

# Playing with (world wide) routing ...for fun (and profit)

**Łukasz Bromirski**

lukasz@bromirski.net

<https://lukasz.bromirski.net>



@LukaszBromirski

# Why we're doing this at all?

- I'm networking geek, living in Europe (Poland)
- CCIE #15929 (R&S/SP) & CCDE #2012::17
- I do like playing with things and ensuring that people have access to knowledge and interesting tools
  - take a look at my blog or prezos (usually in Polish)
  - I created Cisco FAQ PL (unofficial one), BGP Blackholing PL, I'm one of the PLNOG founders
- Doing L3, L4, and currently also firewalls, IPSes and all that fluffy stuff (I'm PM, Engineering@Cisco CNS)
  - ...bootcamps for CCIE SP, network design/architecture (CCDE) and Cisco SDN tools as well



So, BGP in the lab, right?

# Net::BGP, bgpsimple, RIPE feeds and Quagga\*

- Back in 2010 it was hard to get BGP feed „just like that”, so doing RIR network dumps and then feeding them internally was quickest option available (typically)

Virtual BGP open source implementations were neither very robust nor quick, but if everything was done correctly, you could end up with something like this:

- Kevin Myers came up with VM automating 500k BGP entries feed for your lab in 2016\*\*
- This is still fine today, as open implementations matured and can be used even in non-trivial deployments (and there was always BIRD ☺)

```
c1941w-lab#sh ip bgp summary
BGP router identifier 192.168.110.10, local AS number 65100
BGP table version is 11665410, main routing table version 11665410
344999 network entries using 46919864 bytes of memory
344999 path entries using 17939948 bytes of memory
67835/64064 BGP path/bestpath attribute entries using 8411540 bytes of memory
62048 BGP AS-PATH entries using 2683072 bytes of memory
1773 BGP community entries using 80882 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
3970 BGP filter-list cache entries using 47640 bytes of memory
BGP using 76082946 total bytes of memory
BGP activity 1643781/1298782 prefixes, 1644641/1299642 paths, scan interval 60 secs

Neighbor      V      AS MsgRcvd MsgSent  TblVer  InQ OutQ Up/Down  State/PfxRcd
192.168.110.1  4      65000 3252410      4 11665410      0    0 00:31:04   344999
```

\* <https://lukasz.bromirski.net/post/bgp-w-labie/>

\*\* <https://stubarea51.net/2016/01/21/put-500000-bgp-routes-in-your-lab-network-download-this-vm-and-become-your-own-upstream-bgp-isp-for-testing/>

# Some triggers for the idea and today's talk

Blažej Krajňák

3 August 2020 at 22:15

BK

BGP full feed for testing purposes

To: NANOG

Hello,

I'm wondering, if there is any public service I can get full BGP feed from for testing purposes.

I admin multi-homed AS50242 with two default routes for now (fail-over). I'm going to prepare new routing setup with extended validation so really full BGP feed would be useful. Yes, I can ask my upstream provider for it, but I don't want to change settings in production setup.

Thanks

Regards,  
Blažej Krajňák



Daniel Dib

@danieldibsw

...

Interesting number I heard today. It takes around 2h to "converge" if you setup a new router in the default-free zone. Lots of BGP updates...

1:37 PM · Sep 1, 2020 · Twitter Web App

2 Retweets 14 Likes



Łukasz Bromirski @LukaszBromirski · Sep 2

...

Replying to @danieldibsw

That number is completely bogus. I recently helped to setup new AS with new IPv4 prefix (well, both were "recovered" from different places), and newly announced prefix was visible and reachable from distant US, Australian and NZ networks within ~15 minutes.



# I had spare 12 minutes before next webex call...



Łukasz Bromirski  
@LukaszBromirski

...

If you need full BGP feed for your lab - you can have it right away. Just read, configure and be happy: [lukasz.bromirski.net/post/bgp-w-lab...](https://lukasz.bromirski.net/post/bgp-w-lab...) Any comments/feedback more than appreciated. #CCIE #CCIESP #BGP #BGPGECKS



## bgp in the lab #2

recent thread on nanog@ list got me back to old project that i was thinking about long time ago. and here it is - i just ...  
[lukasz.bromirski.net](https://lukasz.bromirski.net)

8:59 PM · Aug 5, 2020 · Twitter Web App

## BGP in the lab - v4 & v6 live feeds from Europe

Łukasz Bromirski | Wed, 07 Oct 2020 13:41:16 -0700

Dear NANOGers,

If you're looking for live, full BGP v4 & v6 feed for your lab or a bit of testing before going live, I just shared a short post on how to get it:

<https://lukasz.bromirski.net/post/bgp-w-labie-3/>

Happy BGPing,

--

Łukasz Bromirski

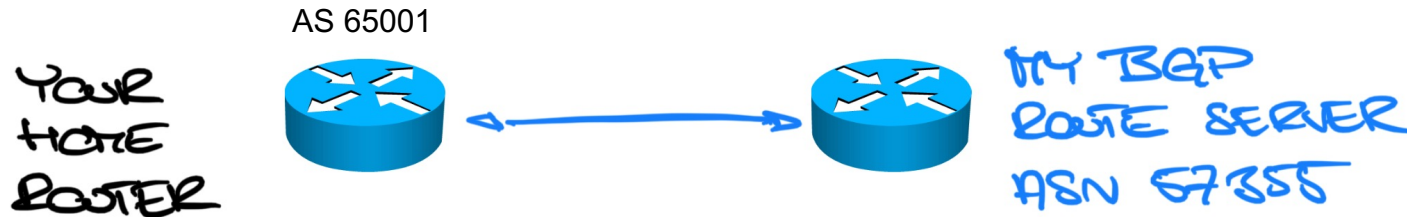
CCIE R&S/SP #15929, CCDE #2012::17, PGP Key ID: 0xFD077F6A

# Let's setup IPv4 and IPv6 "feeder" – route server

- Simplistic setup – I have static IP, you may have dynamic one

you'll come from ASN 65001

I accept nothing, You do whatever you want



# Let's setup IPv4 and IPv6 "feeder" – route server

- Using "Dynamic BGP neighbor" feature of IOS-XE
- Setup on CSR 1000v
- I get feed from two redundant upstream routers
- I also filter whatever comes in from you, so don't send me anything... please

```
!  
router bgp 57355  
  bgp router-id 85.232.240.179  
  bgp asnotation dot  
  bgp log-neighbor-changes  
  BGP Dynamic Peer definition bgp listen range ::/0 peer-group BGP-FF-V6  
  bgp listen limit 100  
  no bgp default ipv4-unicast  
  Peer group definition that will be used below neighbor BGP-FF-V6 peer-group  
  neighbor BGP-FF-V6 remote-as 65001  
  neighbor BGP-FF-V6 ebgp-multihop 255  
  neighbor BGP-FF-V6 version 4  
  neighbor BGP-FF-V6 timers 3600 7200  
  neighbor 2001:[#1 upstream] remote-as 57355  
  neighbor 2001:[#2 upstream] remote-as 57355  
  !  
  address-family ipv4  
  exit-address-family  
  !  
  address-family ipv6  
  Filter RIB->FIB table-map TABLE-SRD filter  
  All peers activated for IPv6 will have this applied neighbor BGP-FF-V6 activate  
  neighbor BGP-FF-V6 send-community both  
  neighbor BGP-FF-V6 remove-private-as  
  neighbor BGP-FF-V6 prefix-list DENY-ALL-V6 in  
  neighbor 2001:[#1 upstream] activate  
  neighbor 2001:[#2 upstream] activate  
  !
```



Ideas – get this in the lab

# Observations – user types and maximums

- Three types of users:

“Good friends”: session on both AFs - IPv4 and IPv6 – staying up for weeks

“Testers”: come, get the feed, reset session couple of times and then disappear forever

“Undecided”: come up from time to time

- “Project” maximums:

74 peers up on IPv4 at the same time

16 peers up on IPv6 at the same time

```
bgp-ff-atman-v6#sh bgp ipv6 unicast summary
BGP router identifier 85.232.240.179, local AS number 57355
```

[...]

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
*2001:470		65001	93	206932	118986159	0	0	2d04h	0
*2001:67C		65001	1121	91539	118986139	0	0	21:42:11	0
2001:1A68		57355	9172901	26752	118986206	0	0	2w2d	43378
2001:1A68		57355	9170815	26747	118986206	0	0	2w2d	43378

\* Dynamically created based on a listen range command  
Dynamically created neighbors: 2, Subnet ranges: 1

```
bgp-ff-atman-v4#sh bgp ipv4 unicast summary
BGP router identifier 85.232.240.179, local AS number 57355
```

[...]

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
46.2	4	57355	9571237	26752	110892405	0	0	2w2d	814746
46.2	4	57355	9564005	26742	110892405	0	0	2w2d	814746
*78.	4	65001	83609	10929705	110892275	0	0	8w2d	0
*78.	4	65001	436	2111019	110892283	0	0	1w3d	0
*87.	4	65001	1114	353830	110892105	0	0	21:33:14	0
*89.	4	65001	91	602125	110892381	0	0	2d04h	0
*99.	4	65001	315	1831015	110892088	0	0	1w1d	0
*103	4	65001	6087284	11168403	110892307	0	0	8w5d	0
*178	4	65001	3836	8816869	110892399	0	0	6w5d	0
*193	4	65001	2536	11439478	110892373	0	0	9w0d	0
*194	4	65001	1964	8747495	110892283	0	0	6w4d	0
*203	4	65001	150	835823	110892222	0	0	3d15h	0

\* Dynamically created based on a listen range command  
Dynamically created neighbors: 10, Subnet ranges: 1

```
bgp-ff-atman-v4#
```

# Observations - typical problems

- I refuse to RTFM (BGP local-as) – ASN 65179 seems to be typical (?)

```
%BGP-3-NOTIFICATION: sent to neighbor *2600:1F16:[...] passive 2/2  
(peer in wrong AS) 2 bytes FE9B
```

- “I’ll send you what I have”

```
For address family: IPv6 Unicast  
Session: *2001:470:[...]  
[...]
```

	Outbound	Inbound
Local Policy Denied Prefixes:	-----	-----
Total:	0	<b>91726</b>

- Please help me help you – if you have problems, let me know at  
lukasz @ bromirski.net

Ideas – home RPKI lab

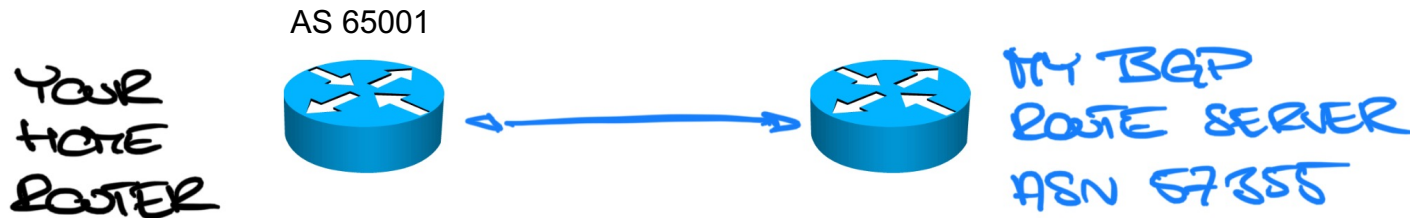
# RPKI?

ROA
172.25.0.0/16-24
12343

- BGP prefixes can be hijacked by third parties, by injecting more specifics or playing with paths
- SDR tries to signal (control plane mechanism only) which prefixes are „authorized“ – with Origin AS and prefix along with prefix length
- ROA objects are kept by RIRs and cache can be downloaded to your own daemon – multiple of them available:
- BGP router can then use two tables to compare BGP prefixes with their „validation“ status
  - VALID – present as ROA object and valid in BGP table
  - INVALID – present as ROA object, but BGP table has invalid info (one of three attributes)
  - NOT FOUND – there's no instance of ROA for matching BGP prefix
- That „status“ validation can then make you influence decisions for forwarding – for example lower Local Preference, or drop prefix info

# RPKI validation at home

- You get feed and can do route validation
- Normal, production sites should do that anyway, but if you're in lab... it's again hard to get the "real" view locally
- You can then add yourself local RPKI validation daemon and you'll know which prefixes are valid, invalid or not-present
- Limited use case for end AS with "home" internet, but may be useful with multiple internet links when only some of the ISPs are advertising invalid entries



Ideas – egress traffic engineering

# Egress traffic load balancing

## Multiple ISP links (1/3)

- Default route to all three links, and rely on per-destination load balancing

```
ip route 0.0.0.0 0.0.0.0 100.64.0.2
ip route 0.0.0.0 0.0.0.0 169.254.0.2
ip route 0.0.0.0 0.0.0.0 195.81.100.2
```

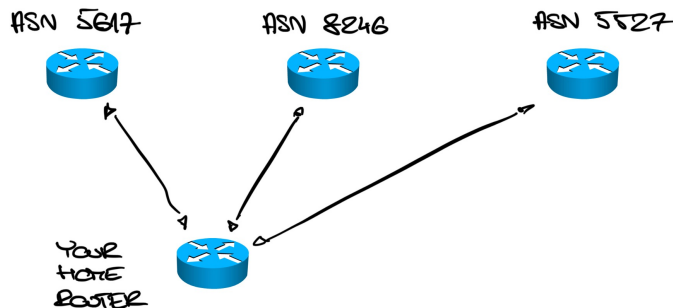
- Do some CEF tricks:

```
rtr-edge(config)#ip cef load-sharing algorithm include-ports ?
destination Use destination port in hash function
source       Use source port in hash function
```

- Default route + more specific, static routes (inconvenient) or download each N hours/days latest per-ASN prefixes and then build static routes from them, uploading them to router (automation challenge)

```
$ bgpq3 -F "ip route %n/%l 100.64.100.2\\n" as8614
ip route 193.231.172.0/24 100.64.100.2
ip route 193.239.64.0/24 100.64.100.2
ip route 193.239.65.0/24 100.64.100.2
ip route 193.239.66.0/24 100.64.100.2
ip route 193.239.67.0/24 100.64.100.2
ip route 217.156.124.0/24 100.64.100.2
```

```
$ whois -h whois.radb.net - "-K -i origin AS8614"
```

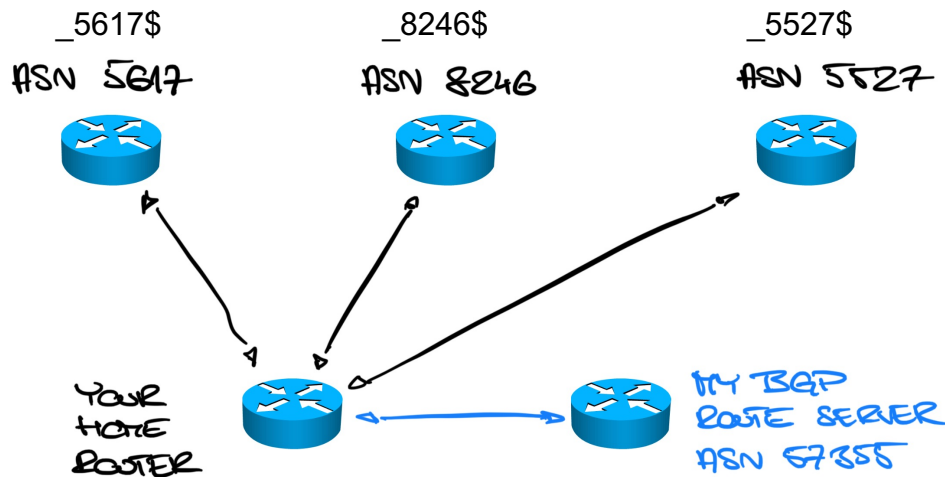




# Egress traffic load balancing

## Multiple ISP links (2/3)

- SDN Unicorn magic
- BGP feed from me



# Egress traffic load balancing

## Multiple ISP links (3/3)

- You accept all or partial table from my feed
- You modify next-hops/interfaces based on the AS\_PATH attribute

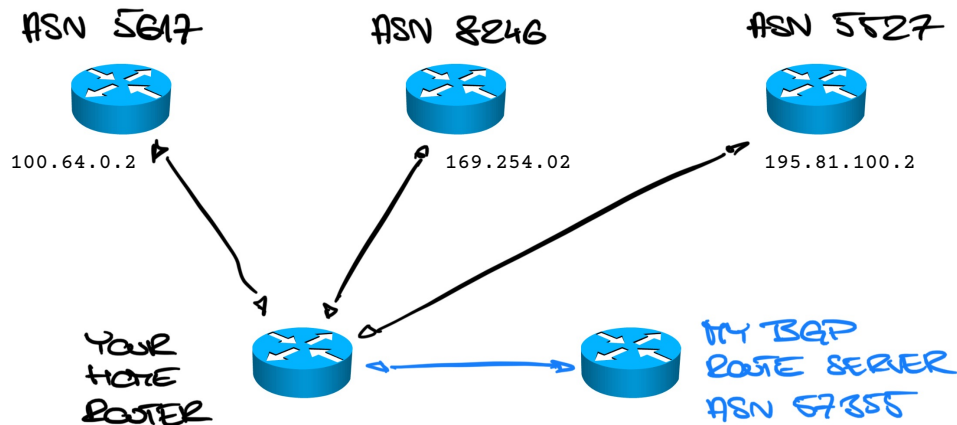
```
route-map BGP-TE permit 10
  description TE towards 5617
  match as-path 101
  set ip next-hop 100.64.0.2

route-map BGP-TE permit 20
  description TE towards 8246
  match as-path 102
  set ip next-hop 169.254.0.2

route-map BGP-TE permit 10
  description TE towards 5527
  match as-path 103
  set ip next-hop 195.81.100.2

ip as-path access-list 101 permit _5617$
ip as-path access-list 102 permit _8246$
ip as-path access-list 103 permit _5527$

router bgp 65001
  address-family ipv4 unicast
    neighbor 85.232.240.179 route-map BGP-TE in
```



- Your traffic will flow accordingly and should fail over if nexthop/interface is gone

Ideas – what else would you find useful?

# See more about BGP and routing

- Security by BGP 101 – distributed, BGP-based security system:  
[https://lukasz.bromirski.net/docs/prezos/certee2017/BGP\\_Security\\_101.pdf](https://lukasz.bromirski.net/docs/prezos/certee2017/BGP_Security_101.pdf)
- Scaling services out using IP anycast:  
[http://lukasz.bromirski.net/docs/prezos/plnog2011/ip\\_anycast.pdf](http://lukasz.bromirski.net/docs/prezos/plnog2011/ip_anycast.pdf)
- BGP in the lab #3 – IPv4 and IPv6 feeds for free:  
<https://lukasz.bromirski.net/post/bgp-w-labie-3/>
- My home network – using anycast and BGP for things like one DNS server, etc.:  
<https://lukasz.bromirski.net/post/moja-siec-domowa-2/>
- Daniel Stocker writeup about the project:  
[https://puffy.nolink.ch/posts/fullbgp\\_at\\_home/](https://puffy.nolink.ch/posts/fullbgp_at_home/)

Playing with (world wide) routing  
...for fun (and profit)

Q&A

**Łukasz Bromirski**

lukasz@bromirski.net

<https://lukasz.bromirski.net>



@LukaszBromirski