

Modern Routing Practices

CS249i: The Modern Internet



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-v112.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-sv1-agg8--sfo-agg4-100gbe.cenic.net (137.164.11.92)  2.532 ms
7  dc-sv1-agg10--sv1-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/>

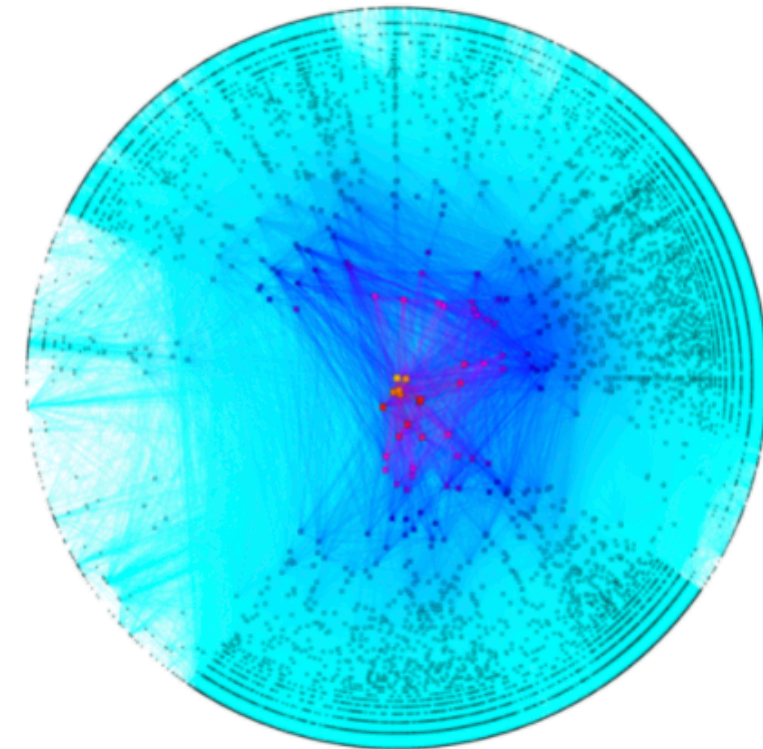
- Data Archives

- [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.napafrika.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/>

1 2 3 4 .. 1844

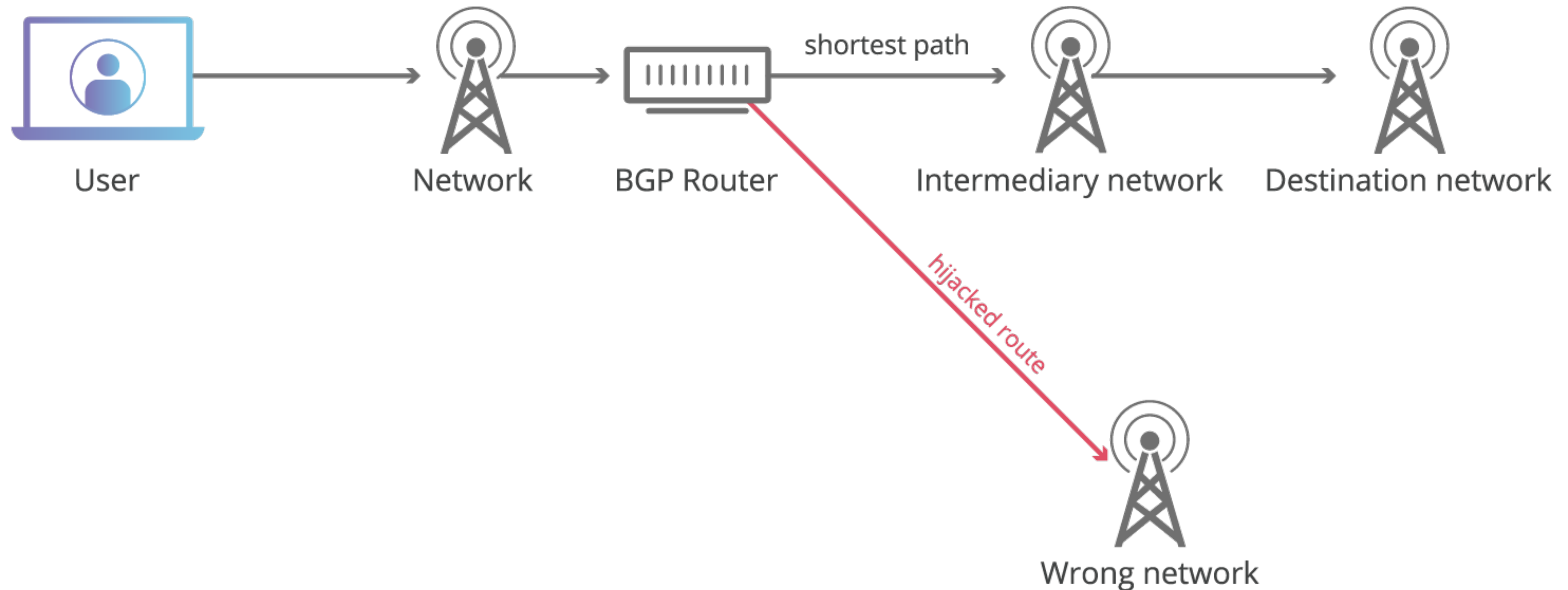
| AS Rank ▲ | AS Number ▼ | Organization | | cone size (ASes) ▼ |
|-----------|-------------|---------------------|--|--------------------|
| 1 | 3356 | Level 3 Parent, LLC | | 46995 |
| 2 | 1299 | Telia Company AB | | 36489 |



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



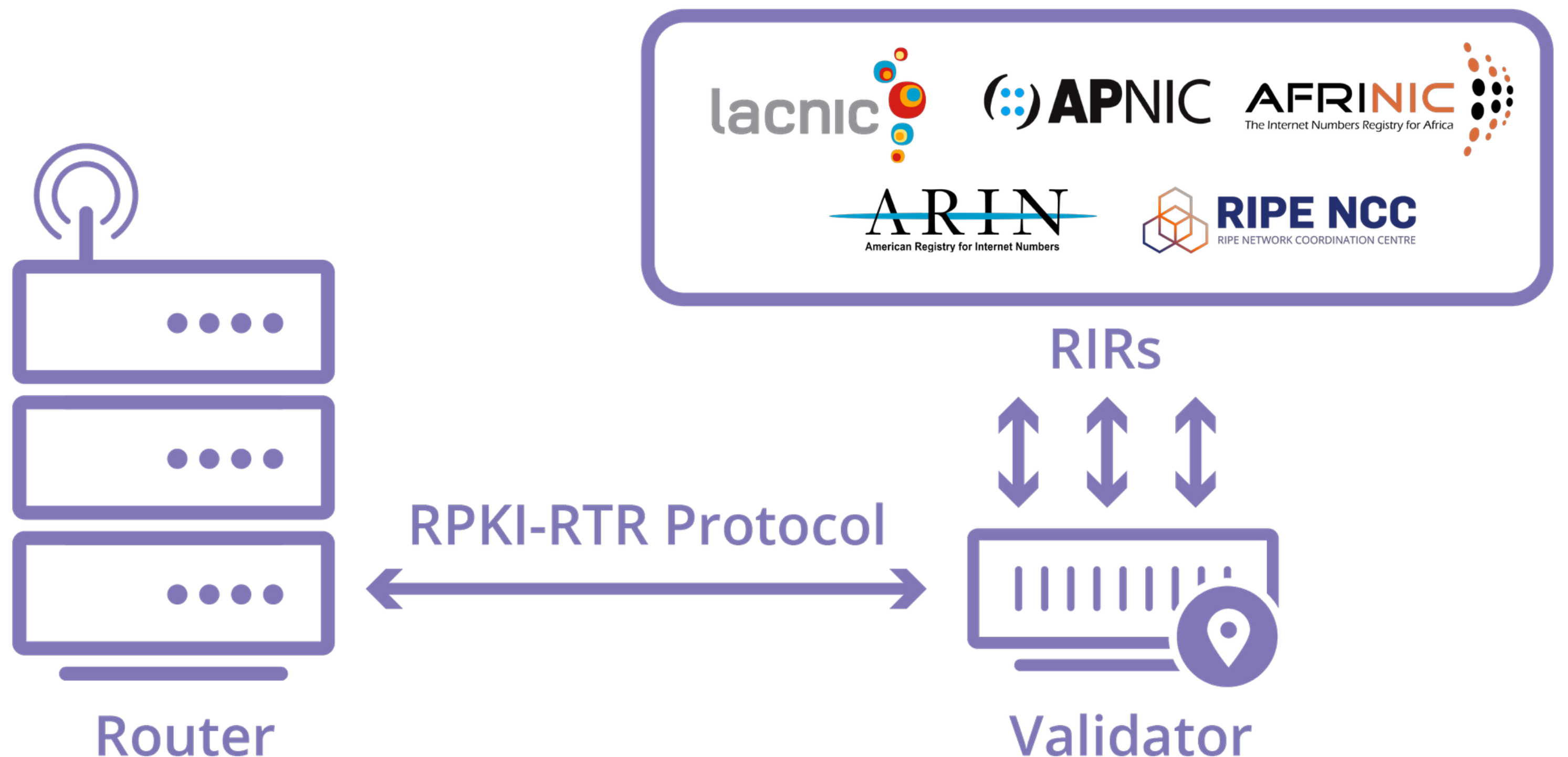
RPKI

Resource Public Key Infrastructure (RPKI)

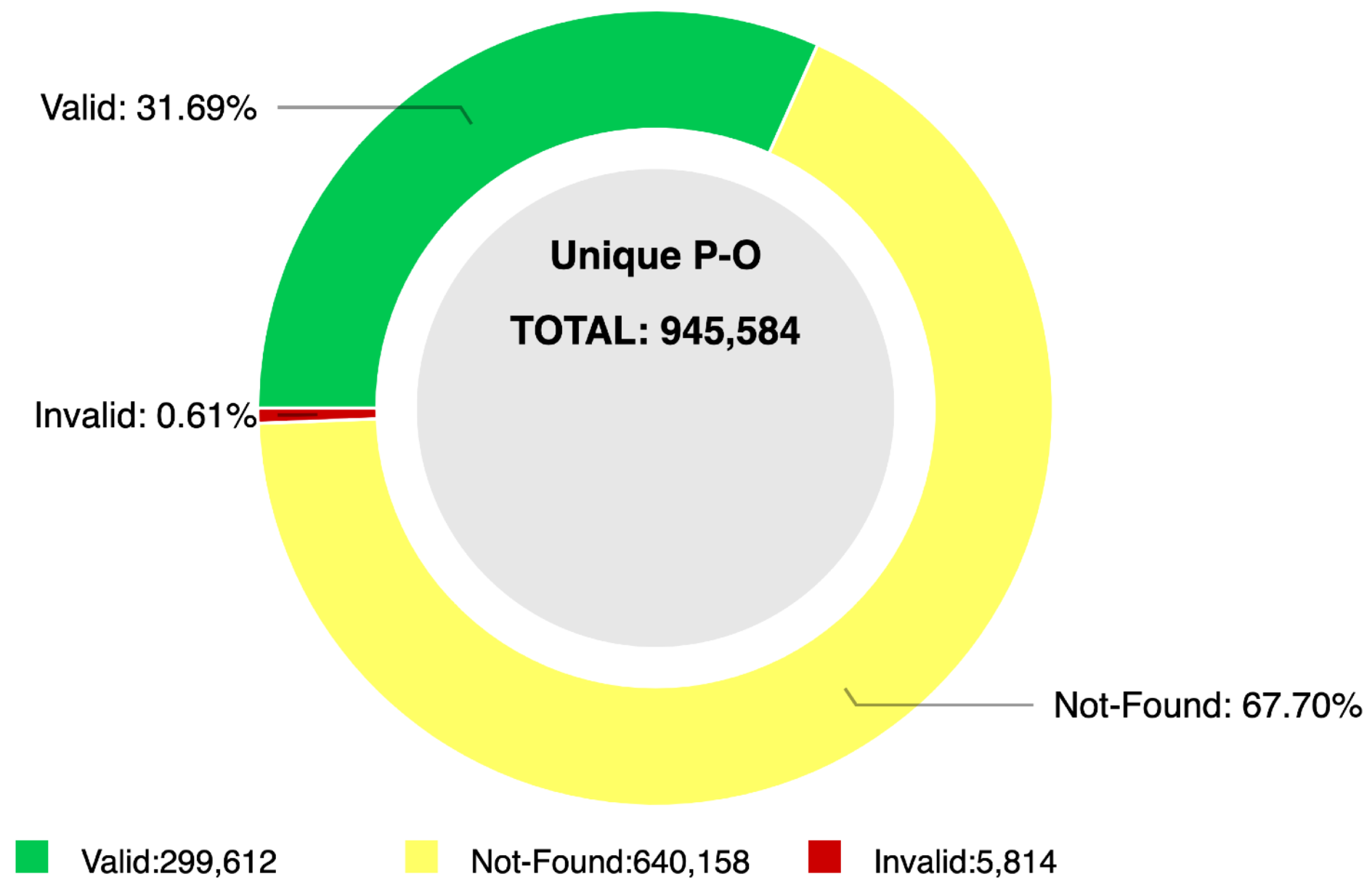
PKI that communicates who owns IP prefixes and the AS number that can originate — in an object known as a Route Origin Authorization (ROA).

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

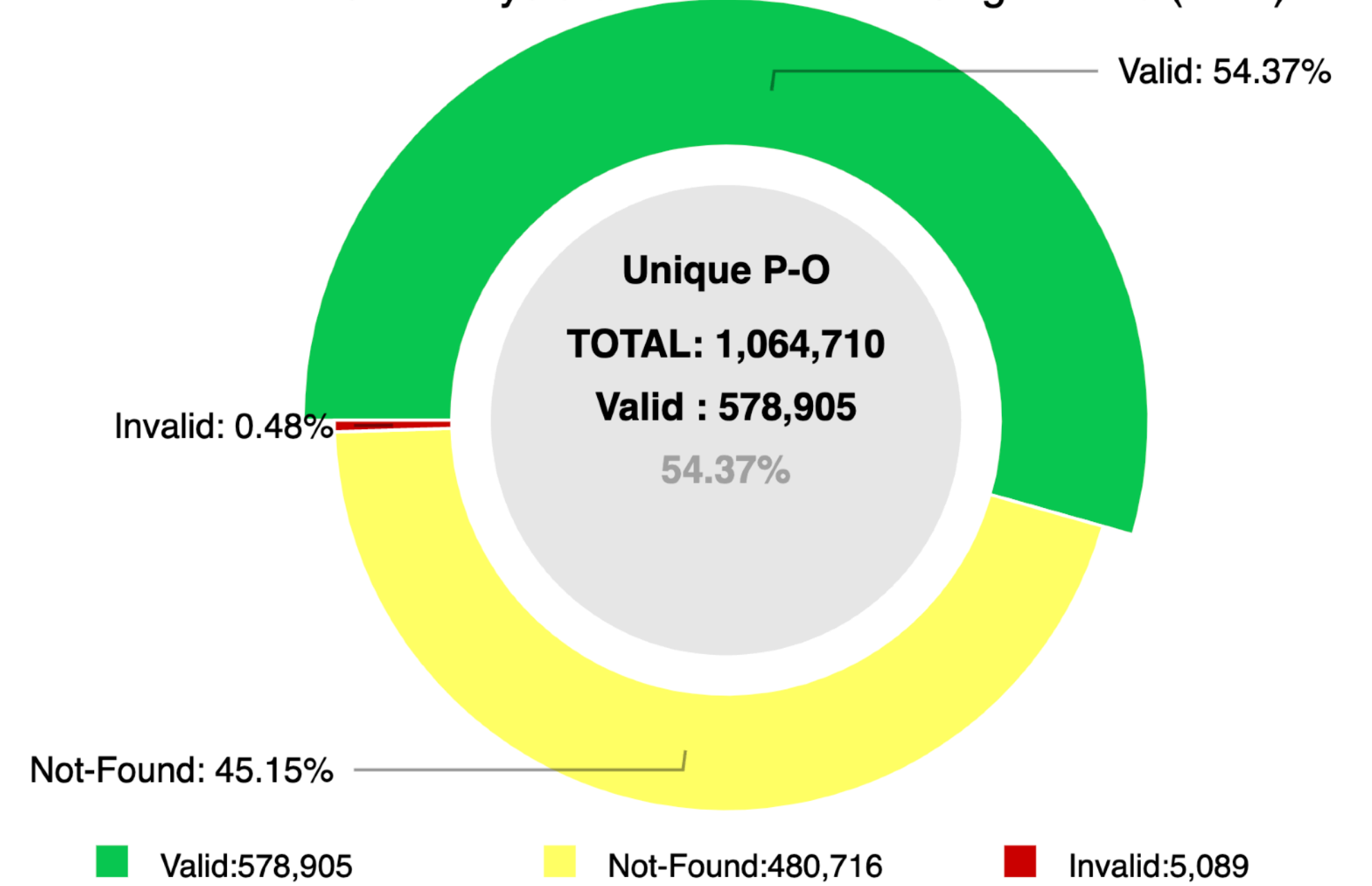


RPKI Deployment



2023

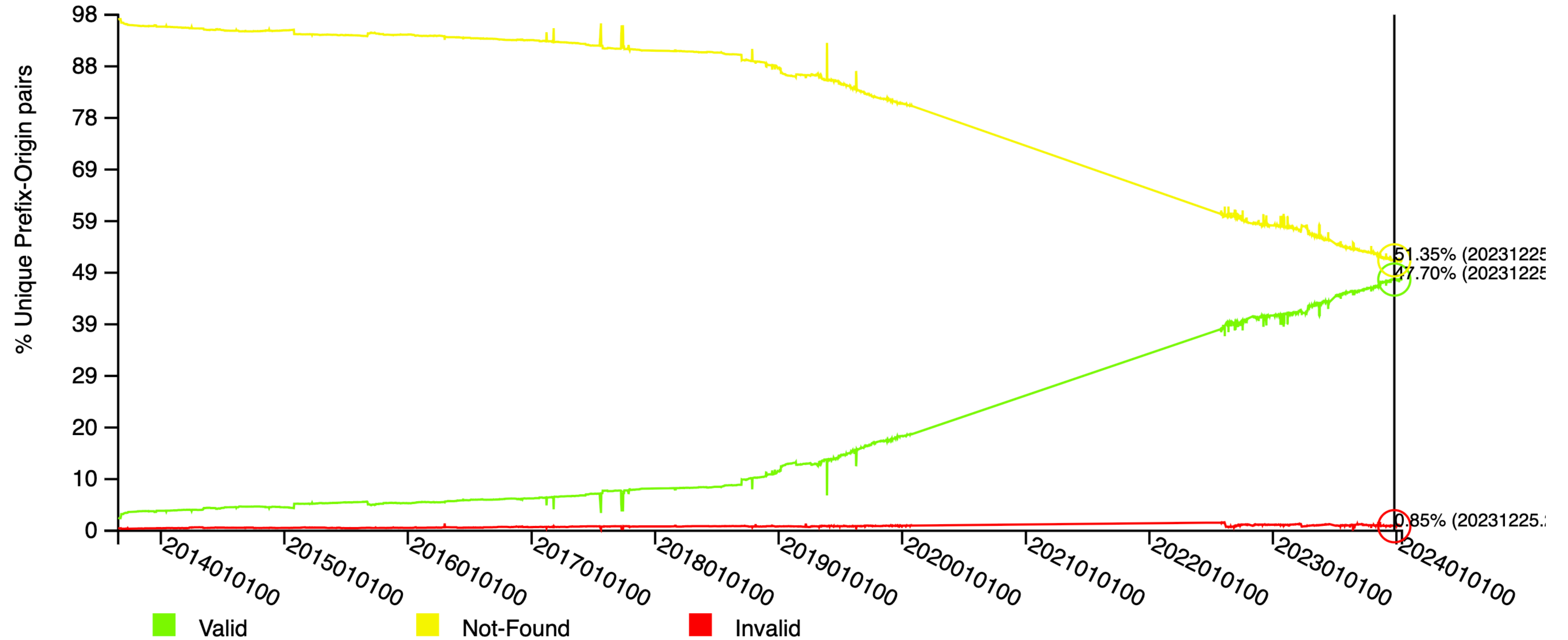
RPKI-ROV Analysis of Unique Prefix-Origin Pairs (IPv4)



Today

RPKI Deployment History

RPKI-ROV History of Unique Prefix-Origin Pairs (IPv4)



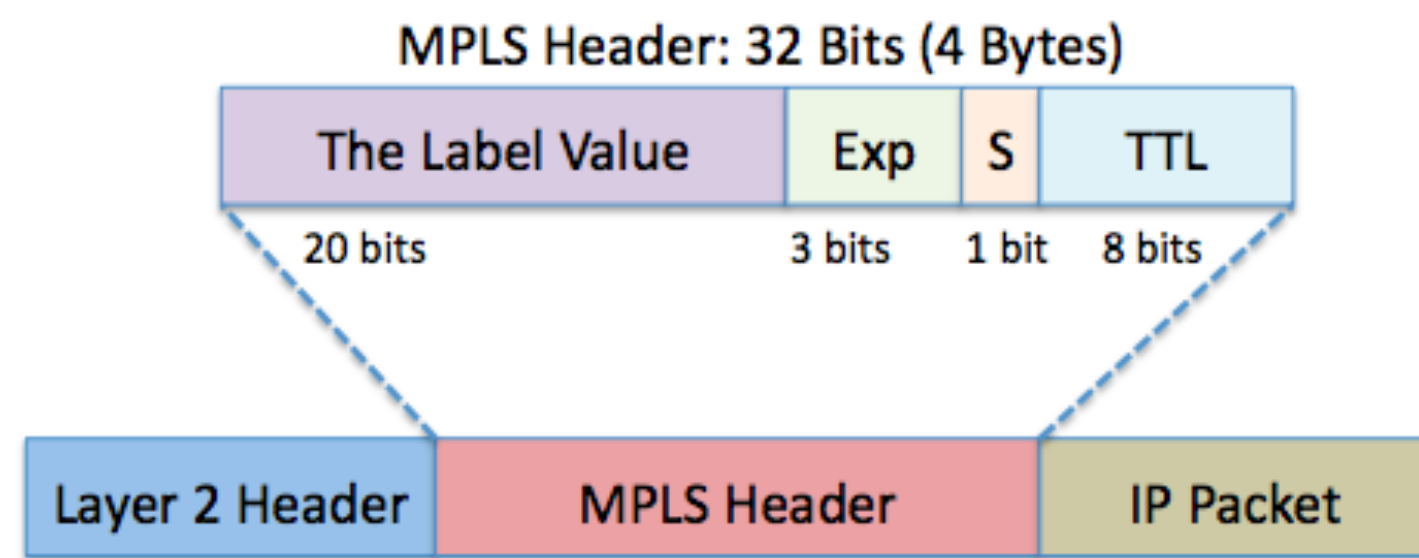


MPLS

MPLS — Multiprotocol Label Switching

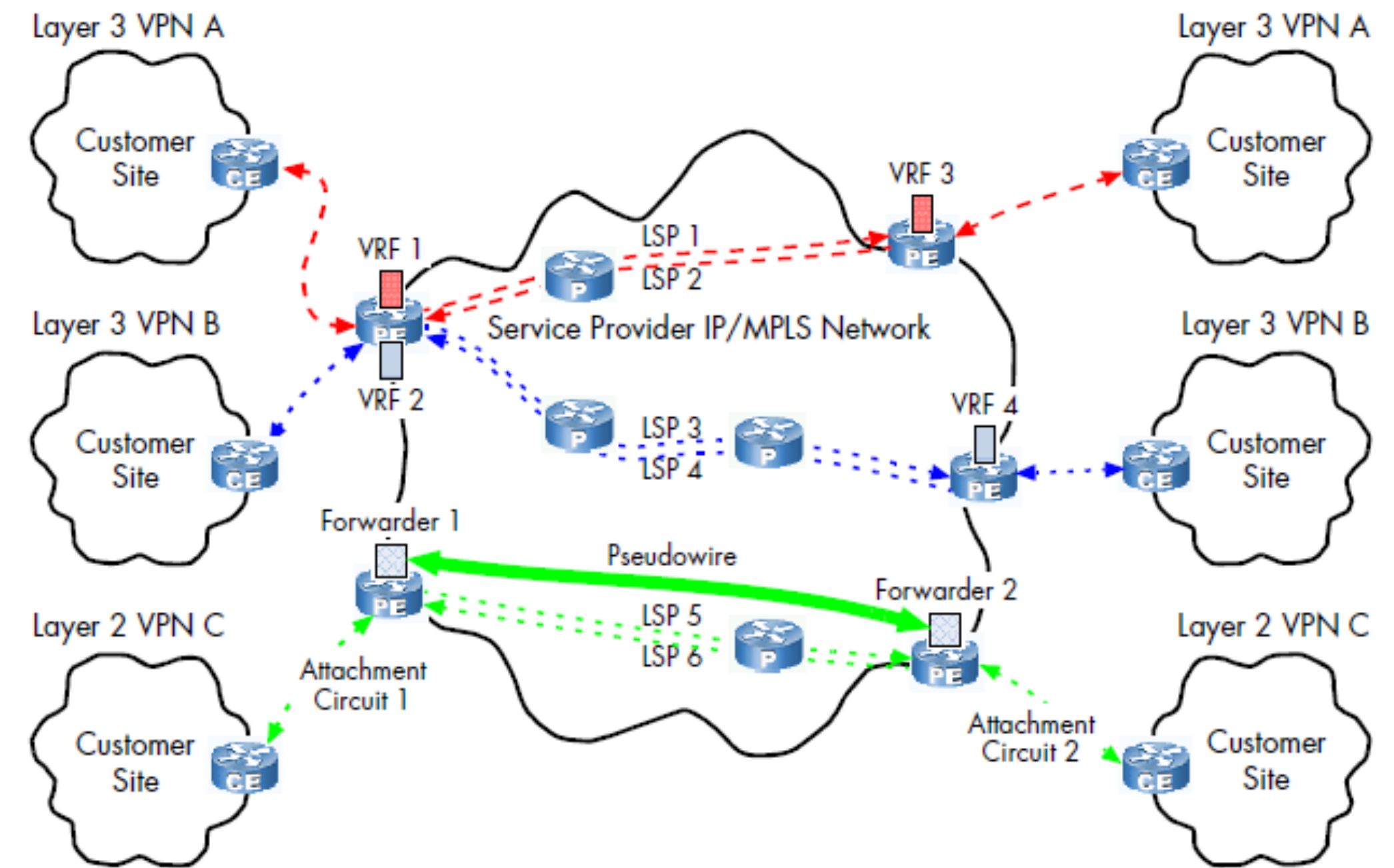
Routing technique where path through network is determined at ingress.

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.



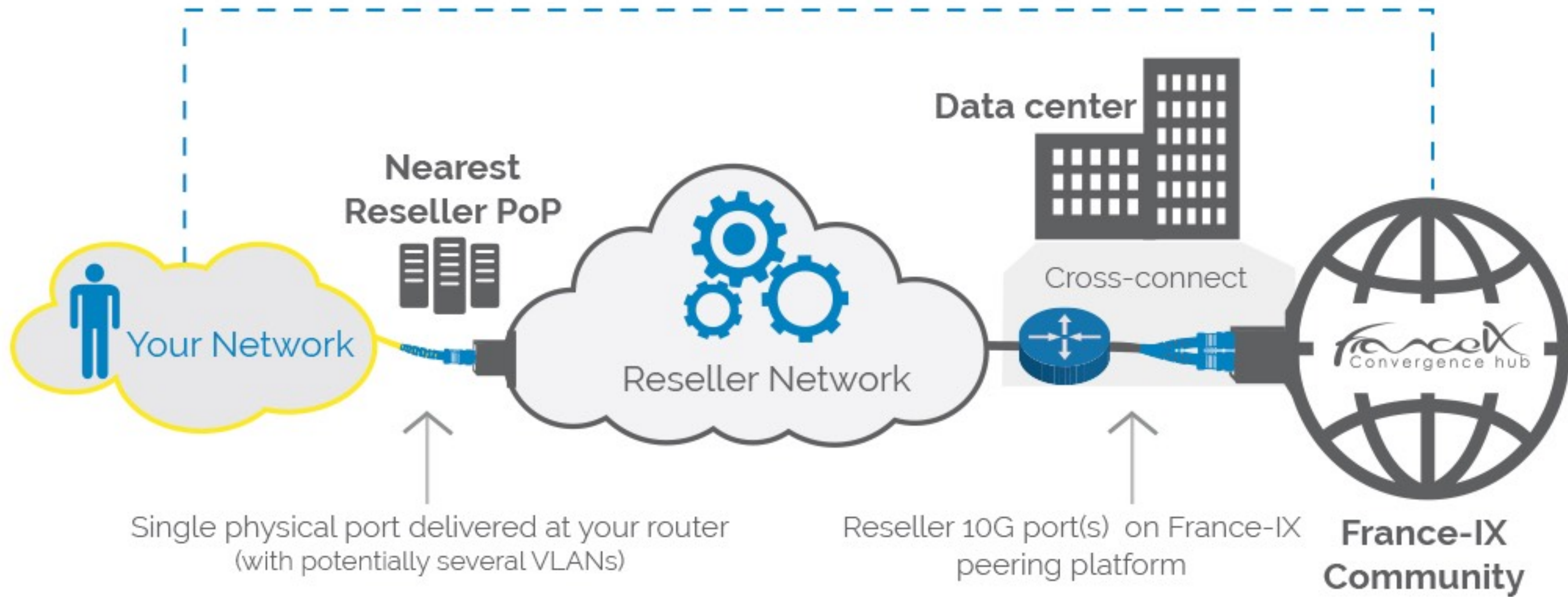
Tier 1s often use MPLS on their backbone



Remote Peering

Remote peering model

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



Remote Peering

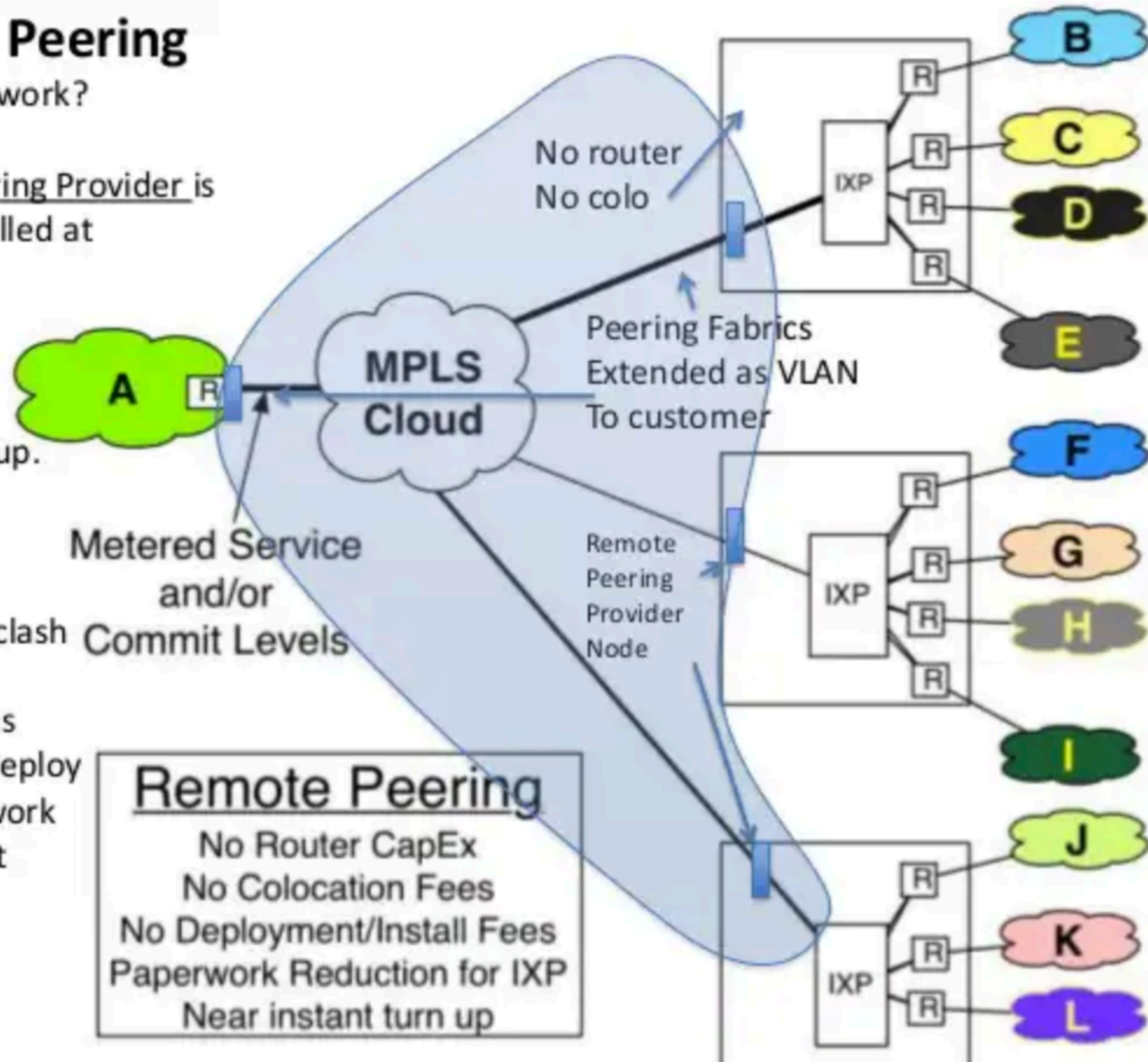
How does it work?

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



Remote Peering
No Router CapEx
No Colocation Fees
No Deployment/Install Fees
Paperwork Reduction for IXP
Near instant turn up



BGP Communities

BGP Communities

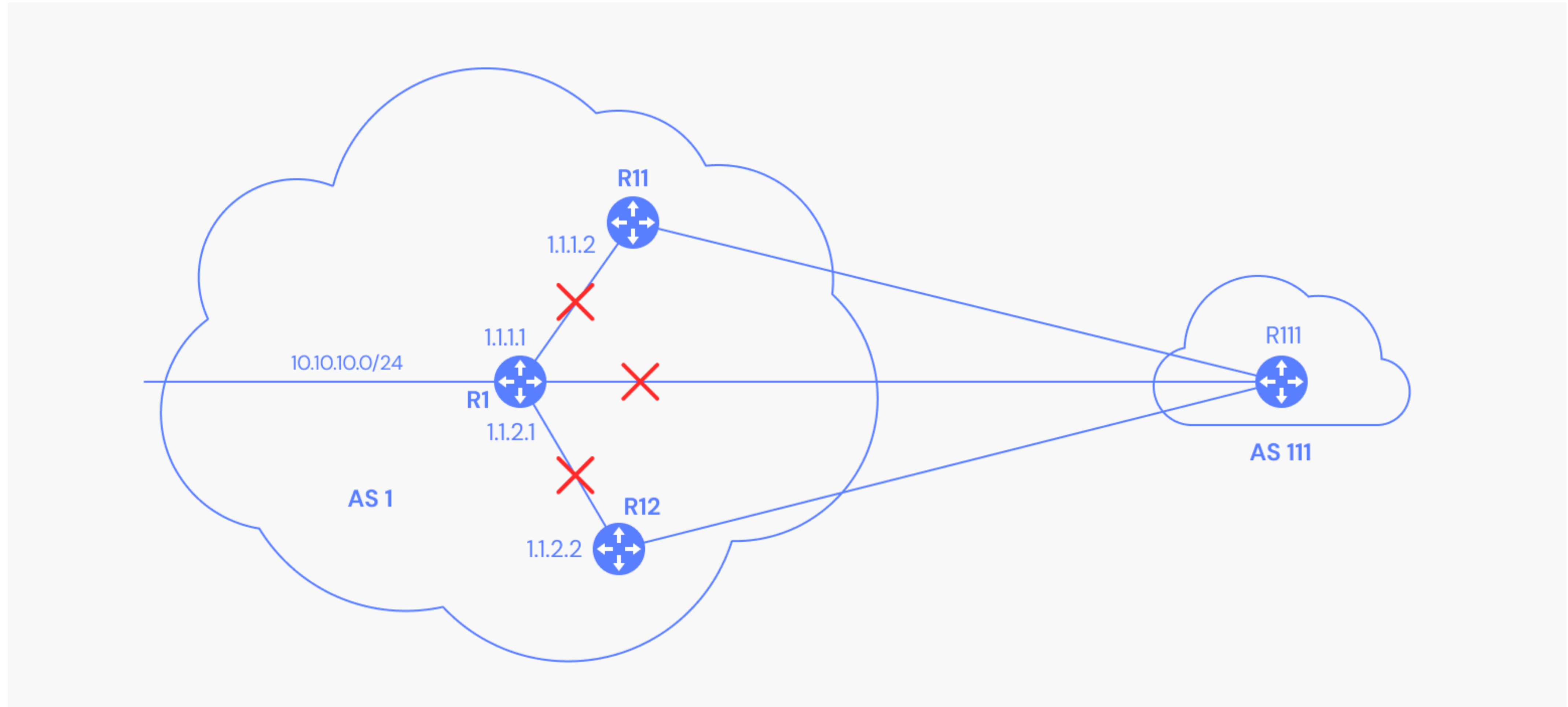
"BGP Communities" — BGP attribute that is parsed and passed to BGP peers

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

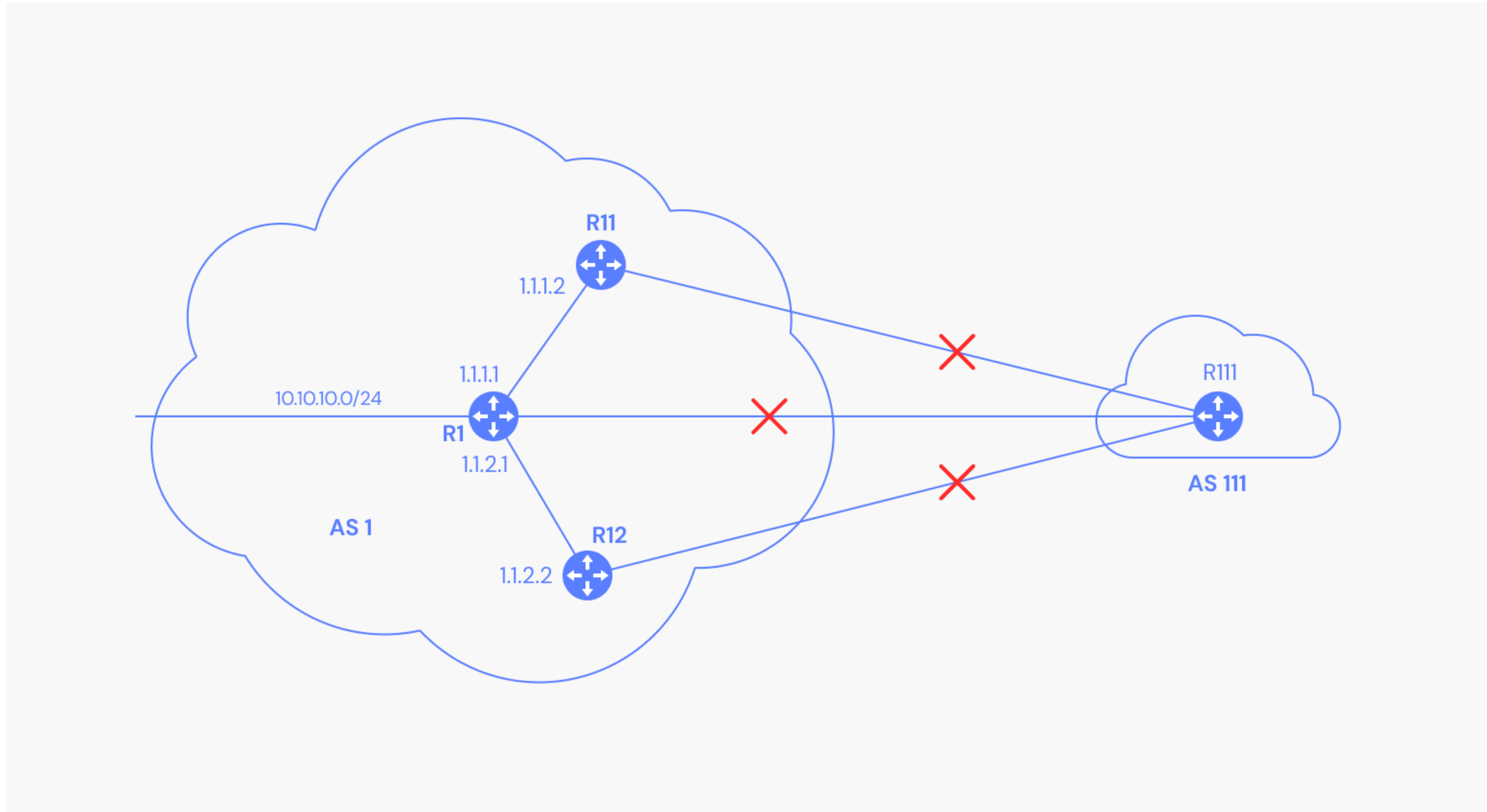
Communities allows an AS to tell its neighbors additional information about the routes it's advertising

Both standardized and non-standard communities exist

No Advertise



No Export

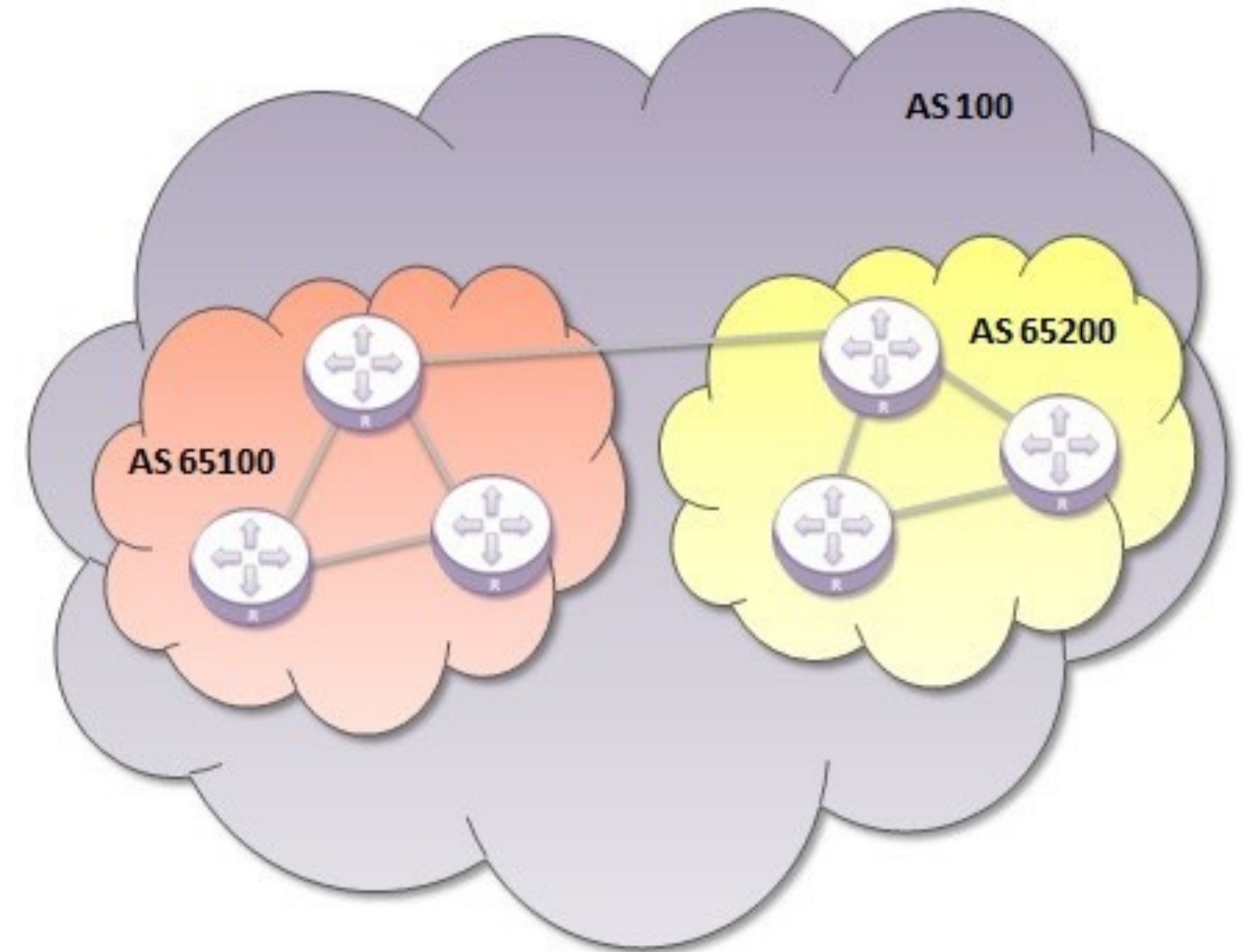


Other Standardized Communities

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 |

IPv4 → IPv6

| | IPv4 | IPv6 |
|------------------------|-------------------------|--------------------|
| Address Size | 32-bit | 128-bit |
| Header Size | 20 bytes | 40 bytes |
| Header Fields | 12 fields | 8 fields |
| Checksum | IP + TCP, Sometimes UDP | TCP + UDP |
| Flow Labeling | — | Flow ID |
| Fragmentation | Host + Router | Host Only |
| Host Addressing | DHCP, ARP, IRDP | SLAC, ICMP, DHCPv6 |
| Broadcast | Yes! | No! |