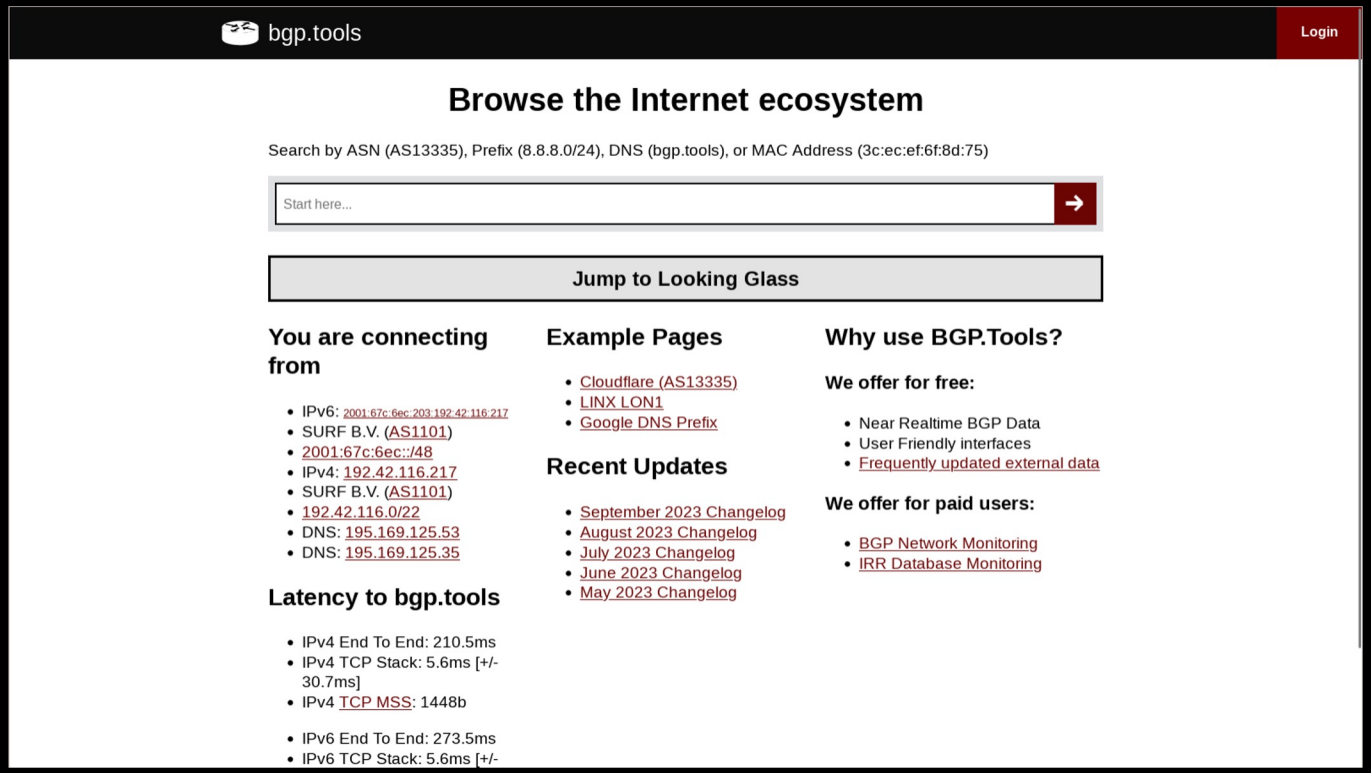
Building and expanding the bgp.tools realtime BGP collector

Ben Cartwright-Cox - NANOG 89 (2023)

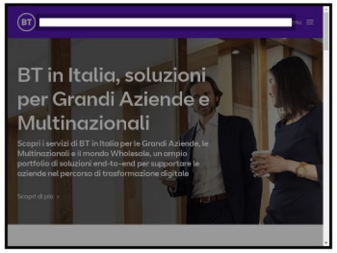
bgp.tools

d6253275ab09f7afeaca74955a6cd164

Quick overview of bgp.tools



ASN Info



BT Italia S.p.A.

AS Number **8968**
Website **http://www.bt.com/italia**

Overview | Prefixes | Connectivity | Whois | IX

Peers
57

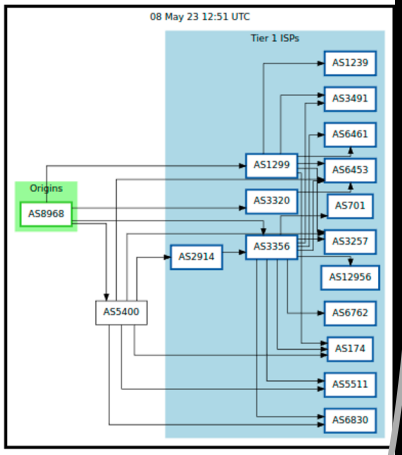
Upstreams
4

Downstreams
31 (Cone: 37)

Network Policy

Click here to learn more about what this graph means and what makes up a network policy

Policy awesome_curie



Prefixes in selected policy

-	Prefix	Description
	85.20.128.0/17	BT Italia S.p.A.
	213.217.128.0/18	BT Italia S.p.A.

How are upstreams and downstreams calculated?

Upstreams

	ASN	Description	IPv4	IPv6
	AS1299	Arelion (fka. Telia Carrier)	✔	✔
	AS3356	Lumen (Level 3)	✔	✔
	AS5400	British Telecommunications PLC	✔	✔
	AS3320	Deutsche Telekom AG	✔	✔

Peers

	ASN	Description	IPv4	IPv6
	AS1299	Arelion (fka. Telia Carrier)	✔	✔
	AS3356	Lumen (Level 3)	✔	✔
	AS3313	BT Italia S.p.A.	✔	✔
	AS5400	British Telecommunications PLC	✔	✔
	AS33942	IREN ENERGIA S.P.A	✔	✘
	AS3303	Swisscom (Schweiz) AG	✘	✔
	AS28716	Retelit Digital Services S.p.A.	✘	✔
	AS1239	T-Mobile US (Sprint)	✘	✔
	AS13237	euNetworks GmbH	✘	✔

Downstreams

	ASN	Description	IPv4	IPv6
	AS3313	BT Italia S.p.A.	✔	✔
	AS33942	IREN ENERGIA S.P.A	✔	✘
	AS47224	Telecom Italia S.p.A.	✔	✘

Prefix Data (+DNS)

Overview | Connectivity | Whois | DNS | Validation

Show Forward DNS

A	DNS
198.148.78.23	avapdproxy-01prd.vrt.sourcefire.com
198.148.78.82	confluence.vrt.sourcefire.com
198.148.78.217	avavpn02.vrt.sourcefire.com, avavpn.vrt.sourcefire.com (3 total...)
198.148.79.54	clamav.net
198.148.79.55	updates.vrt.sourcefire.com
198.148.79.58	intelligence.sourcefire.com
198.148.79.63	jira.talos.cisco.com, jira.vrt.sourcefire.com
198.148.79.67	snapshot.clamav.net, www.snapshot.clamav.net

2620:121::/44

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

Overview | Connectivity | Whois | DNS | Validation

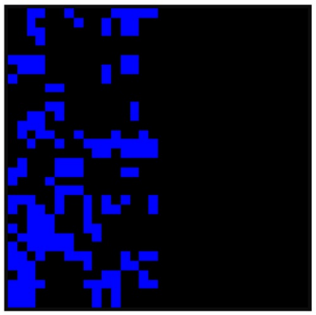
AAAA	DNS
2620:121:0:23::77	regsvc.sco.cisco.com
2620:121:0:500::217	scavpn.vrt.sourcefire.com, v
vm 2620:121:1:59:250:56ff:fe96:bb7a	stage.regsvc.sco.cisco.com
2620:121:1:500::225	cilvpn.vrt.sourcefire.com
2620:121:4:500::217	dtxvpn.vrt.sourcefire.com

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

bgp.tools Start here...

Logged in as AS206924

View | Super LG



198.148.78.0/23

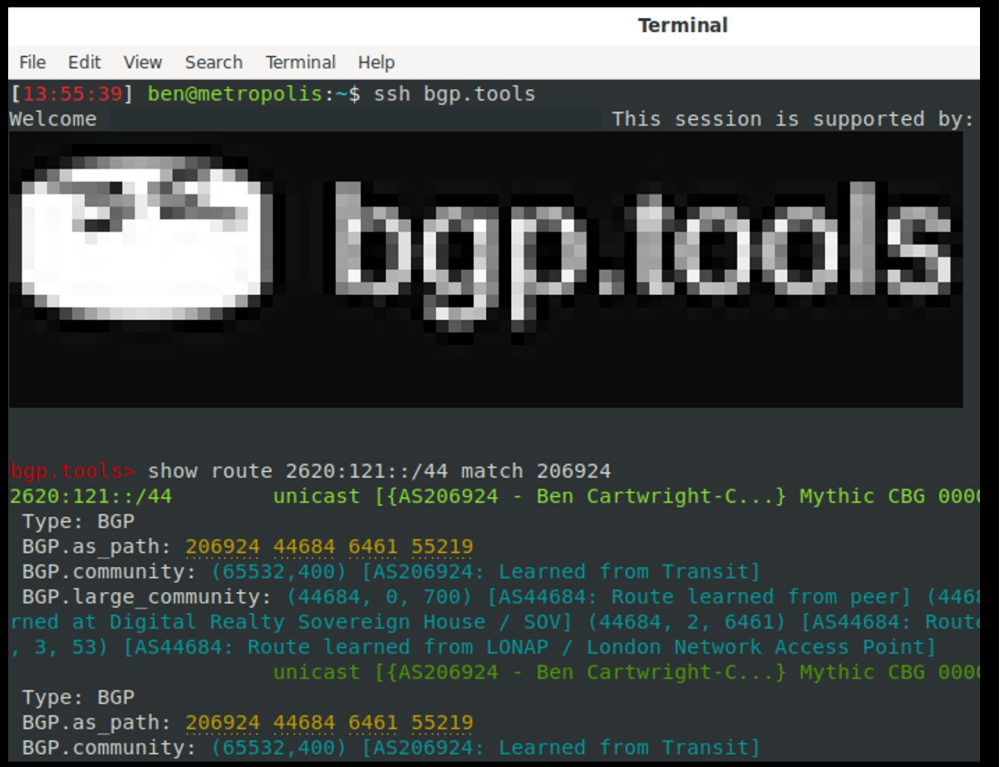
Originated by AS55219
AS Name: **Cisco Systems, Inc.**

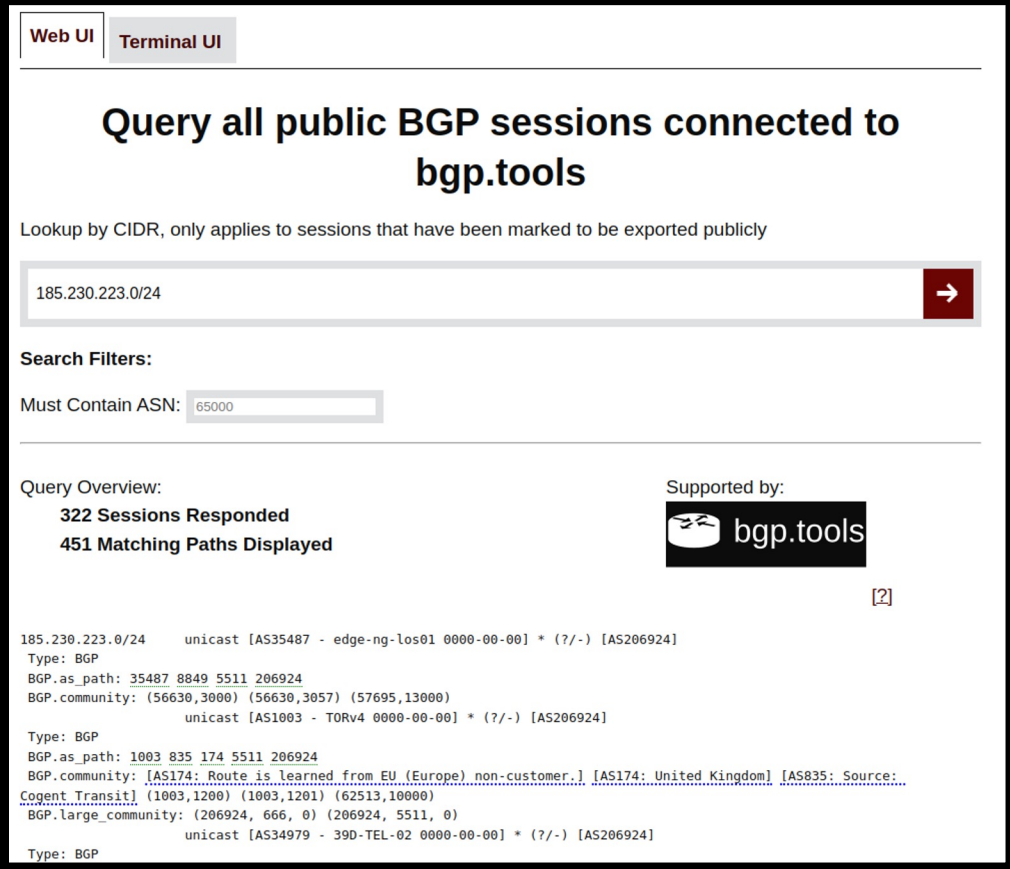
Overview | Connectivity | Whois | DNS | Validation

Registered on
9 May 2013 (10 years old)

Registered to
ARIN-CS-985 **(ARIN)**

Global Looking Glass





IXP Info Pages

NYIIX New York

Go to PeeringDB page

Go to IXP-DB page

Data Feeds Available:

RS Feed, Ping, MAC Address

Do you run this IX and want to help with feeds? Contact Us!

List of members (236 routers over 211 ASNs):

ASN	Description	IPv4	IPv6
AS45437	Real World - The Core	198.32.161.115	2001:504:1::a504
AS52772	SJNET TELECOMUNICACOES - EIRELI	198.32.161.89	2001:504:1::a526
AS53180	INFORTEL COMUNICACOES LTDA	198.32.161.51	2001:504:1::a505
AS41327	Fiber Telecom S.p.A.	198.32.161.50	2001:504:1::a504
AS53667	FranTech Solutions	198.32.161.45	2001:504:1::a505
AS1031	Peer 1 Internet Service LLC	198.32.161.44	2001:504:1::a500
AS271253	LINK BRASIL TELECOMUNICACOES LTDA	198.32.161.43	2001:504:1::a527
AS2734	CoreSite	198.32.161.41	2001:504:1::a500

PIT-IX

Go to PeeringDB page

Go to IXP-DB page

Route Server ASN: AS30365

Data Feeds Available:

RS Feed, Ping, MAC Address

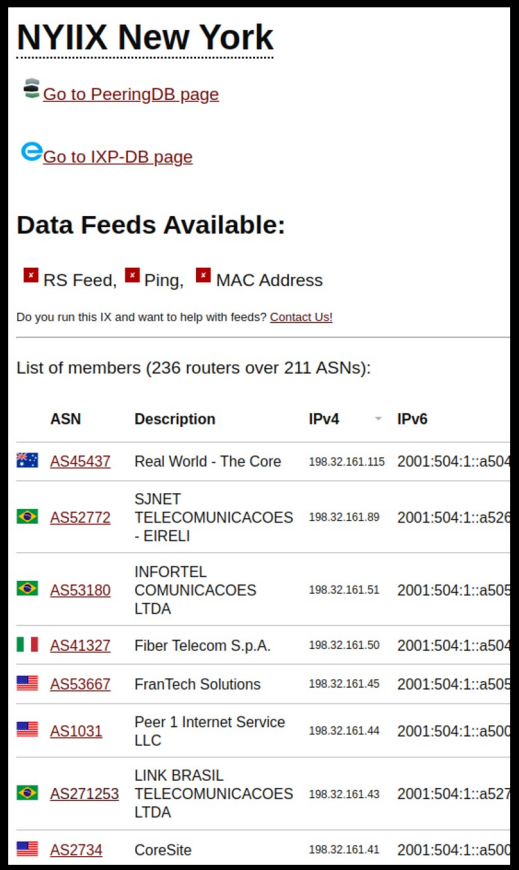
Top Vendors

Vendor	%
Cisco Systems, Inc	28%
Juniper Networks	15%
Arista Networks	12%
Edgecore Networks Corporation	5%
Other	15%

List of members (39 routers over 31 ASNs):

ASN	Description	IPv4	IPv6	Speed
AS400798	Pittsburgh Internet Exchange	206.71.141.6	2001:504:77::6	100 gbps
AS400798	Pittsburgh Internet Exchange	206.71.141.7	2001:504:77::7	100 gbps
AS212232	bgp.tools Route Collector	206.71.141.9	2001:504:77::9	10 gbps
AS20326	TeraSwitch Networks Inc.	206.71.141.10	2001:504:77::10	100 gbps
AS13335	Cloudflare, Inc.	206.71.141.11	2001:504:77::11	10 gbps

IXP Info Pages



NYIIX New York

Go to PeeringDB page

Go to IXP-DB page

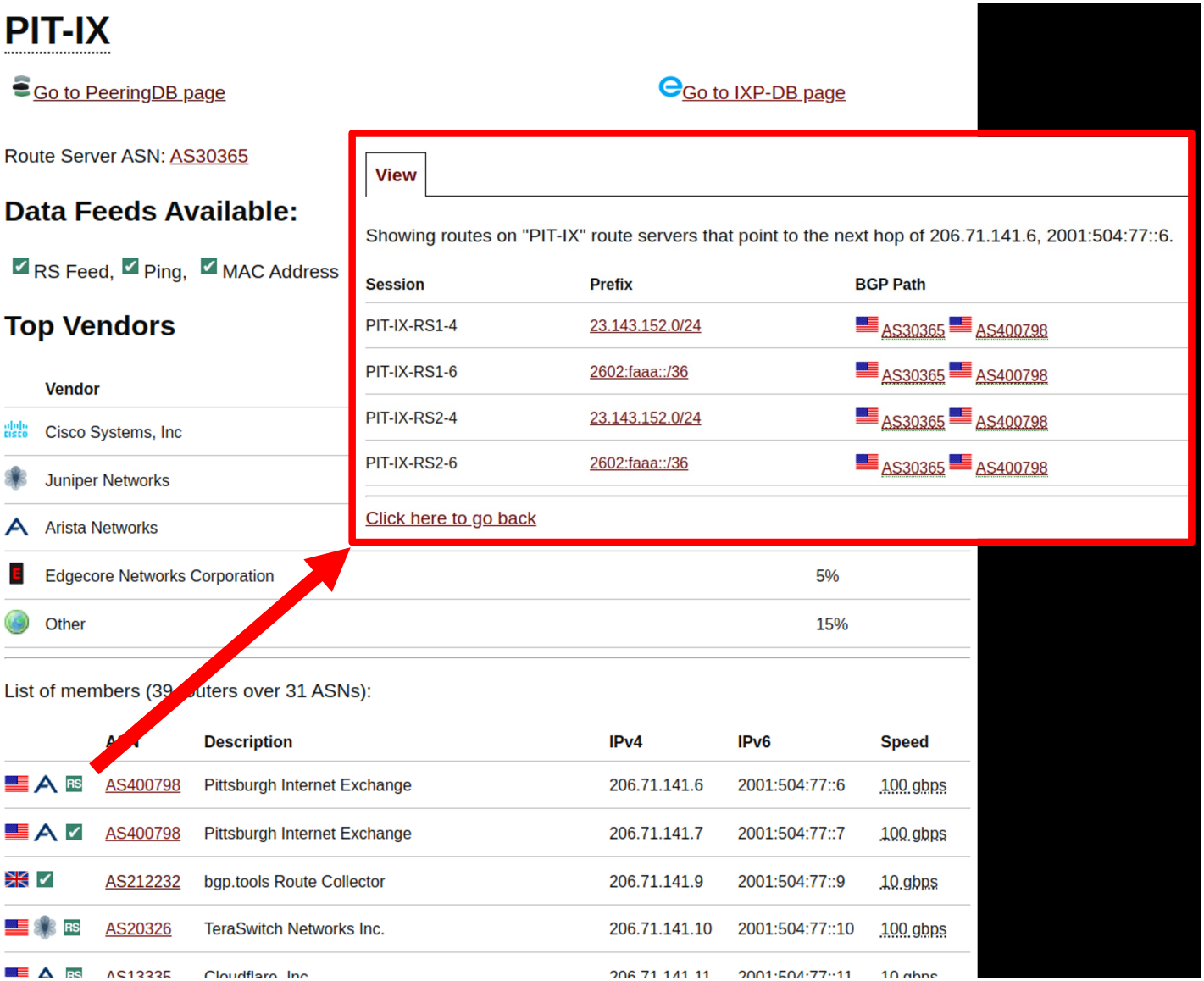
Data Feeds Available:

RS Feed, Ping, MAC Address

Do you run this IX and want to help with feeds? Contact Us!

List of members (236 routers over 211 ASNs):

ASN	Description	IPv4	IPv6
AS45437	Real World - The Core	198.32.161.115	2001:504:1::a504
AS52772	SJNET TELECOMUNICACOES - EIRELI	198.32.161.89	2001:504:1::a526
AS53180	INFORTEL COMUNICACOES LTDA	198.32.161.51	2001:504:1::a505
AS41327	Fiber Telecom S.p.A.	198.32.161.50	2001:504:1::a504
AS53667	FranTech Solutions	198.32.161.45	2001:504:1::a505
AS1031	Peer 1 Internet Service LLC	198.32.161.44	2001:504:1::a500
AS271253	LINK BRASIL TELECOMUNICACOES LTDA	198.32.161.43	2001:504:1::a527
AS2734	CoreSite	198.32.161.41	2001:504:1::a500



PIT-IX

Go to PeeringDB page

Go to IXP-DB page

Route Server ASN: AS30365

Data Feeds Available:

RS Feed, Ping, MAC Address

Top Vendors

Vendor	
Cisco Systems, Inc	
Juniper Networks	
Arista Networks	
Edgecore Networks Corporation	5%
Other	15%

List of members (39 routers over 31 ASNs):

ASN	Description	IPv4	IPv6	Speed
AS400798	Pittsburgh Internet Exchange	206.71.141.6	2001:504:77::6	100 gbps
AS400798	Pittsburgh Internet Exchange	206.71.141.7	2001:504:77::7	100 gbps
AS212232	bgp.tools Route Collector	206.71.141.9	2001:504:77::9	10 gbps
AS20326	TeraSwitch Networks Inc.	206.71.141.10	2001:504:77::10	100 gbps
AS13335	Cloudflare, Inc.	206.71.141.11	2001:504:77::11	10 gbps

View

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

Session	Prefix	BGP Path
PIT-IX-RS1-4	23.143.152.0/24	AS30365 AS400798
PIT-IX-RS1-6	2602:faaa::/36	AS30365 AS400798
PIT-IX-RS2-4	23.143.152.0/24	AS30365 AS400798
PIT-IX-RS2-6	2602:faaa::/36	AS30365 AS400798

Click here to go back

Traceroutes/Looking Glass/Agents

Orange S.A.

AS Number **5511**

BGP

Select BGP Session to query:

London [IPv4] [IPv6]

Input Prefix:

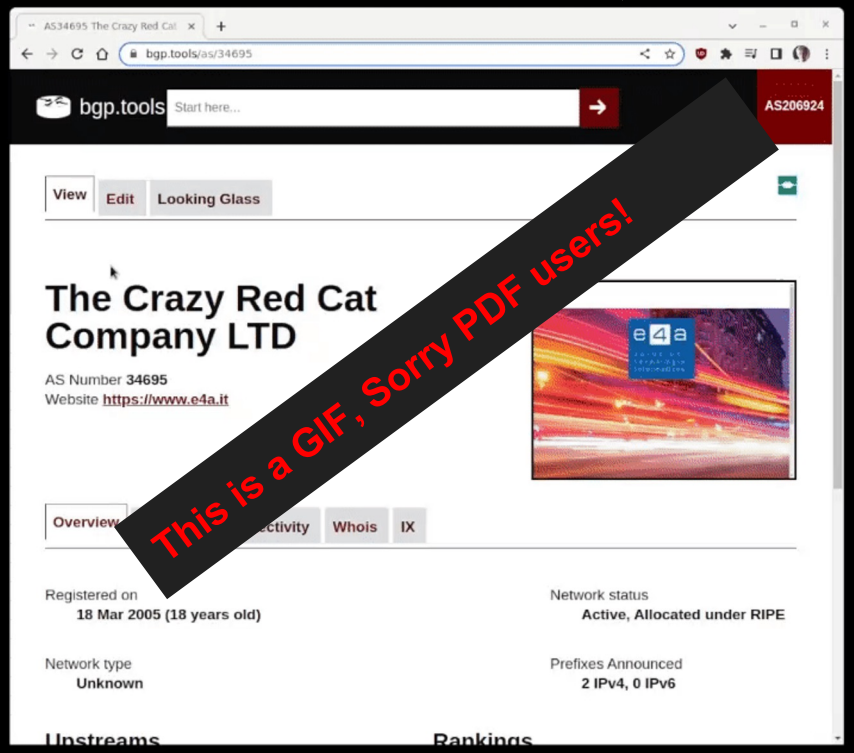
80.80.80.80

-

Query

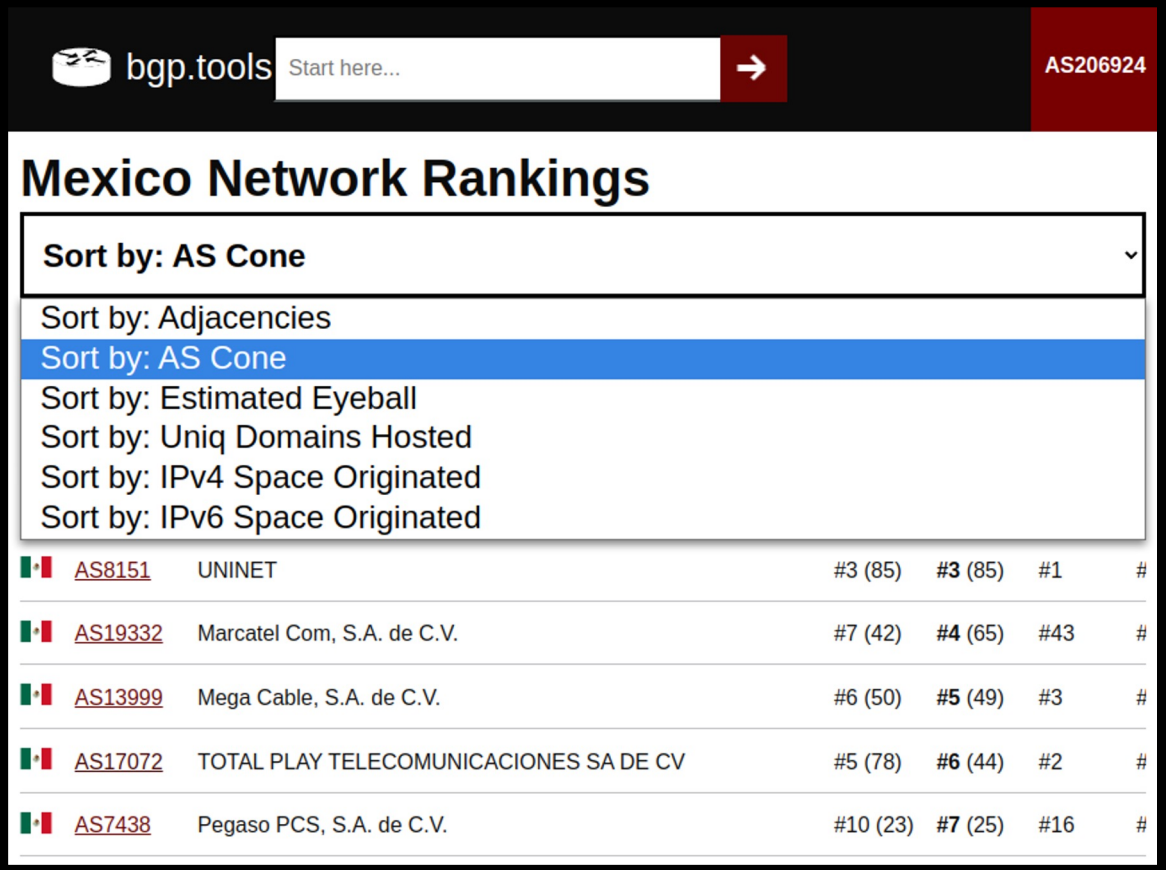
```
80.80.80.0/24        unicast [London 0000-00-00] * (?/-) [AS60679]
 Type: BGP
 BGP.as_path: 5511 3356 30247 60679
 BGP.community: [AS5511: United Kingdom] [AS5511: Route received from peering partner]
[AS5511: Route received in Europe from peering] [AS5511: TUNE announce to US peers]
```

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



Network Ranking

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



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

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

```
# 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

DIY?

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

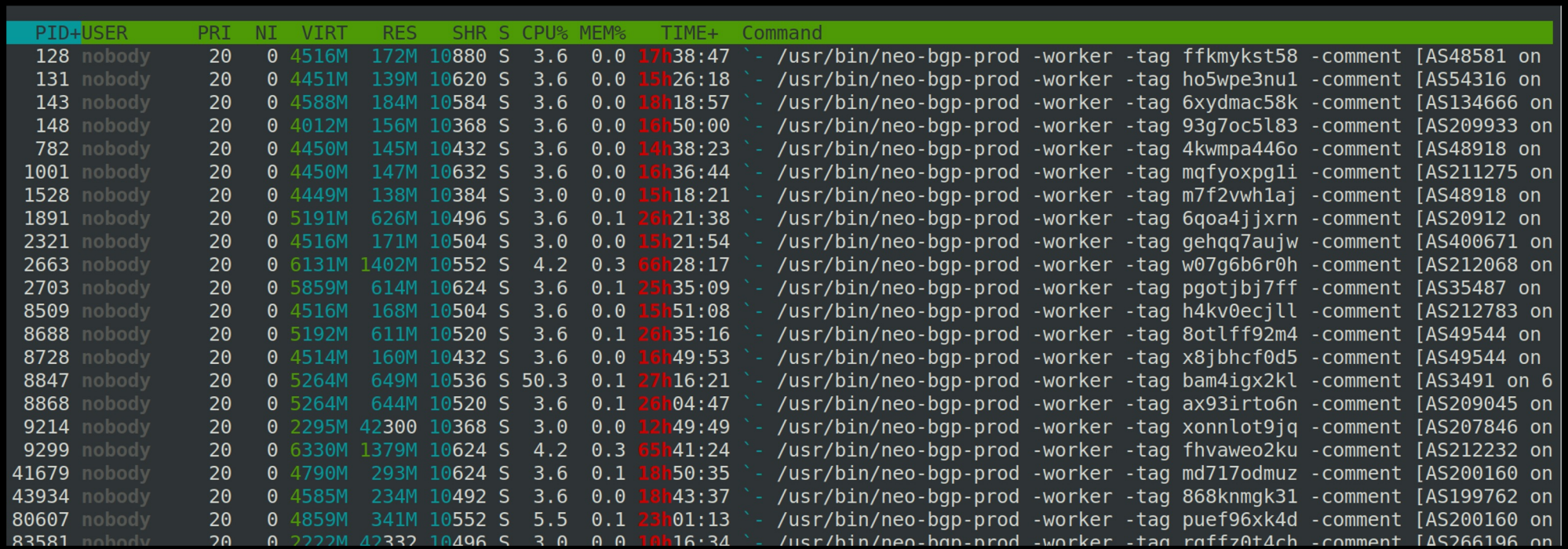
- Table dumps
  - Only show up every 4 - 8 hours
  - Make it hard to remove individual sessions that are known to be bad
- Message dumps (never used by bgp.tools production in the end)
  - If a message file never shows up, you have to wait until the next dump file (4 - 8 hours) before becoming reliability 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"

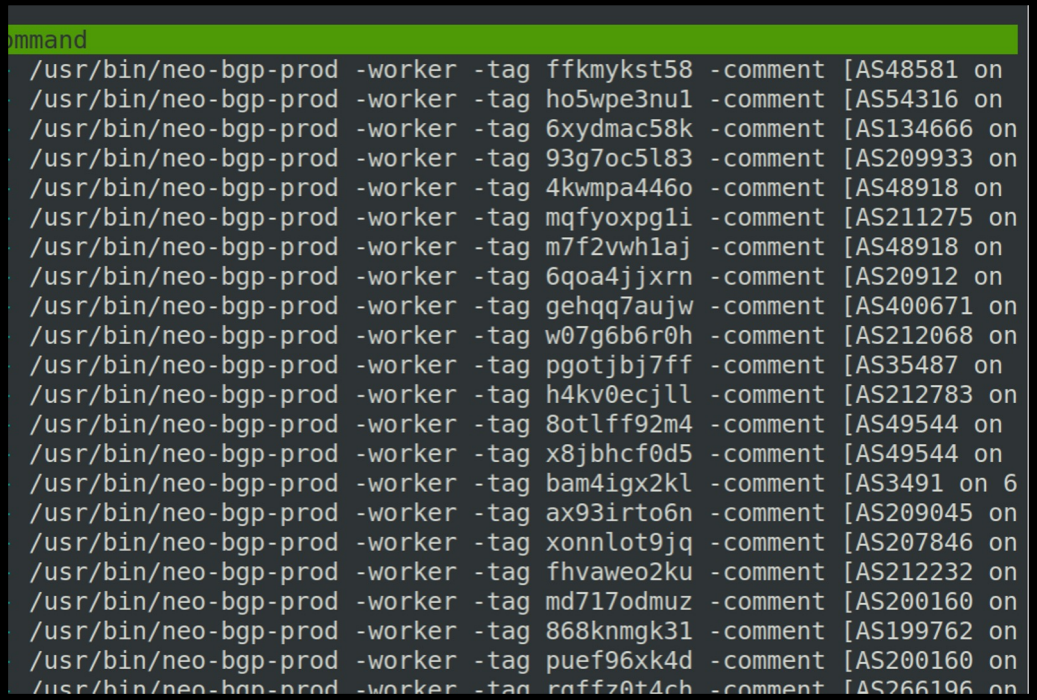


```
  PID+USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
  128 nobody     20   0 4516M  172M 10880 S  3.6  0.0 17h38:47 `- /usr/bin/neo-bgp-prod -worker -tag ffkmykst58 -comment [AS48581 on
  131 nobody     20   0 4451M  139M 10620 S  3.6  0.0 15h26:18 `- /usr/bin/neo-bgp-prod -worker -tag ho5wpe3nu1 -comment [AS54316 on
  143 nobody     20   0 4588M  184M 10584 S  3.6  0.0 18h18:57 `- /usr/bin/neo-bgp-prod -worker -tag 6xydmac58k -comment [AS134666 on
  148 nobody     20   0 4012M  156M 10368 S  3.6  0.0 16h50:00 `- /usr/bin/neo-bgp-prod -worker -tag 93g7oc5l83 -comment [AS209933 on
  782 nobody     20   0 4450M  145M 10432 S  3.6  0.0 14h38:23 `- /usr/bin/neo-bgp-prod -worker -tag 4kwmpa446o -comment [AS48918 on
 1001 nobody     20   0 4450M  147M 10632 S  3.6  0.0 16h36:44 `- /usr/bin/neo-bgp-prod -worker -tag mqfyoxpg1i -comment [AS211275 on
 1528 nobody     20   0 4449M  138M 10384 S  3.0  0.0 15h18:21 `- /usr/bin/neo-bgp-prod -worker -tag m7f2vwh1aj -comment [AS48918 on
 1891 nobody     20   0 5191M  626M 10496 S  3.6  0.1 26h21:38 `- /usr/bin/neo-bgp-prod -worker -tag 6qoa4jjxrn -comment [AS20912 on
 2321 nobody     20   0 4516M  171M 10504 S  3.0  0.0 15h21:54 `- /usr/bin/neo-bgp-prod -worker -tag gehqq7aujw -comment [AS400671 on
 2663 nobody     20   0 6131M 1402M 10552 S  4.2  0.3 66h28:17 `- /usr/bin/neo-bgp-prod -worker -tag w07g6b6r0h -comment [AS212068 on
 2703 nobody     20   0 5859M  614M 10624 S  3.6  0.1 25h35:09 `- /usr/bin/neo-bgp-prod -worker -tag pgotjbj7ff -comment [AS35487 on
 8509 nobody     20   0 4516M  168M 10504 S  3.6  0.0 15h51:08 `- /usr/bin/neo-bgp-prod -worker -tag h4kv0ecjll -comment [AS212783 on
 8688 nobody     20   0 5192M  611M 10520 S  3.6  0.1 26h35:16 `- /usr/bin/neo-bgp-prod -worker -tag 8otlff92m4 -comment [AS49544 on
 8728 nobody     20   0 4514M  160M 10432 S  3.6  0.0 16h49:53 `- /usr/bin/neo-bgp-prod -worker -tag x8jbhcf0d5 -comment [AS49544 on
 8847 nobody     20   0 5264M  649M 10536 S 50.3  0.1 27h16:21 `- /usr/bin/neo-bgp-prod -worker -tag bam4igx2kl -comment [AS3491 on 6
 8868 nobody     20   0 5264M  644M 10520 S  3.6  0.1 26h04:47 `- /usr/bin/neo-bgp-prod -worker -tag ax93irto6n -comment [AS209045 on
 9214 nobody     20   0 2295M 42300 10368 S  3.0  0.0 12h49:49 `- /usr/bin/neo-bgp-prod -worker -tag xonnlot9jq -comment [AS207846 on
 9299 nobody     20   0 6330M 1379M 10624 S  4.2  0.3 65h41:24 `- /usr/bin/neo-bgp-prod -worker -tag fhvaweo2ku -comment [AS212232 on
41679 nobody     20   0 4790M  293M 10624 S  3.6  0.1 18h50:35 `- /usr/bin/neo-bgp-prod -worker -tag md717odmuz -comment [AS200160 on
43934 nobody     20   0 4585M  234M 10492 S  3.6  0.0 18h43:37 `- /usr/bin/neo-bgp-prod -worker -tag 868knmgk31 -comment [AS199762 on
80607 nobody     20   0 4859M  341M 10552 S  5.5  0.1 23h01:13 `- /usr/bin/neo-bgp-prod -worker -tag puef96xk4d -comment [AS200160 on
83581 nobody     20   0 2222M 42332 10496 S  3.0  0.0 10h16:34 `- /usr/bin/neo-bgp-prod -worker -tag rqffz0t4ch -comment [AS266196 on
```

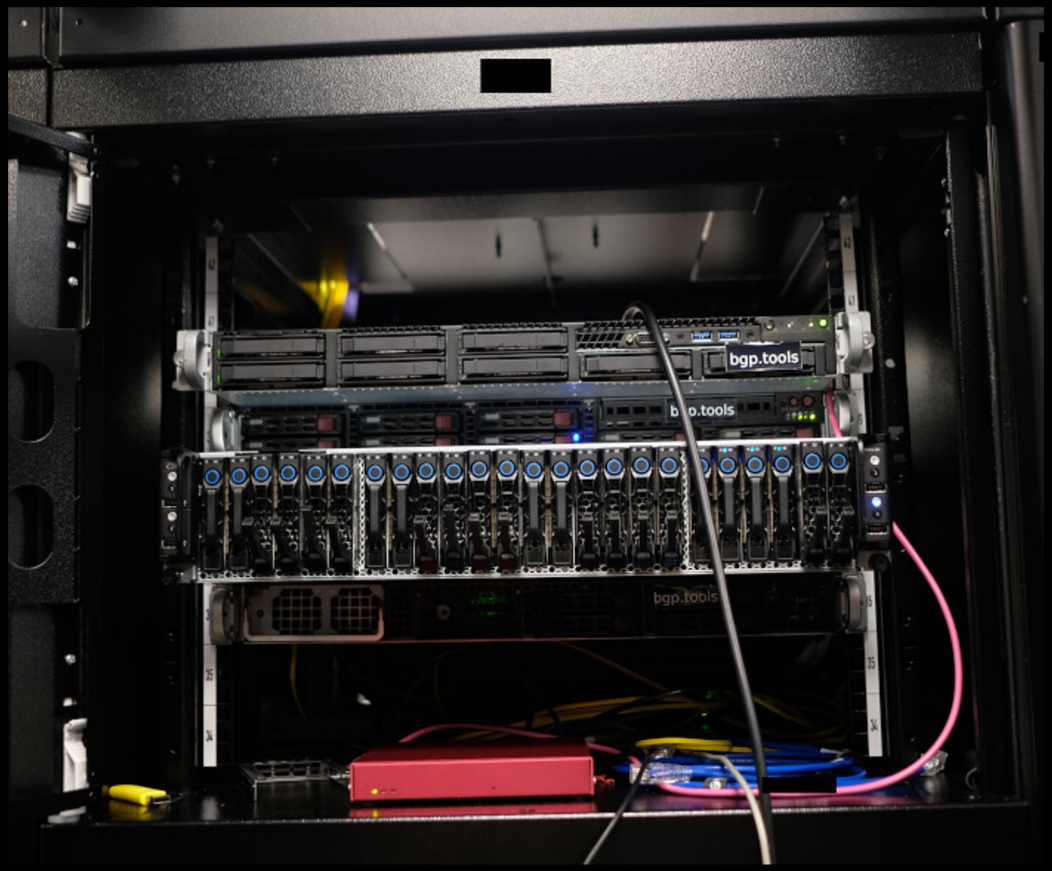
A purpose built "bgpd" for the exact use case that bgp.tools wants

"neo-bgp"

- Each BGP Session is in it's 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



"neo-bgp"



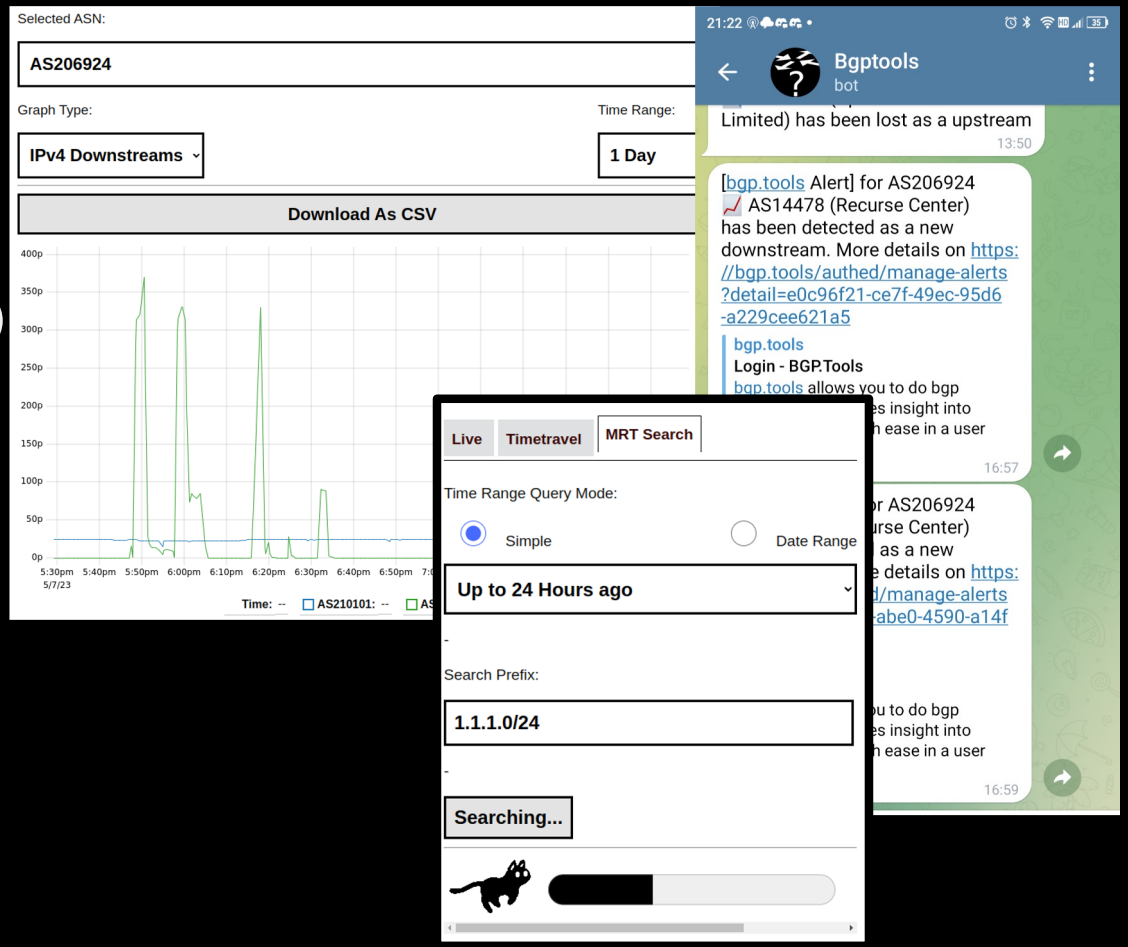
- Bgp.tools currently sits at 1360~ BGP sessions
  - 750~ full IPv4 tables
  - 1210~ full IPv6 tables
  - 1,000,000,000+ BGP Paths stored in RAM
- Hardware is modest, entire site operates inside ¼ cab with room to spare
  - 512G DDR4 per machine
  - 32~ cores per machine
  - 3 active machines right now, 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

- [x] Getting low latency and accurate BGP data to use
- [x] 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

- [x] Getting low latency and accurate BGP data to use
- [x] Building a scalable system to avoid being picky on feeds
- [ ] Collecting relevant data
- [x] 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

Name	Physical Location	Type	Scope	Raw Data
RRC00	Amsterdam, NL	multihop	global	data
RRC01	London, GB	IXP	LINX, LONAP	data
RRC03	Amsterdam, NL	IXP	AMS-IX, NL-IX	data
RRC04	Geneva, CH	IXP	CIXP	data
RRC05	Vienna, AT	IXP	VIXP	data
RRC06	Otemachi, JP	IXP	DIX-IE	data
RRC07	Stockholm, SE	IXP	Netnod	data
RRC10	Milan, IT	IXP	MIX	data
RRC11	New York, NY, US	IXP	NYIIX	data
RRC12	Frankfurt, DE	IXP	DE-CIX	data
RRC13	Moscow, RU	IXP	MSK-IX	data
RRC14	Palo Alto, CA, US	IXP	PAIX	data
RRC15	Sao Paolo, BR	IXP	PTTMetro-SP	data
RRC16	Miami, FL, US	IXP	Equinix Miami	data
RRC18	Barcelona, ES	IXP	CATNIX	data
RRC19	Johannesburg, ZA	IXP	NAP Africa JB	data

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

  - 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

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 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, Frys-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 a IXP not listed here?
admin@bgp.tools 🙏

Challenges running bgp.tools

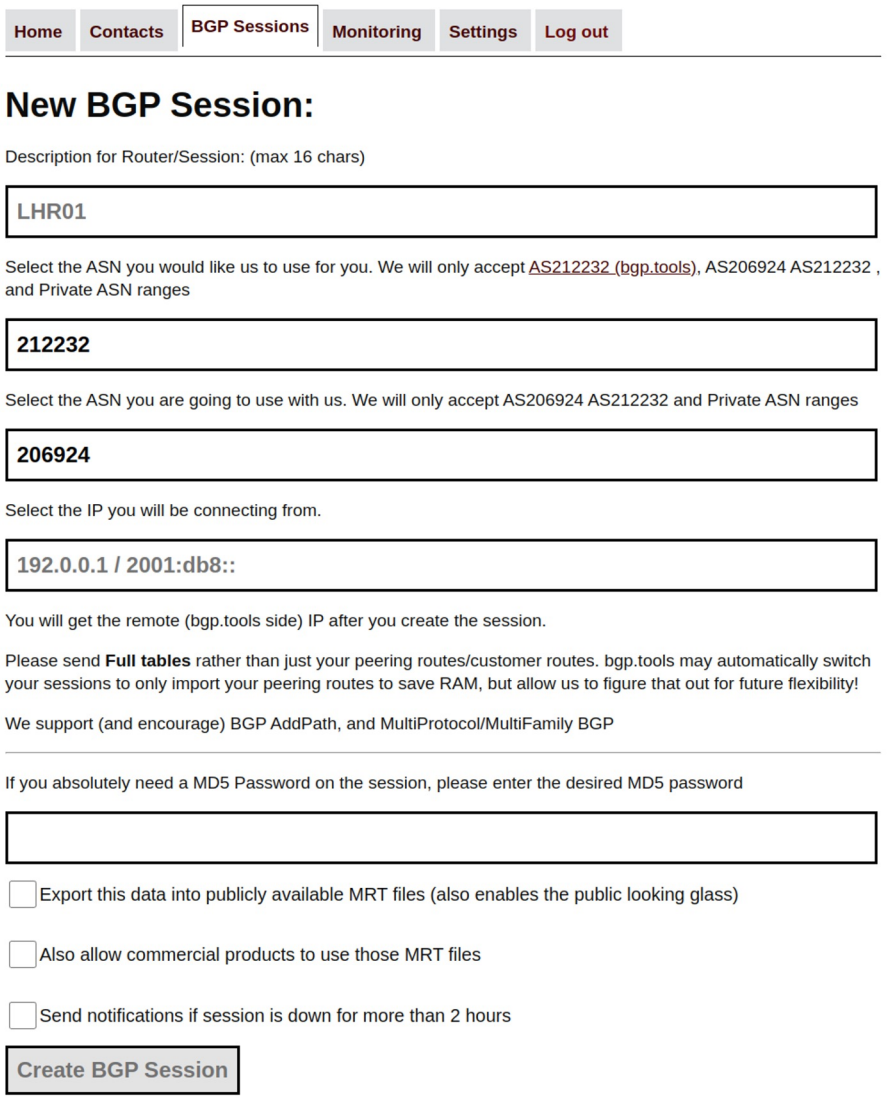
- ✅ Getting low latency and accurate BGP data to use
- ✅ Building a scalable system to avoid being picky on feeds
- ✅ Collecting relevant data
- ✅ Not going bankrupt

Setting up feeds is easy

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

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 to bottom link "Contribute Data"

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

EOF