INEX

IXP Route Servers with RPKI

UKNOF44, Belfast.

September 10th, 2019.

Barry O’Donovan

@ComePeerWithMe / @barryo79

https://www.inex.ie/

IX Route Servers

- An IXP is (usually) a shared broadcast domain
  (think of it as a *big switch)*
- IXP participants arrange bilateral BGP peering sessions to exchange routes and thus traffic.
- BGP sessions required if everyone peers with everyone:

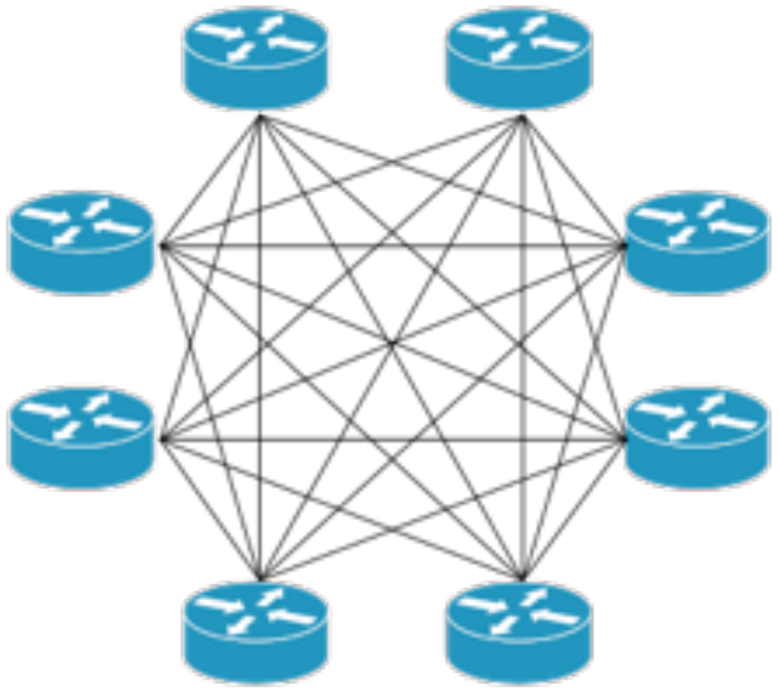
$$\frac{n(n-1)}{2}$$

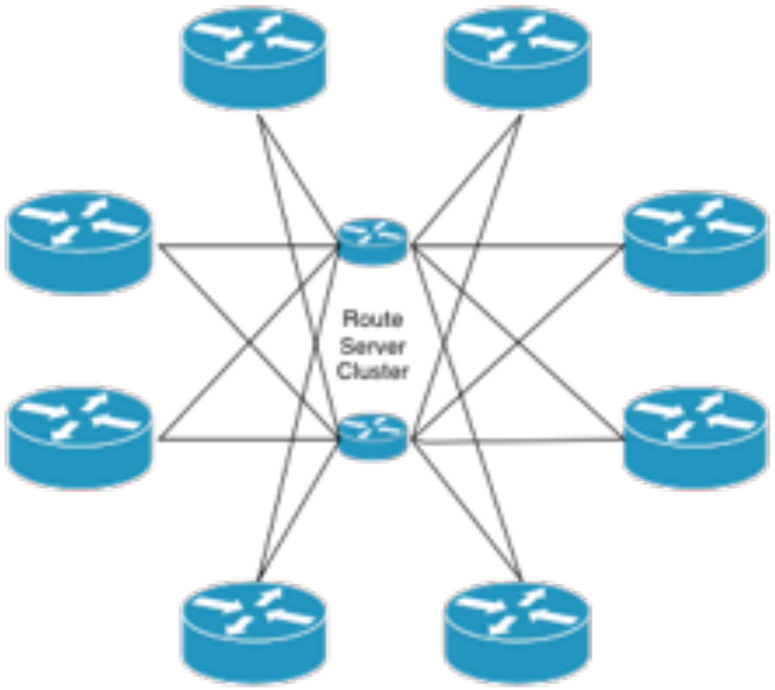
  - 10 participants: 45 sessions
  - 100 participants: 4,950 sessions

More info: RFC's 7947 and 7948.

IX Route Servers

Peering on IXP without Route Servers





Peering on IXP with Route Servers

IRRDB vs. RPKI ROAs

```
route6:         2001:db8::/32
descr:          Example IPv6 route object
origin:         AS65500
created:        2006-07-12T16:11:58Z
last-modified:  2011-02-22T15:58:03Z
source:         SOME-IRRDB
```

```
route:          192.0.2.0/24
descr:          Example IPv4 route object
origin:         AS65500
created:        2004-12-06T11:43:57Z
last-modified:  2016-11-16T22:19:51Z
source:         SOME-IRRDB
```

1. BGP filtering automation tool: https://github.com/snar/bgpq3

RPKI ROAs - Route Origin Authorisations

- A cryptographically secure replacement for route[6] objects
- Adds maximum prefix length
- Yields route origin triplets that have been validated

```
( Origin AS, Prefix       ,  Max Length )
( AS65500,   2001:db8::/32,  /48        )
( AS65501,   192.0.2.0/24,   /24        )
```

Valid ROAs on INEX LAN2

```
bird> show route
  filter {
    if bgp_large_community ~ [( 2128, 1000, 1 )] then accept;
  }
  table master4 count
```

```
5868 of 21920 routes for 16944 networks in table master4
```

=> 35% of IPv4 routes on INEX LAN2 have a valid ROA

```
902 of 2868 routes for 1943 networks in table master6
```

=> 46% of IPv6 routes on INEX LAN2 have a ROA

Invalid ROAs on INEX LAN2

```
bird> show route
  filter {
    if bgp_large_community ~ [(2128, 1101, 13)] then accept;
  }
  table master4 count
```

```
106 of 21918 routes for 16942 networks in table master4
```

=> 0.6% of IPv4 routes on INEX LAN2 have a valid ROA

```
12 of 2866 routes for 1941 networks in table master6
```

=> 0.6% of IPv6 routes on INEX LAN2 have a ROA

IXP Manager

- An INEX project
- Full-stack management system for IXPs
- FOSS - GPL v2 license
- Complete route server automation
- In use at >70 IXPs worldwide

https://www.ixpmanager.org/

github.com/inex/IXP-Manager

Route Servers with RPKI

Route Server Refresh at INEX & IXP Manager

- RPKI just one element
- Upgrade configuration from Bird v1.6 to Bird v2.0
- Complete rewrite of filtering workflow
  - Large communities used extensively within the route server
- Upgrade Bird's Eye[1] for Bird v2 BGP
- Overhaul IXP Manager looking glass

[1] A secure micro service for querying Bird - https://github.com/inex/birdseye

Bird v1 to v2 Changes

- RPKI-RTR supported
- Collapsed separate daemons for IPv4 and IPv6 into a single daemon
    - master route table becomes master4 / master6
    - new protocol blocks: ipv4 { ... }   /   ipv6 { ... }
- Other very minor configuration changes

IXP Manager v5 Route Server Filtering

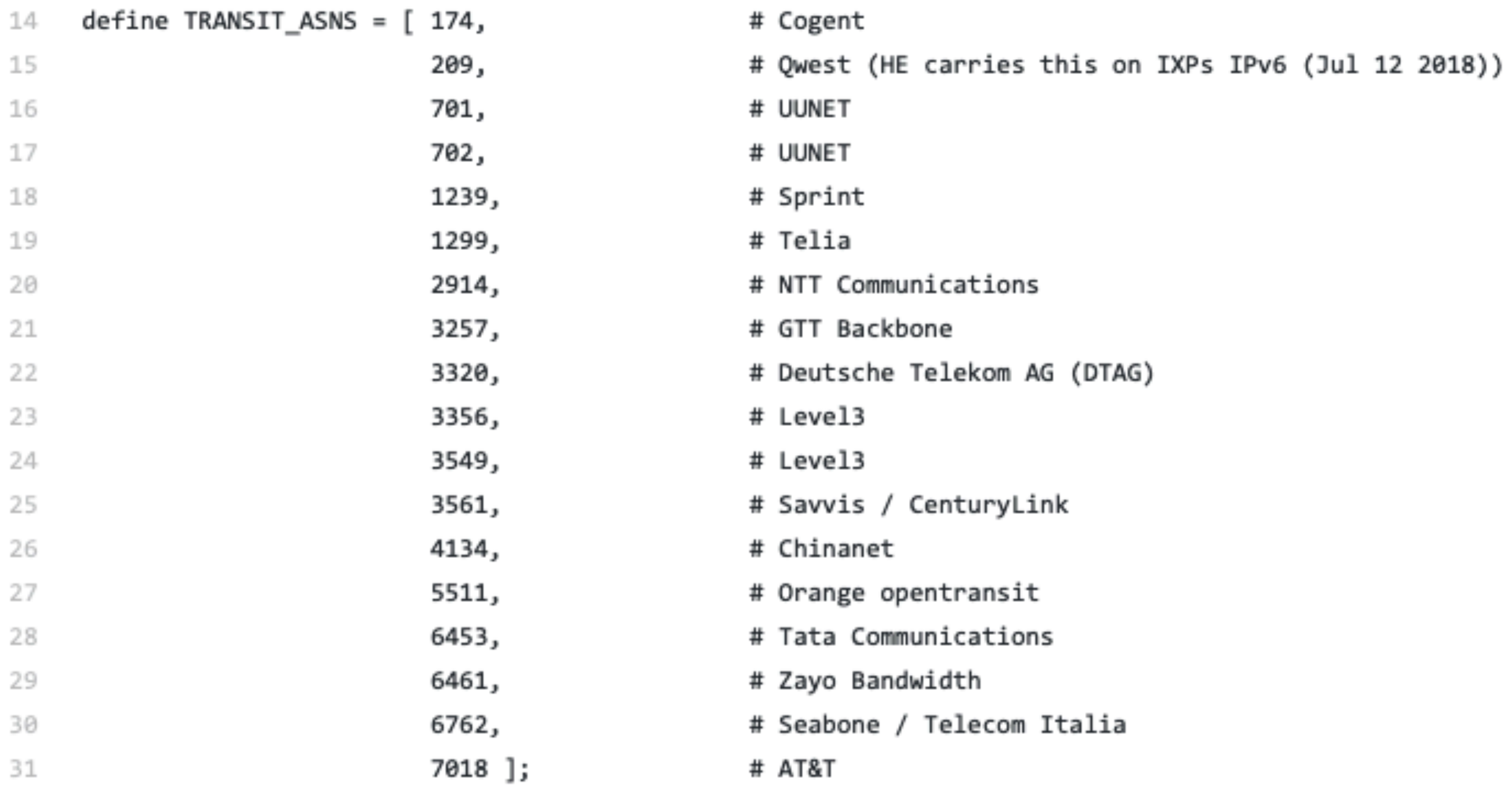
1. Small prefixes (default is > /24 / /48 for ipv4 / ipv6)
2. Martians / bogons
3. Ensure at least 1 ASN and <= 64 ASNs in path
4. Ensure peer AS is the same as first AS in the prefix's AS path
5. Prevent next-hop hijacking
6. Filter known transit networks
7. Ensure origin AS is in set of ASNs from member AS-SET
8. RPKI:
   - Valid -> accept
   - Invalid -> drop
9. RPKI Unknown -> revert to standard IRRDB prefix filtering

IXP Manager v5 Route Server Filtering

1. Small prefixes (default is > /24 / /48 for ipv4 / ipv6)
2. Martians / bogons
3. Ensure at least 1 ASN and <= 64 ASNs in path
4. Ensure peer AS is the same as first AS in the prefix's AS path
5. Prevent next-hop hijacking
6. **Filter known transit networks**
7. Ensure origin AS is in set of ASNs from member AS-SET
8. RPKI:
    - Valid -> accept
    - Invalid -> drop
9. RPKI Unknown -> revert to standard IRRDB prefix filtering

Filter Known Transit Networks

These do not peer at IX’s and they aren’t typically customers of IX participants

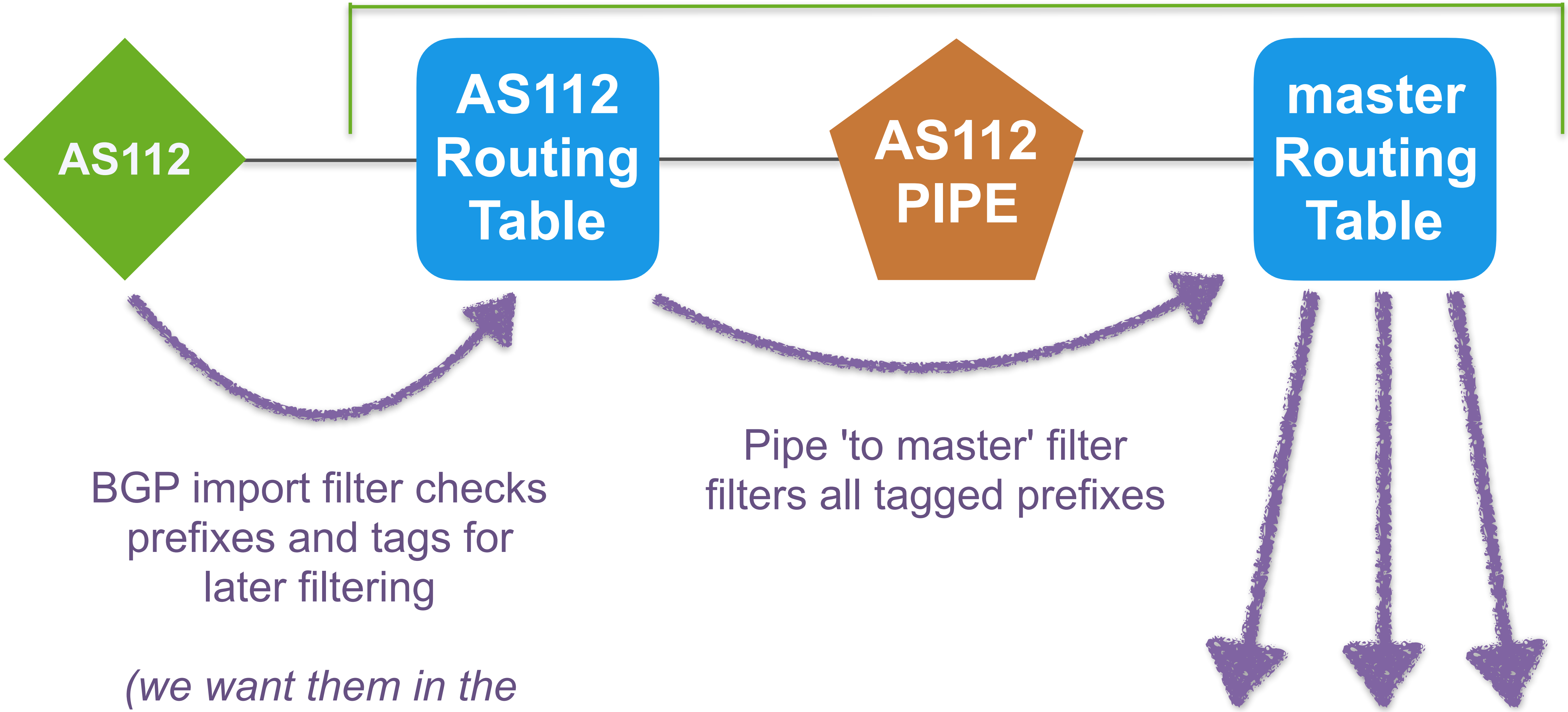


```
define TRANSIT_ASNS = [ 174,                   # Cogent
                        209,                   # Qwest (HE carries this on IXPs IPv6 (Jul 12 2018))
                        701,                   # UUNET
                        702,                   # UUNET
                        1239,                  # Sprint
                        1299,                  # Telia
                        2914,                  # NTT Communications
                        3257,                  # GTT Backbone
                        3320,                  # Deutsche Telekom AG (DTAG)
                        3356,                  # Level3
                        3549,                  # Level3
                        3561,                  # Savvis / CenturyLink
                        4134,                  # Chinanet
                        5511,                  # Orange opentransit
                        6453,                  # Tata Communications
                        6461,                  # Zayo Bandwidth
                        6762,                  # Seabone / Telecom Italia
                        7018 ];                # AT&T
```

IXP Manager v5 Route Server Filtering

1. Small prefixes (default is > /24 / /48 for ipv4 / ipv6)
2. Martians / bogons
3. Ensure at least 1 ASN and <= 64 ASNs in path
4. Ensure peer AS is the same as first AS in the prefix's AS path
5. Prevent next-hop hijacking
6. Filter known transit networks
7. Ensure origin AS is in set of ASNs from member AS-SET
8. **RPKI:**
    - **Valid -> accept**
    - **Invalid -> drop**
9. **RPKI Unknown -> revert to standard IRRDB prefix filtering**

IXP Manager v5 Bird Topology - Import From Member



BGP import filter checks prefixes and tags for later filtering

(we want them in the AS112 RT for the looking glass / analysis)

Pipe 'to master' filter filters all tagged prefixes

Route Server BGP Community Usage

Side note

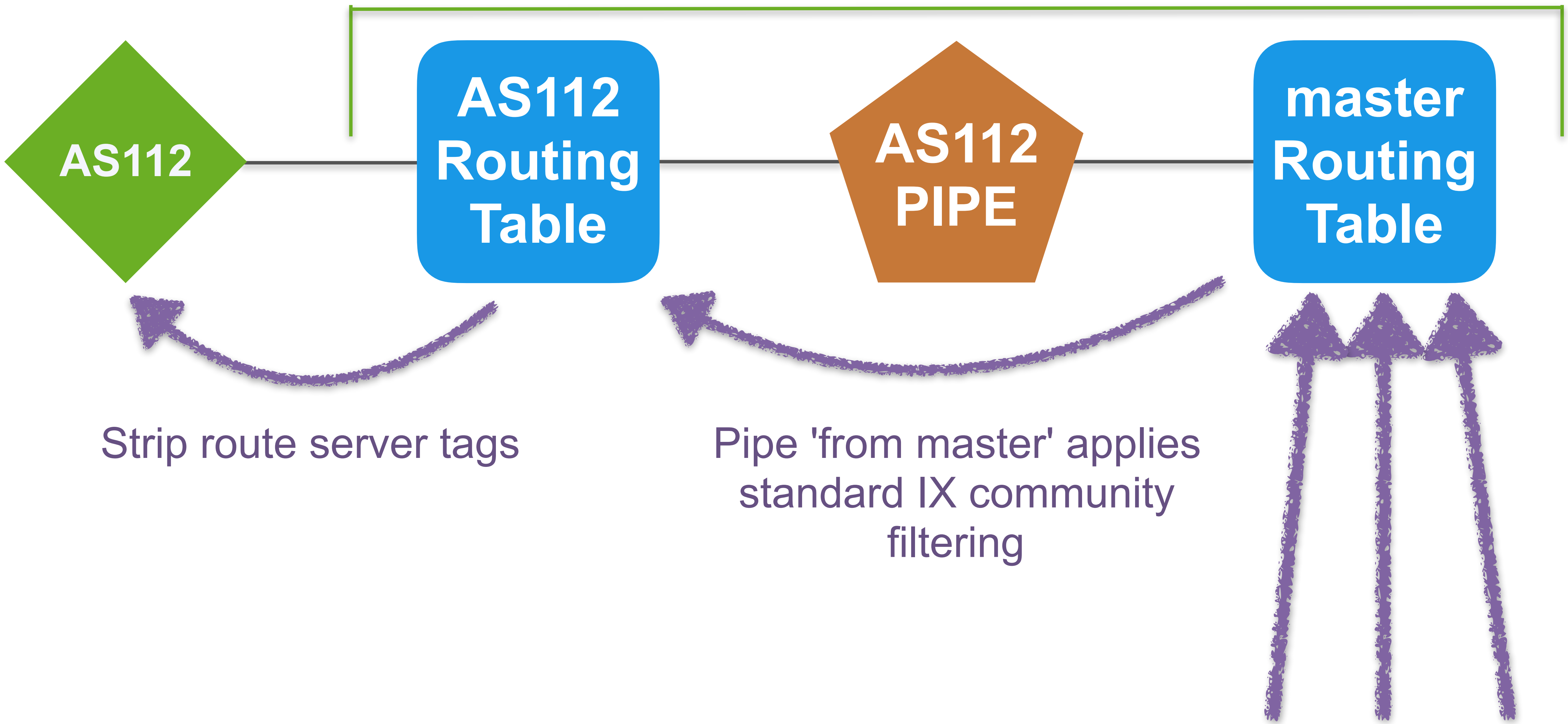
Description	Large Community
RPKI Valid	43760:1000:1
RPKI Unknown	43760:1000:2
IRRDB Valid	43760:1001:1
…	…

Description	Large Community
Bogon Prefix	43760:1101:3
IRRDB Invalid	43760:1101:9
RPKI Invalid	43760:1101:13
…	…

43760:1101:* are filtered

1. https://github.com/euro-ix/rs-workshop-july-2017/wiki/Route-Server-BGP-Community-usage

IXP Manager v5 Bird Topology - Export To Member



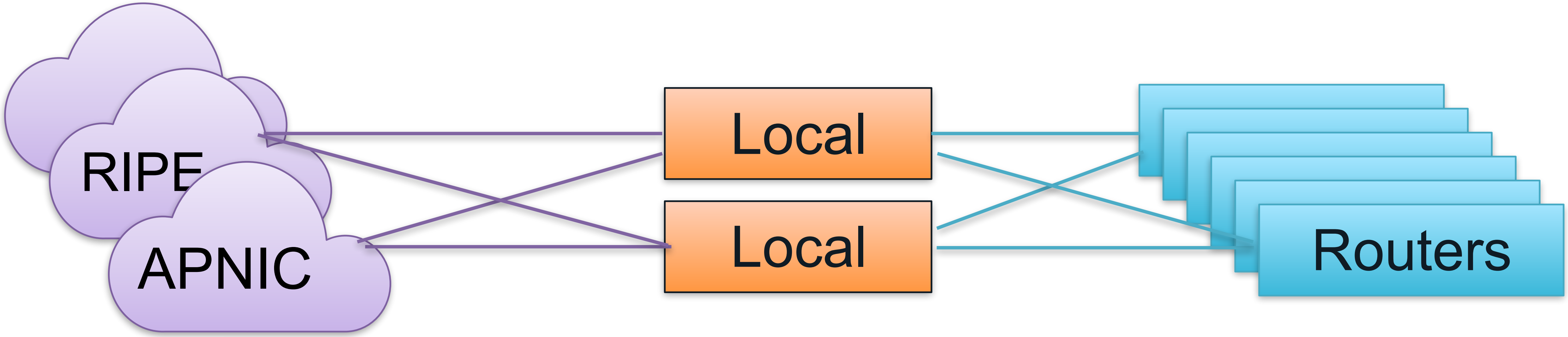
Strip route server tags

Pipe 'from master' applies standard IX community filtering

RPKI Implementation Notes

Validating BGP Routing with RPKI-RTR

- A cache server *(validator)* does the cryptographic heavy lifting
- Routers receive and maintain the set of ROAs via RPKI-RTR from the cache
- RPKI gives three validation results: VALID, INVALID, UNKNOWN



Validator Software - RIPE NCC RPKI Validator 3

- RIPE NCC RPKI Validator 3 released in 2018
  - https://github.com/RIPE-NCC/rpki-validator-3
- Dramatically reduces installation complexity
- Modest VM requirements, runs on standard OS distributions

```
$ wget https://ftp.ripe.net/tools/rpki/validator3/rc/generic/rpki-validator-latest-dist.tar.gz
$ tar zxf rpki-validator-latest-dist.tar.gz
$ ./rpki-validator-3.0-x/rpki-validator-3.sh
$ open http://localhost:8080

$ wget https://ftp.ripe.net/tools/rpki/validator3/rc/generic/rpki-rtr-server-latest-dist.tar.gz
$ tar zxf rpki-rtr-server-latest-dist.tar.gz
$ ./rpki-rtr-server/rpki-rtr-server-3.sh
```

Validator Software - Routinator 3000

- Routinator 3000 by NLnet Labs
  - https://github.com/NLnetLabs/routinator
- First impressions: low overheard, installation simplicity, stable, "just works"

```
$ curl https://sh.rustup.rs -sSf | sh
$ source ~/.cargo/env
$ cargo install routinator
$ routinator rtrd -al 127.0.0.1:3323
```

Validator Software - Cloudflare's RPKI Toolkit

- RPKI Toolkit by Cloudflare
  - https://github.com/cloudflare/cfrpki#octorpki
  - https://github.com/cloudflare/gortr
- First impressions: low overheard, installation simplicity, stable, "just works"

```
$ go get github.com/cloudflare/cfrpki/cmd/octorpki
$ mkdir tals && mkdir cache && touch rrdp.json
$ cp go/src/github.com/cloudflare/cfrpki/cmd/octorpki/tals/* tals/
$ ./go/bin/octorpki -mode server

$ go get github.com/cloudflare/gortr/cmd/gortr
$ ./go/bin/gortr -bind :3323 -cache http://localhost:8080/output.json
```

Validator Software - RPKI-RTR and Bird

```
roa4 table t_roa;

protocol rpki rpki1 {

    roa4 { table t_roa; };

    remote "192.0.2.67" port 3323;

    retry keep 90;
    refresh keep 900;
    expire keep 172800;
}
```

Validator Software - RPKI-RTR and Bird

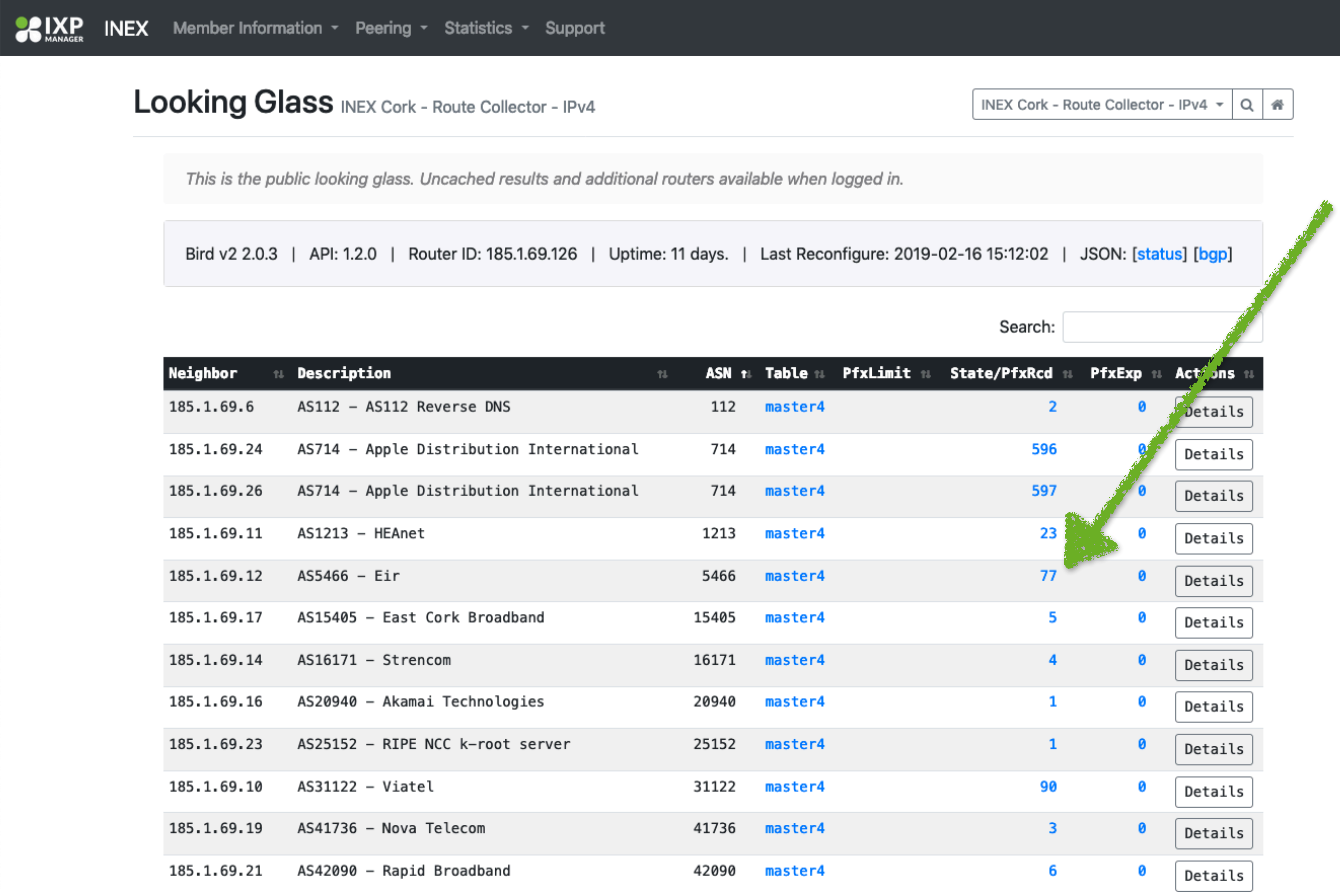
```
# RPKI check
rpki_result = roa_check( t_roa, net, bgp_path.last );


if( rpki_result = ROA_INVALID ) then {
    ...
}


# or ROA_VALID / ROA_UNKNOWN

# consider bgp_path.last_nonaggregated
```

Implementation Process at INEX

- INEX has two route servers and a route collector per LAN
- Upgrade route collector to Bird v2 + RPKI first
  - identify members who peer on the route server with RPKI invalid prefixes
  - found 4 members of ~80 with issues
    - 1 x more specific advertised than ROA allowed for
    - 1 x origin AS not matching ROA
    - 1 x member still advertising transferred space, new owners had ROAs
    - 1 x member created ROA for upstream peer-as rather than origin-as
  - members alerted to this on a "FYI basis" (i.e. non-blocking for INEX)
- Route server #1 completed Feb 7th
- Route server #2 completed Feb 14th

Implementation Process at INEX

- Outside of the four members with issues, no other member issues
- No issues to date with Bird v2
- Some issues with RIPE's validator (crashing, disk space)
- No issues with Routinator 3000, or OctoRPKI
- There's a lot in this (Bird v2, route collector vs server, large community tagging and filtering, RPKI vs IRRDB, etc.)



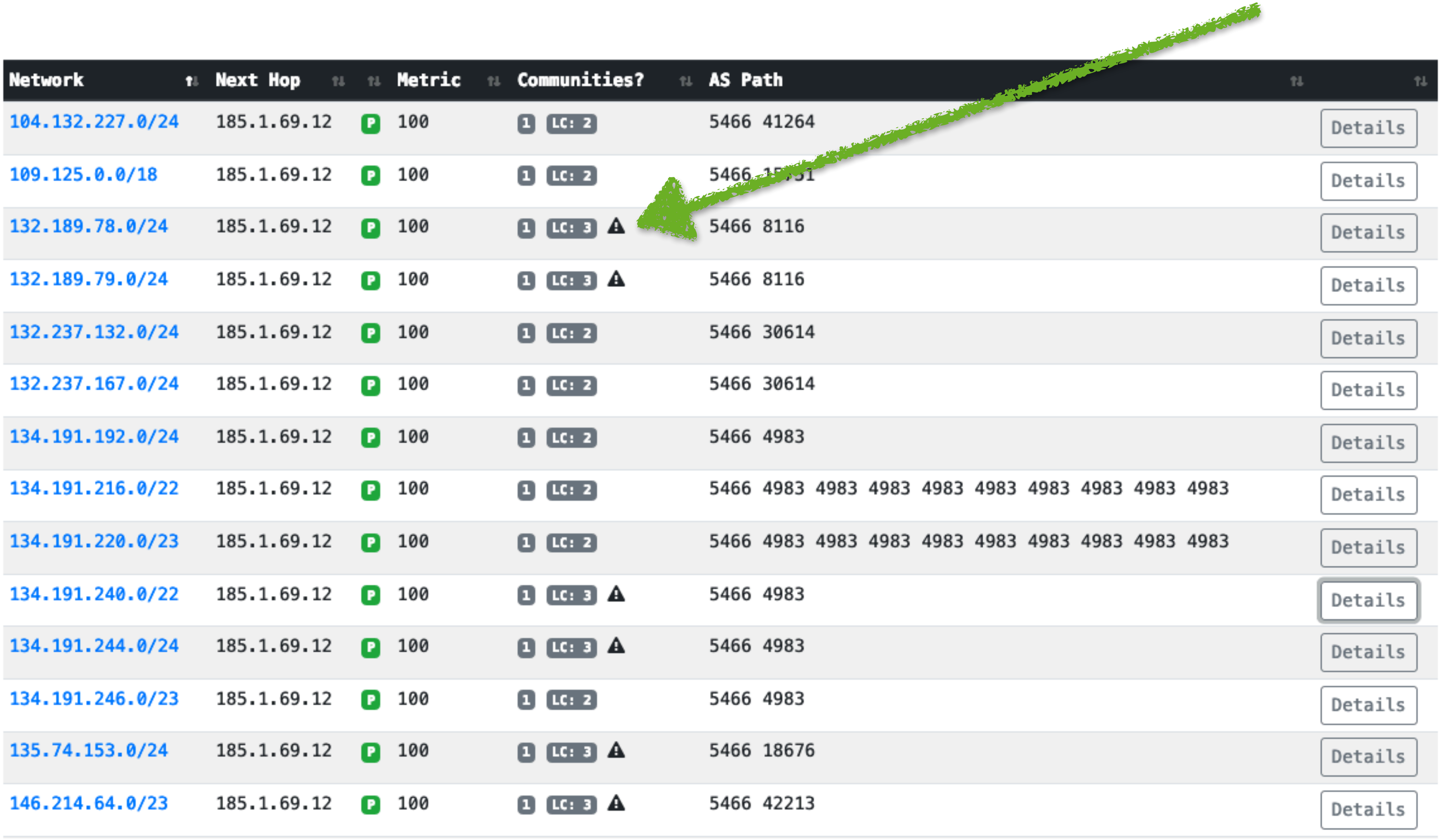
Looking Glass INEX Cork - Route Collector - IPv4

This is the public looking glass. Uncached results and additional routers available when logged in.

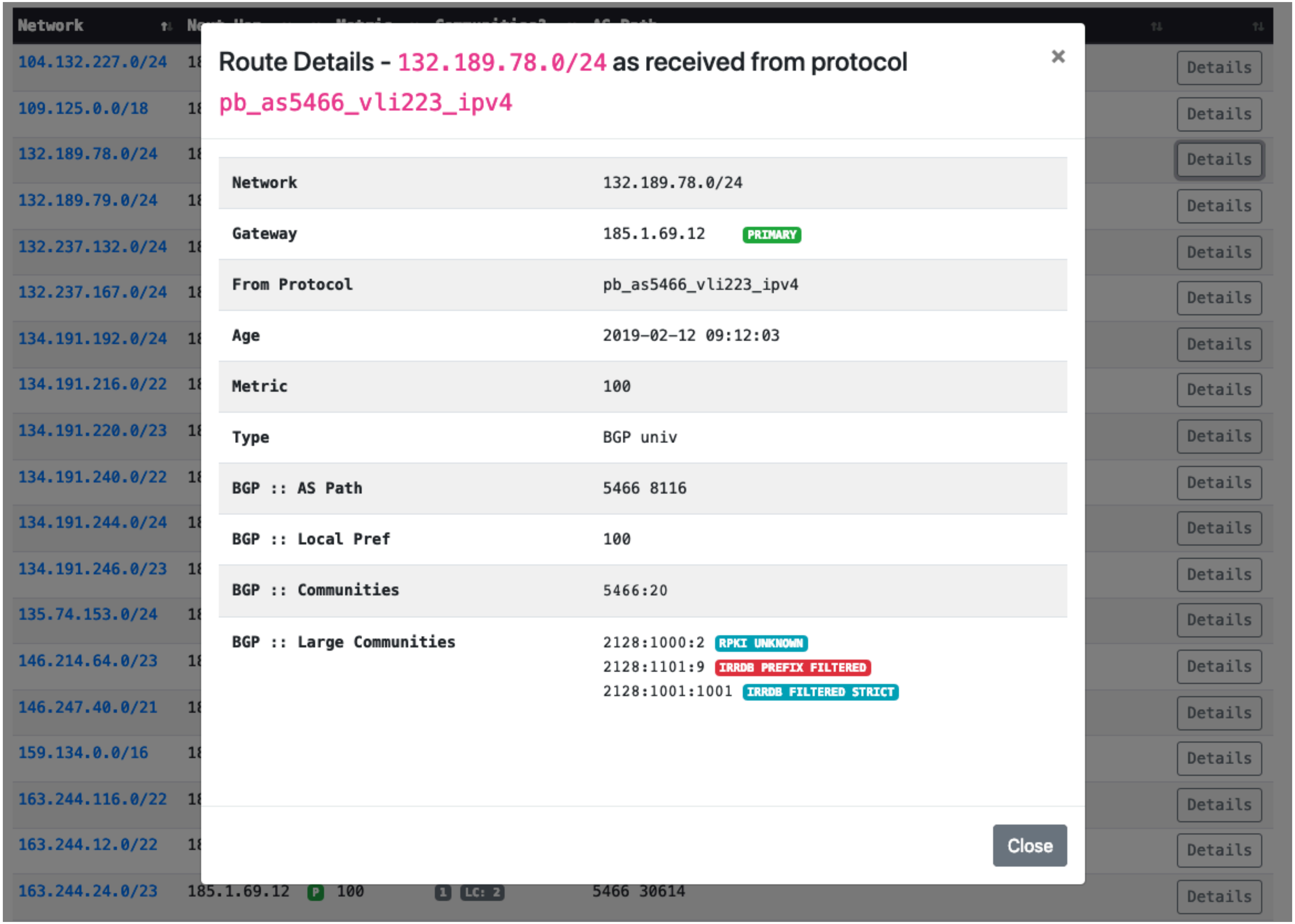
Bird v2 2.0.3 | API: 1.2.0 | Router ID: 185.1.69.126 | Uptime: 11 days. | Last Reconfigure: 2019-02-16 15:12:02 | JSON: [status] [bgp]

Search:

Neighbor	Description	ASN	Table	PfxLimit	State/PfxRcd	PfxExp	Actions
185.1.69.6	AS112 – AS112 Reverse DNS	112	master4		2	0	Details
185.1.69.24	AS714 – Apple Distribution International	714	master4		596	0	Details
185.1.69.26	AS714 – Apple Distribution International	714	master4		597	0	Details
185.1.69.11	AS1213 – HEAnet	1213	master4		23	0	Details
185.1.69.12	AS5466 – Eir	5466	master4		77	0	Details
185.1.69.17	AS15405 – East Cork Broadband	15405	master4		5	0	Details
185.1.69.14	AS16171 – Strencom	16171	master4		4	0	Details
185.1.69.16	AS20940 – Akamai Technologies	20940	master4		1	0	Details
185.1.69.23	AS25152 – RIPE NCC k-root server	25152	master4		1	0	Details
185.1.69.10	AS31122 – Viatel	31122	master4		90	0	Details
185.1.69.19	AS41736 – Nova Telecom	41736	master4		3	0	Details
185.1.69.21	AS42090 – Rapid Broadband	42090	master4		6	0	Details



Network	Next Hop		Metric	Communities?	AS Path	
104.132.227.0/24	185.1.69.12	P	100	1 LC: 2	5466 41264	Details
109.125.0.0/18	185.1.69.12	P	100	1 LC: 2	5466 15751	Details
132.189.78.0/24	185.1.69.12	P	100	1 LC: 3 ⚠	5466 8116	Details
132.189.79.0/24	185.1.69.12	P	100	1 LC: 3 ⚠	5466 8116	Details
132.237.132.0/24	185.1.69.12	P	100	1 LC: 2	5466 30614	Details
132.237.167.0/24	185.1.69.12	P	100	1 LC: 2	5466 30614	Details
134.191.192.0/24	185.1.69.12	P	100	1 LC: 2	5466 4983	Details
134.191.216.0/22	185.1.69.12	P	100	1 LC: 2	5466 4983 4983 4983 4983 4983 4983 4983 4983 4983	Details
134.191.220.0/23	185.1.69.12	P	100	1 LC: 2	5466 4983 4983 4983 4983 4983 4983 4983 4983 4983	Details
134.191.240.0/22	185.1.69.12	P	100	1 LC: 3 ⚠	5466 4983	Details
134.191.244.0/24	185.1.69.12	P	100	1 LC: 3 ⚠	5466 4983	Details
134.191.246.0/23	185.1.69.12	P	100	1 LC: 2	5466 4983	Details
135.74.153.0/24	185.1.69.12	P	100	1 LC: 3 ⚠	5466 18676	Details
146.214.64.0/23	185.1.69.12	P	100	1 LC: 3 ⚠	5466 42213	Details

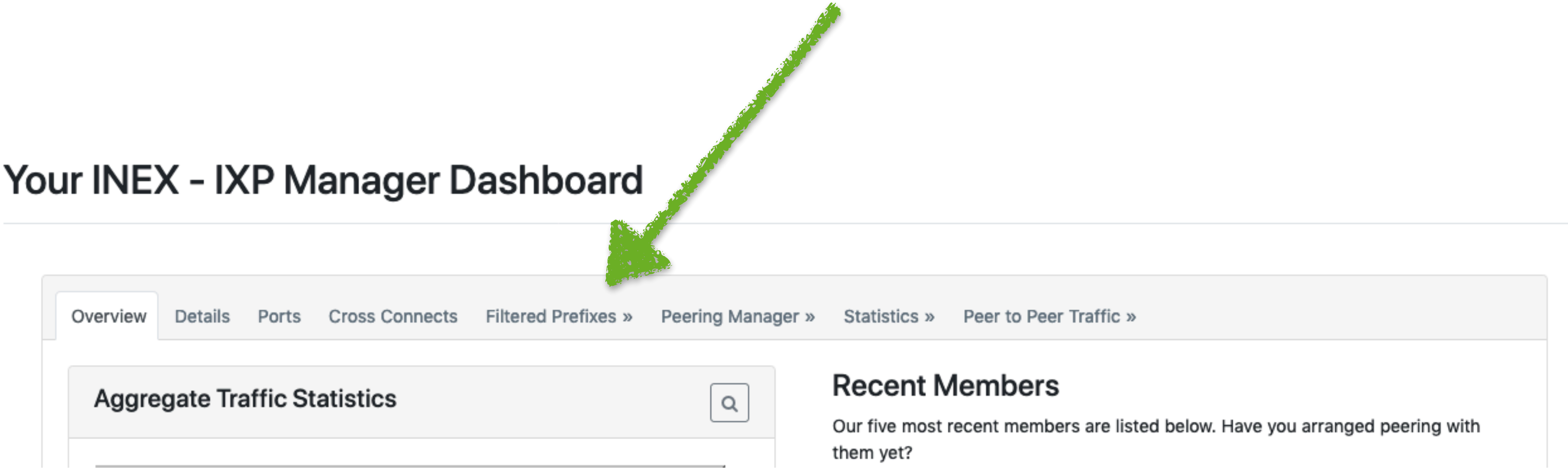


Route Details - 132.189.78.0/24 as received from protocol pb_as5466_vli223_ipv4

Network	132.189.78.0/24
Gateway	185.1.69.12 PRIMARY
From Protocol	pb_as5466_vli223_ipv4
Age	2019-02-12 09:12:03
Metric	100
Type	BGP univ
BGP :: AS Path	5466 8116
BGP :: Local Pref	100
BGP :: Communities	5466:20
BGP :: Large Communities	2128:1000:2 RPKI UNKNOWN 2128:1101:9 IRRDB PREFIX FILTERED 2128:1001:1001 IRRDB FILTERED STRICT

Network					
104.132.227.0/24	18				Details
109.125.0.0/18	18				Details
132.189.78.0/24	18				Details
132.189.79.0/24	18				Details
132.237.132.0/24	18				Details
132.237.167.0/24	18				Details
134.191.192.0/24	18				Details
134.191.216.0/22	18				Details
134.191.220.0/23	18				Details
134.191.240.0/22	18				Details
134.191.244.0/24	18				Details
134.191.246.0/23	18				Details
135.74.153.0/24	18				Details
146.214.64.0/23	18				Details
146.247.40.0/21	18				Details
159.134.0.0/16	18				Details
163.244.116.0/22	18				Details
163.244.12.0/22	18				Details
163.244.24.0/23	185.1.69.12 P	100	1 LC: 2	5466 30614	Details

New *Route Server Filtered Prefixes* Tool

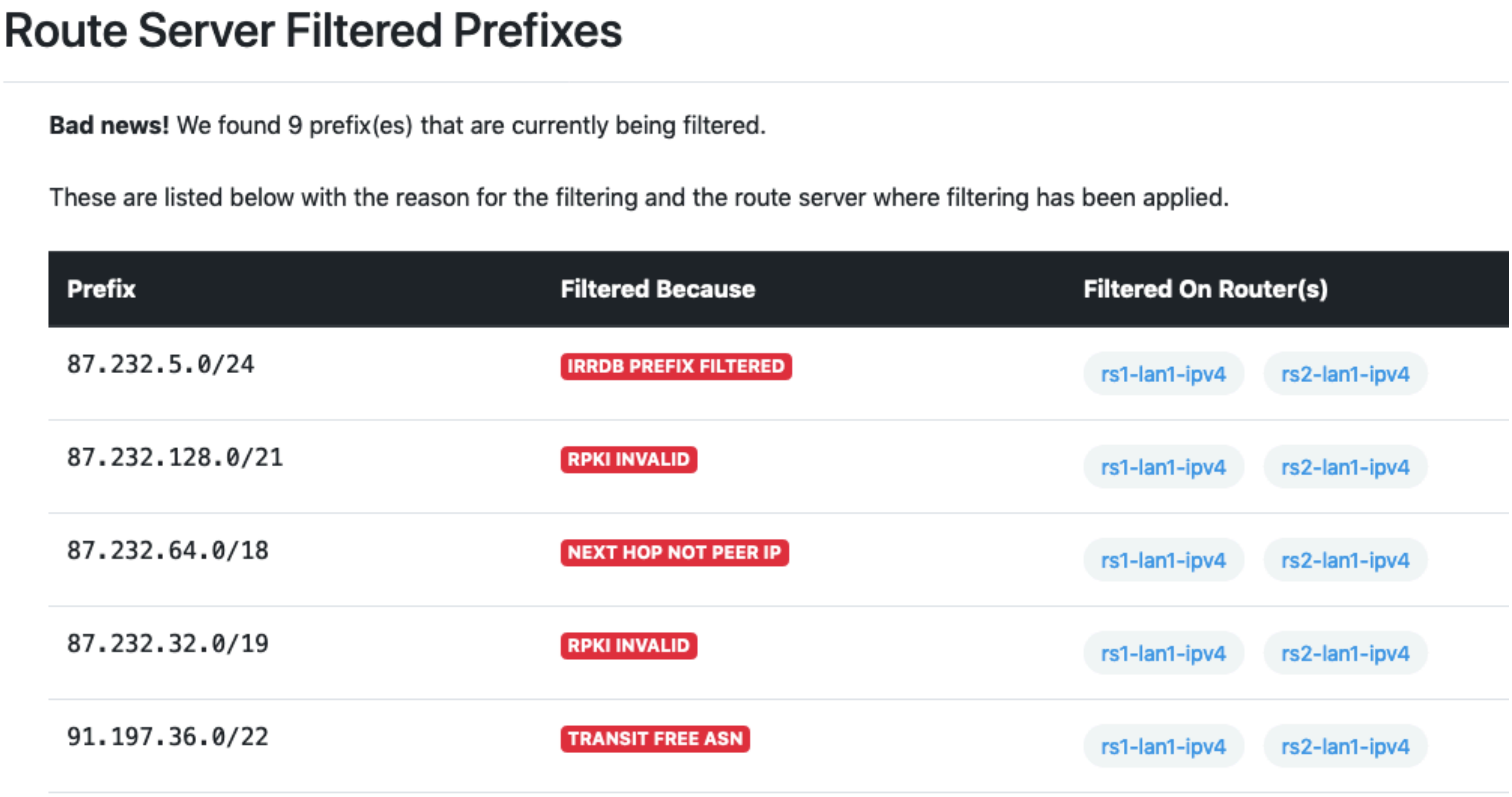


Your INEX - IXP Manager Dashboard

Aggregate Traffic Statistics

Recent Members

Our five most recent members are listed below. Have you arranged peering with them yet?



Route Server Filtered Prefixes

Bad news! We found 9 prefix(es) that are currently being filtered.

These are listed below with the reason for the filtering and the route server where filtering has been applied.

Prefix	Filtered Because	Filtered On Router(s)
87.232.5.0/24	IRRDB PREFIX FILTERED	rs1-lan1-ipv4 rs2-lan1-ipv4
87.232.128.0/21	RPKI INVALID	rs1-lan1-ipv4 rs2-lan1-ipv4
87.232.64.0/18	NEXT HOP NOT PEER IP	rs1-lan1-ipv4 rs2-lan1-ipv4
87.232.32.0/19	RPKI INVALID	rs1-lan1-ipv4 rs2-lan1-ipv4
91.197.36.0/22	TRANSIT FREE ASN	rs1-lan1-ipv4 rs2-lan1-ipv4

THANK YOU

Any Questions?

@ComePeerWithMe / @barryo79

barry.odonovan@inex.ie

https://www.inex.ie/

https://www.ixpmanager.org/
https://docs.ixpmanager.org/