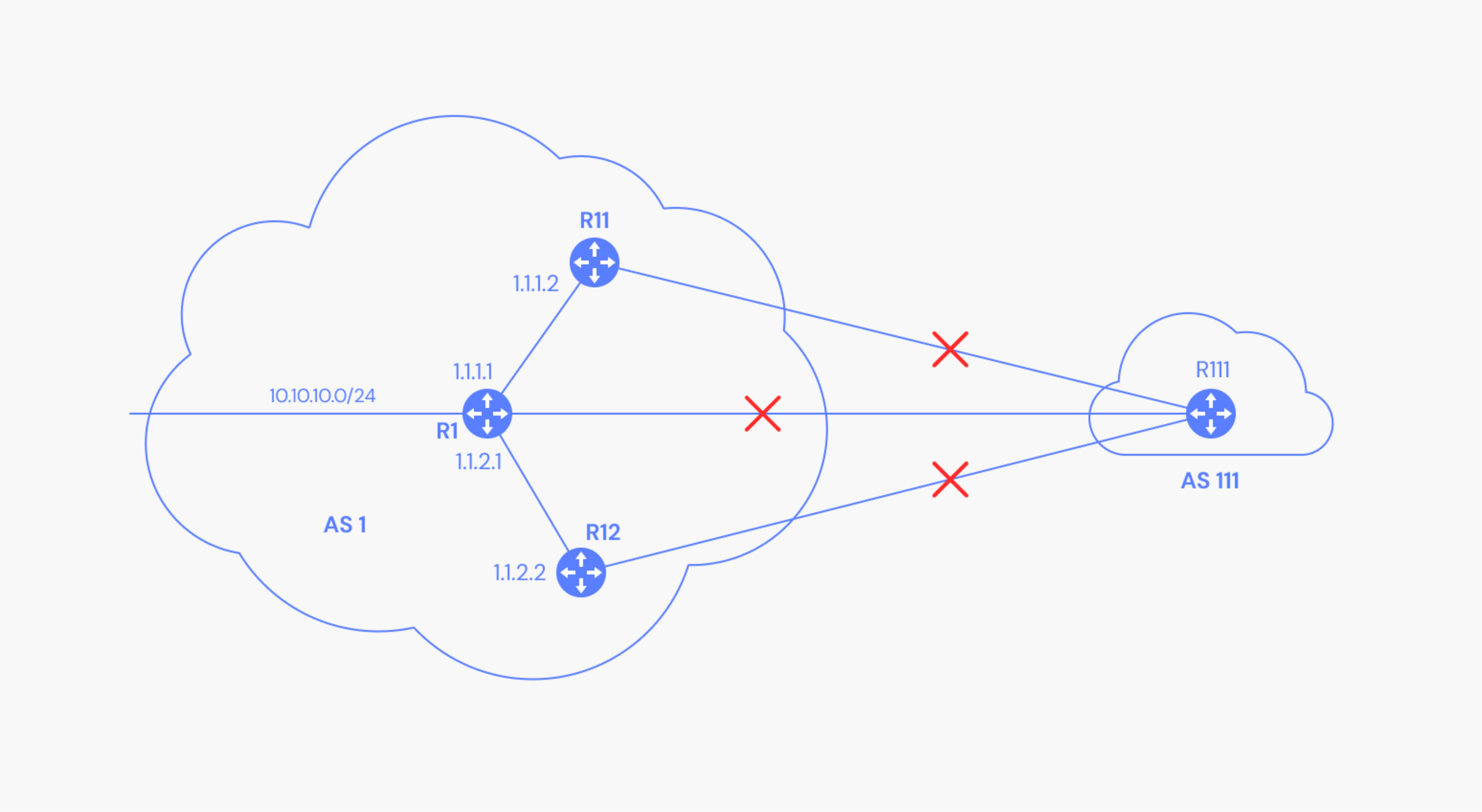
Communities allows an AS to tell its neighbors additional information about the routes it's advertising

Both standardized and non-standard communities exist

No Advertise



No Export

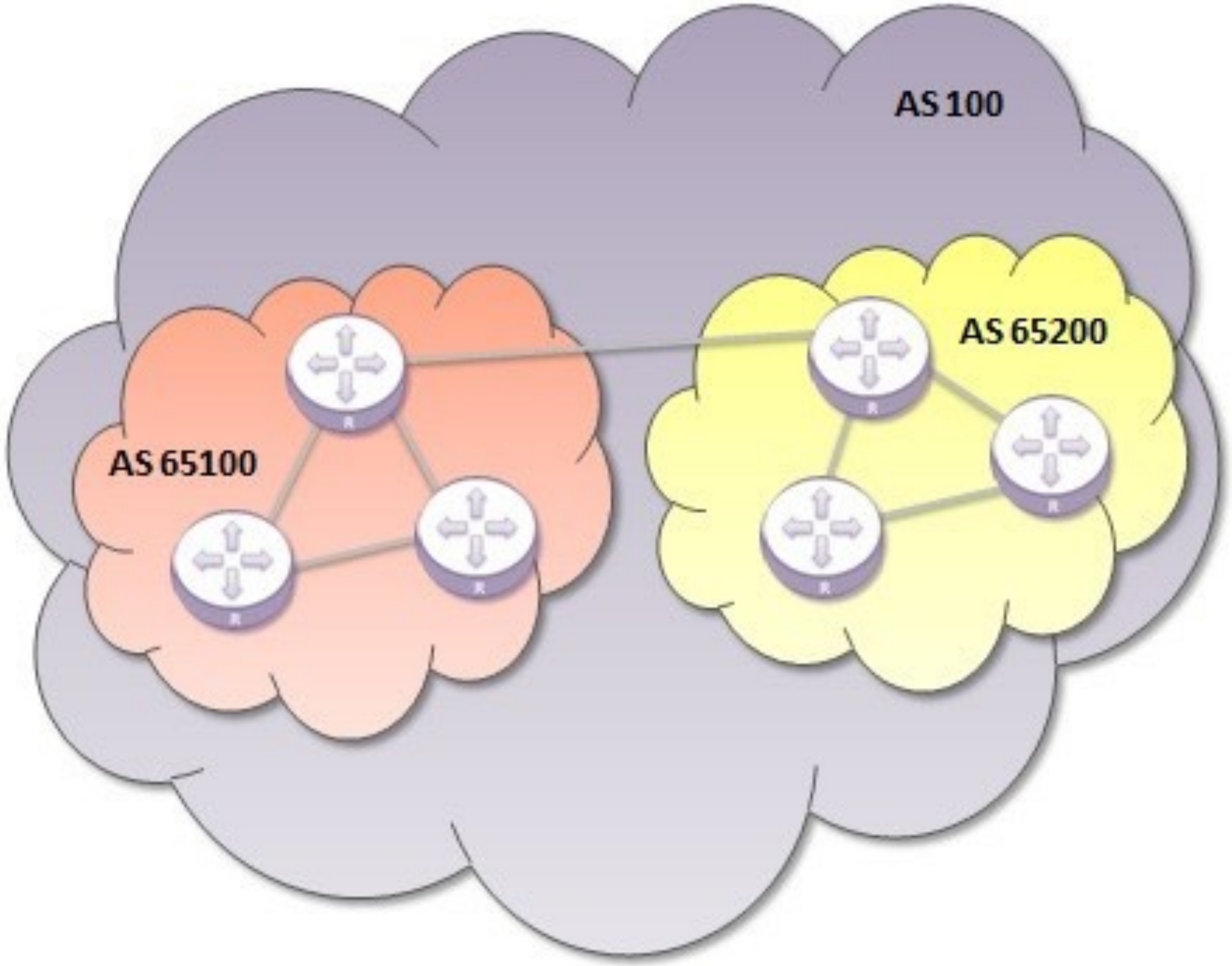


Other Standardized Communities

NO_EXPORT_SUBCONFED: Do not advertise outside of your BGP confederation

NOPEER: Other routers don't *have to* propagate the prefix

BLACKHOLE: Drop all traffic for this prefix (used to protect against DDoS)



Some NTT Communities

Customers wanting to alter their route announcements to selected peers

NTT BGP customers may choose to prepend to selected peers with the following communities, where nnn is the peer's ASN:

Community	Description
65400:nnn	do not advertise to peer nnn in North America
65401:nnn	prepends o/b to peer nnn 1x in North America
65402:nnn	prepends o/b to peer nnn 2x in North America
65403:nnn	prepends o/b to peer nnn 3x in North America
65410:nnn	announce to peer nnn in North America, disregards 2914:429 and 65500:nnn
65420:nnn	do not advertise to peer nnn in Europe
65421:nnn	prepends o/b to peer nnn 1x in Europe