Building and expanding the bgp.tools realtime BGP collector

Ben Cartwright-Cox - NANOG 89 (2023)
Quick overview of bgp.tools

Browse the Internet ecosystem

Start here...

You are connecting from
- IPv6: 2601:67c:6ec::/48
- IPv4: 192.42.116.217
- SURF B.V. (AS1101)
- DNS: 195.169.125.51
- DNS: 195.169.125.35

Latency to bgp.tools
- IPv4 End To End: 210.5ms
- IPv4 TCP Stack: 5.6ms [+/- 30.7ms]
- IPv4 TCP MSS: 1448b
- IPv6 End To End: 273.5ms
- IPv6 TCP Stack: 5.6ms [+/- 30.7ms]

Example Pages
- Cloudflare (AS13335)
- LINX LON1
- Google DNS Prefix

Recent Updates
- September 2023 Changelog
- August 2023 Changelog
- July 2023 Changelog
- June 2023 Changelog
- May 2023 Changelog

Why use BGP.Tools?

We offer for free:
- Near Realtime BGP Data
- User Friendly Interfaces
- Frequent updated external data

We offer for paid users:
- BGP Network Monitoring
- IRR Database Monitoring

Telecom Italia San Marino S.p.A
AS Number 15433
Website https://www.telecomitalia.sm

Registered on 28 Jun 2000 (22 years old)
Network status Active, Allocated under RIPE
Network type Eyeab
Prefixes Announced 75 IPv4, 0 IPv6

Upstreams
- AS8762 - Telecom Italia Sparkle S.p.A
Prefix Data (+DNS)

**Show Forward DNS**

<table>
<thead>
<tr>
<th>A</th>
<th>DNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>198.148.78.23</td>
<td>avapdproxy-01prd.vrt.sourcetfire.com</td>
</tr>
<tr>
<td>198.148.78.82</td>
<td>confluence.vrt.sourcetfire.com</td>
</tr>
<tr>
<td>198.148.78.217</td>
<td>avapvn02.vrt.sourcetfire.com, avapvn.vrt.sourcetfire.com (3 total...)</td>
</tr>
<tr>
<td>198.148.79.54</td>
<td>clamav.net</td>
</tr>
<tr>
<td>198.148.79.55</td>
<td>updates.vrt.sourcetfire.com</td>
</tr>
<tr>
<td>198.148.79.58</td>
<td>intelligence.sourcetfire.com</td>
</tr>
<tr>
<td>198.148.79.63</td>
<td>jira.talos.cisco.com, jira.vrt.sourcetfire.com</td>
</tr>
<tr>
<td>198.148.79.67</td>
<td>snapshot.clamav.net, <a href="http://www.snapshot.clamav.net">www.snapshot.clamav.net</a></td>
</tr>
</tbody>
</table>

**2620:121::/44**

Originated by **AS55219**
AS Name: **Cisco Systems, Inc.**

**Validation**

<table>
<thead>
<tr>
<th>AAAA</th>
<th>DNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2620:121:0:23::77</td>
<td>regsvc.sco.cisco.com</td>
</tr>
<tr>
<td>2620:121:0:500:217</td>
<td>scavpn.vrt.sourcetfire.com, vm</td>
</tr>
<tr>
<td>2620:121:1:500::225</td>
<td>ctyvpn.vrt.sourcetfire.com</td>
</tr>
<tr>
<td>2620:121:4:500::217</td>
<td>dtxvpn.vrt.sourcetfire.com</td>
</tr>
</tbody>
</table>

Last Update: 2023-08-31T08:25:15Z UTC

**198.148.78.0/23**

Originated by **AS55219**
AS Name: **Cisco Systems, Inc.**

Registered on 9 May 2013 (10 years old)
Registered to **ARIN-CS-985** (ARIN)
Global Looking Glass

---

Terminal

Welcome

This session is supported by:

---

Web UI

Query all public BGP sessions connected to bgp.tools

Lookup by CIDR, only applies to sessions that have been marked to be exported publicly

185.230.223.0/24

Search Filters:
Must Contain ASN: 65000

---

Supported by:

bgp.tools

---

bgp.tools> show route 2620:121::/44 match 206924
2620:121::/44  unicast [(AS206924 - Ben Cartwright-C...) Mythic CBG 0000]
Type: BGP
BGP.as_path: 206924 44684 6461 55219
BGP.community: (65532,400) [AS206924: Learned from Transit]
BGP.large.community: (44684, 0, 700) [AS44684: Route learned from peer] (44684: Route learned at Digital Realty Sovereign House / SOV) (44684, 2, 6461) [AS44684: Route learned from LONAP / London Network Access Point]
BGP.community: (65532,400) [AS206924: Learned from Transit]

Type: BGP
BGP.as_path: 206924 44684 6461 55219
BGP.community: (65532,400) [AS206924: Learned from Transit]

---

bgp.tools> show route 2620:121::/44 match 206924
2620:121::/44  unicast [(AS206924 - Ben Cartwright-C...) Mythic CBG 0000]
Type: BGP
BGP.as_path: 206924 44684 6461 55219
BGP.community: (65532,400) [AS206924: Learned from Transit]
BGP.large.community: (44684, 0, 700) [AS44684: Route learned from peer] (44684: Route learned at Digital Realty Sovereign House / SOV) (44684, 2, 6461) [AS44684: Route learned from LONAP / London Network Access Point]
BGP.community: (65532,400) [AS206924: Learned from Transit]

Type: BGP
BGP.as_path: 206924 44684 6461 55219
BGP.community: (65532,400) [AS206924: Learned from Transit]
# PIT-IX

Route Server ASN: **AS30365**

## Data Feeds Available:
- RS Feed
- Ping
- MAC Address

## Top Vendors

<table>
<thead>
<tr>
<th>Vendor</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco Systems, Inc</td>
<td>28%</td>
</tr>
<tr>
<td>Juniper Networks</td>
<td>15%</td>
</tr>
<tr>
<td>Arista Networks</td>
<td>12%</td>
</tr>
<tr>
<td>Edgecore Networks Corporation</td>
<td>5%</td>
</tr>
<tr>
<td>Other</td>
<td>15%</td>
</tr>
</tbody>
</table>

## List of members (236 routers over 211 ASNs):**

<table>
<thead>
<tr>
<th>ASN</th>
<th>Description</th>
<th>IPv4</th>
<th>IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS45437</td>
<td>Real World - The Core</td>
<td>198.32.141.115</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS52772</td>
<td>S.JE.TELECOMM. EIREL</td>
<td>198.32.141.89</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS53180</td>
<td>INFOTEL, COMM. LTD</td>
<td>198.32.141.51</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS53667</td>
<td>FranTech Solutions</td>
<td>198.32.141.45</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS1031</td>
<td>Peer 1 Internet Service LLC</td>
<td>198.32.141.44</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS271253</td>
<td>TELECOMM. EIREL</td>
<td>198.32.141.43</td>
<td>2001:504:1:a604</td>
</tr>
<tr>
<td>AS2734</td>
<td>CoreSite</td>
<td>198.32.141.41</td>
<td>2001:504:1:a604</td>
</tr>
</tbody>
</table>

## List of members (39 routers over 31 ASNs):**

<table>
<thead>
<tr>
<th>ASN</th>
<th>Description</th>
<th>IPv4</th>
<th>IPv6</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS40078</td>
<td>Pittsburgh Internet Exchange</td>
<td>206.71.141.6</td>
<td>2001:504:77::6</td>
<td>100 Gbps</td>
</tr>
<tr>
<td>AS40078</td>
<td>Pittsburgh Internet Exchange</td>
<td>206.71.141.7</td>
<td>2001:504:77::7</td>
<td>100 Gbps</td>
</tr>
<tr>
<td>AS21232</td>
<td>bgp.tools Route Collector</td>
<td>206.71.141.9</td>
<td>2001:504:77::9</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>AS20326</td>
<td>TeraSwitch Networks Inc.</td>
<td>206.71.141.10</td>
<td>2001:504:77::10</td>
<td>100 Gbps</td>
</tr>
<tr>
<td>AS18225</td>
<td>Cloudflare Inc</td>
<td>206.71.141.11</td>
<td>2001:504:77::11</td>
<td>10 Gbps</td>
</tr>
</tbody>
</table>
## PIT-IX

**Route Server ASN:** AS30365

### Data Feeds Available:
- RS Feed
- Ping
- MAC Address

### Top Vendors
- Cisco Systems, Inc
- Juniper Networks
- Arista Networks
- Edgecore Networks Corporation

### List of members (236 routers over 211 ASNs):

<table>
<thead>
<tr>
<th>ASN</th>
<th>Description</th>
<th>IPv4</th>
<th>IPv6</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS45437</td>
<td>Real World - The Core</td>
<td>196.32.141.115</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
<tr>
<td>AS25177</td>
<td>SJET TELECOMunicaciones - IRELI</td>
<td>196.32.141.89</td>
<td>2001:504:1:a604</td>
<td>15%</td>
</tr>
<tr>
<td>AS55310</td>
<td>INFOPORT Telecommunication Ltda</td>
<td>196.32.141.51</td>
<td>2001:504:1:a604</td>
<td>15%</td>
</tr>
<tr>
<td>AS41327</td>
<td>Fiber Telecom s.p.a.</td>
<td>196.32.141.50</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
<tr>
<td>AS35857</td>
<td>FanTech Solutions</td>
<td>196.32.141.45</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
<tr>
<td>AS1031</td>
<td>Peer 1 Internet Service LLC</td>
<td>196.32.141.44</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
<tr>
<td>AS271253</td>
<td>LINK BRASIL TELECOMunicaciones Ltda</td>
<td>196.32.141.43</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
<tr>
<td>AS2734</td>
<td>CoreSite</td>
<td>196.32.141.41</td>
<td>2001:504:1:a604</td>
<td>5%</td>
</tr>
</tbody>
</table>

### List of members (39 routers over 31 ASNs):

<table>
<thead>
<tr>
<th>ASN</th>
<th>Description</th>
<th>IPv4</th>
<th>IPv6</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS400798</td>
<td>Pittsburgh Internet Exchange</td>
<td>206.71.141.6</td>
<td>2001:504:77::6</td>
<td>100 Gbps</td>
</tr>
<tr>
<td>AS400798</td>
<td>Pittsburgh Internet Exchange</td>
<td>206.71.141.7</td>
<td>2001:504:77::7</td>
<td>100 Gbps</td>
</tr>
<tr>
<td>AS212232</td>
<td>bgp.tools Route Collector</td>
<td>206.71.141.9</td>
<td>2001:504:77::9</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>AS20326</td>
<td>TeraSwitch Networks Inc.</td>
<td>206.71.141.10</td>
<td>2001:504:77::10</td>
<td>10 Gbps</td>
</tr>
</tbody>
</table>

### Showing routes on "PIT-IX" route servers that point to the next hop of 206.71.141.6, 2001:504:77::6.

<table>
<thead>
<tr>
<th>Session</th>
<th>Prefix</th>
<th>BGP Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIT-IX-RS1-4</td>
<td>23.143.152.0/24</td>
<td>AS30365, AS400798</td>
</tr>
<tr>
<td>PIT-IX-RS1-6</td>
<td>2602:faaa:/36</td>
<td>AS30365, AS400798</td>
</tr>
<tr>
<td>PIT-IX-RS2-4</td>
<td>23.143.152.0/24</td>
<td>AS30365, AS400798</td>
</tr>
<tr>
<td>PIT-IX-RS2-6</td>
<td>2602:faaa:/36</td>
<td>AS30365, AS400798</td>
</tr>
</tbody>
</table>

Click here to go back
Traceroutes/Looking Glass/Agents

You need a bgp.tools (free) + RIPE Atlas account for this

Orange S.A.
AS Number 5511

Select BGP Session to query:

London [IPv4] [IPv6]

Input Prefix:

80.80.80.80

Query

80.80.80.0/24  unicast [London 0680-08-08] * (-/-) [AS66679]
Type: BGP
BGP.as path: 5511 3356 39247 60679
BGP.community: [AS5511: United Kingdom] [AS5511: Route received from peering partner] [AS5511: Route received in Europe from peering] [AS5511: TUNE announce to US peers]

This is a GIF, Sorry PDF users!
Network Ranking

Can be ranked by Global or ASN Country using:

- Peer Count (*)
- AS Cone
- Eyeball Population
- Domain Records
- IPv4/IPv6 space originated

* is improved by feeding bgp.tools BGP data

https://bgp.tools/rankings/MX?sort=cone
Core points

- Bgp.tools was built out of the frustration I had with similar tools
- Practically realtime BGP data, updates fast enough to use as live feedback
- The horrors of WHOIS is handled, and in some cases is updated in near real time
- Most data is frequently updated:
  - ICMP Ping data scans of IPv4 /0
  - IPv4 and IPv6 RDNS data
  - Forward DNS data (Looking what A or AAAA records point to a prefix)
- Peering IXP data is provided:
  - Like what people are sending to Route Servers
  - What vendors they are running on the exchange
  - If they are doing (very) remote peering on the exchange
Making the bgp.tools I want possible
Challenges running bgp.tools

- Getting low latency and accurate BGP data to use
- Building a scalable system to avoid being picky on feeds
- Collecting relevant data
- Not going bankrupt
The inner runnings of bgp.tools

- Most critically BGP path data

```bash
# birdc s ro 80.80.80.0/24 all
BIRD 2.0.7 ready.
Table master4:
80.80.80.0/24  unicast [transit4_velox_2 2023-09-19] * (100) [AS60679i]
    via 193.35.59.46 on eno1.601
Type: BGP univ
BGP.origin: IGP
BGP.as_path: 3170 6461 7385 30247 60679
BGP.next_hop: 193.35.59.46
BGP.local PREF: 10
BGP.community: (60945,0) (60945,5459) (65532,400)
```
Standard BGP data sources
Using public data sources

- RIPE RIS (RIS) and RouteViews (RV) export MRT dumps
  - MRT Dumps come in two types, a RIB (aka a full table dump) and "messages" (a copy of all BGP messages in the last 15 mins)
  - Table dumps are done 4 to 8 hours, message files are provided every 15 mins
    - (Most of the time)

- Bgp.tools started in 2018 by using RIS and RV MRT table dumps.
  - I quickly learned the quirks of using RIS and RV as "Production" data sources…
RIS and RV quirks to control for

- **Table dumps**
  - Only show up every 4 - 8 hours
  - Make it hard to remove individual sessions that are known to be bad

- **Message dumps** *(never used by bgp.tools production in the end)*
  - If a message file never shows up, you have to wait until the next dump file (4 - 8 hours) before becoming reliability in sync again.
  - People "UPDATE flood" collectors by mistake, making these archives sometimes huge and a pain to decode

- **General**
  - Huge bias to AS6939 (HE)
  - They are on almost all of the large IXPs, and provide you 180k+ of peered v4 routes that will likely be preferred over transit, hiding transit paths from the collector
Going beyond RIS and RV

- Eventually in 2021 after a number of issues with MRT files from RV and RIS, bgp.tools started to build its own route collector
  - Issues like moderation, bad data, stuck routes
  - Reducing site data latency to be less than 8 hours behind with message files would be the same effort as building my own collector
- Decided that a multihop eBGP only collector was viable to start with

- It was clear that no "normal" BGPd was going to work for the scale I wanted, a custom suite of bgp software needed to be written
- Bootstrapped with a live copy of the NLNOG Ring route collector: https://lg.ring.nlnog.net/
"neo-bgp"

A purpose built "bgpd" for the exact use case that bgp.tools wants
"neo-bgp"

- Each BGP Session is in its own process
  - PIDs crash independently, memory per process is manageable
  - Upgrades can happen on a single BGP session at a time
  - Entire system scales to as many CPU cores as your system has
- No need to implement router-useful functions
  - Bgp.tools is only interested in getting BGP paths and AS summary computations as fast as possible
- Feature implementation moves to the bgpd, not a polling worker over N many sessions
Bgp.tools currently sits at 1360~ BGP sessions
  ○ 750~ full IPv4 tables
  ○ 1210~ full IPv6 tables
  ○ 1,000,000,000+ BGP Paths stored in RAM

Hardware is modest, entire site operates inside ¼ cab with room to spare
  ○ 512G DDR4 per machine
  ○ 32~ cores per machine
  ○ 3 active machines right now, 3 available to turn on (when I want to pay for power)

Running bgp.tools on a cloud provider would cost around $12k USD a month
  ○ In reality it costs 15-20x less than that in colo
Challenges running bgp.tools

- ✔ Getting low latency and accurate BGP data to use
- ✔ Building a scalable system to avoid being picky on feeds
- □ Collecting relevant data
- □ Not going bankrupt
"neo-bgp"

- The site is funded by offering rapid BGP/IRR/RPKI monitoring (and historical searching on your own feeds)
- The neo-bgp architecture allows me to send alerts as fast as I can get data for them!
- There are a bunch of other paid user features, but I don't want to turn this into a major sales pitch
Challenges running bgp.tools

- ☑ Getting low latency and accurate BGP data to use
- ☑ Building a scalable system to avoid being picky on feeds
- ☐ Collecting relevant data
- ☑ Not going bankrupt
Internet Exchange Route Collection
Status Quo

- Most of the RIS / RouteViews collectors live on internet exchanges

- This has some advantages, as networks can peer with route collectors over shared L2 fabrics
Issues with IXP route collection

- RIPE RIS has ~1535 BGP sessions online,
  - But 372 / 407 Full IPv4/IPv6 tables
  - (by their own calculations)
  - $372 + 407 = 779$. Far off the 1535 total session count
  - Many people peer with RIS, but only send their customer routes
    - This is not entirely helpful…
Other problems with IXP Route Collection

- Really expensive if you don't have friends
  - IXP Membership fees + XC fees + colo fees
  - IXP membership alone can be more than the last two
  - [https://peering.exposed](https://peering.exposed)

- Even if the IXP can be done for free, the power to power the machine or transport to another place is likely also non trivially expensive
Getting creative to solve for XC Fees / Colo

- What is the cheapest, smallest, most insane thing we could ship to a *willing* IXP?
Getting creative to solve for XC Fees / Colo

- What is the cheapest, smallest, most insane thing we could ship to a *willing* IXP?

https://blog.benjojo.co.uk/post/smart-sfp-linux-inside
Getting creative to solve for XC Fees / Colo

- What is the cheapest, smallest, most insane thing we could ship to a willing IXP?

  - No XC, The switch is the power supply, you can hitch backhaul either via someone friendly on the IXP, or relaying via a VPS or something
  - Cheap, Around 150 USD all in
  - Single core ARMv7, with 512M of RAM running Debian Jessie
  - **Completely crazy.** Everyone is going to look at you like you lost the plot!
  - Made by a Russian/Dubai company since the Russian Invasion of Ukraine

[https://blog.benjojo.co.uk/post/smart-sfp-linux-inside](https://blog.benjojo.co.uk/post/smart-sfp-linux-inside)
Creative solutions are available
Creative solutions are available

- Runs a 400Mhz~ 32bit MIPS core, 32MB of RAM
- The constrained RAM and MIPS CPU μArch makes this a challenge to program for
- Thankfully Zig lang has a mostly working MIPS target!
- To use as a generic "Linux box" you must perform some software changes
- Vendor has been really keen and helpful with modding these

- Similar tech is available via Huawei/Nokia/FS.COM (they share a chipset and design) for 80 USD~ per optic
The actual preference tree

1. Some IXPs have VM infrastructure on the exchange that is easy to use, bgp.tools can run a relay in 128MB of RAM and very low CPU requirements
2. Those magic Linux optics are easy and convenient to ship around
   ○ But are mildly scary for some, also 1G only, and IXPs are sunsetting 1G ports
3. At worst I can ship physical 1U hardware around
   ○ Ideally want to try and land as many IXPs in a single machine to conserve funds
All sessions lead back to London

- You have noticed it isn't really possible to store a *modern* full internet table on 32MB of RAM.

- Instead of storing sessions locally, the local collector will "rehost" the BGP session back in London where all of the website infrastructure is.

- This is because with how bgp.tools is designed, all BGP data has to be within 3ms~ of the web server to ensure a enjoyable experience.
Current Deployments

- JINX, DINX, CINX, NMBINX
- ONIX, QIX
- NL-IX, INTERIX, Fry's-IX
- BCIX, Stuttgart-IX, DE-CIX {FRA,DUS,LEJ,HAM,MUC}
- RomandIX
- LONAP
- PIT-IX, DE-CIX {New York, Dallas, Chicago, Richmond, Phoenix}
- BIX.BG
- DE-CIX {Madrid, Barcelona}
- DE-CIX Marseille
- DE-CIX Palermo
- LU-CIX
- SOX Serbia
- MSK-IX Moscow
- DE-CIX Istanbul
- DE-CIX Lisbon

In the pipeline:
- IX.BR (?)
- THINX WAW
- SF-MIX (?)
- STH-IX (?)
- NIX.CZ
- Interlan
- FIXO (?)

Bgp.tools is always looking for better visibility into IXPs!

Do you run an IXP not listed here?
admin@bgp.tools
Challenges running bgp.tools

- ✔ Getting low latency and accurate BGP data to use
- ✔ Building a scalable system to avoid being picky on feeds
- ✔ Collecting relevant data
- ✔ Not going bankrupt
Setting up feeds is easy

Go to (PeeringDB SSO is supported): [https://bgp.tools/kb/setup-sessions](https://bgp.tools/kb/setup-sessions)

You can **instantly** setup sessions to bgp.tools. Where you **should** export a full table. You can peer using eBGP Multihop or via a IXP collector where available.

Export to 3rd parties/Looking Glass visibility is entirely optional!
Questions?

Want to feed bgp.tools?
go to bgp.tools and go to bottom link "Contribute Data"

More complex queries:
IRC: Benjojo-bgptools (terahertz) / benjojo (everything else)
Or email: admin@bgp.tools