Zabawy z lokalnym i globalnym routingiem...

(dla zabawy, bezpieczeństwa, lepszej inżynierii ruchu i for profit też)

Łukasz Bromirski
lukasz@bromirski.net
https://lukasz.bromirski.net
@LukaszBromirski

PLNOG
where ipeople meet

„Chciałbym mieć BGP w domu”

(a mam tylko łącze „klienckie” i nikt BGP mi nie da)

Net::BGP, bgpsimple, feedy RIPE'a i Quagga*

- W 2010 ciężko było dostać feed BGP „ot tak", więc najlepszą opcją dla kogoś „bez znajomości" było przekształcenie dumpów globalnej tablicy od RIRa/ów i wstrzyknięcie ich lokalnie

  Implementacje open source nie były zbyt skalowalne czy szybkie, ale jeśli zrobiłeś wszystko porządnie – dawało się tego używać i wyglądało to mniej więcej tak:

- Kevin Myers zbudował całą VMkę automatyzującą 500k prefixów w BGP w 2016**
- Takie podejście jest nadal OK, a większość rozwiązań Open Source dojrzała (i jest 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/

“Bo mnie sprowokowali”

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 reall full BGP feed
would be usefull. 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
@danieldibswe

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

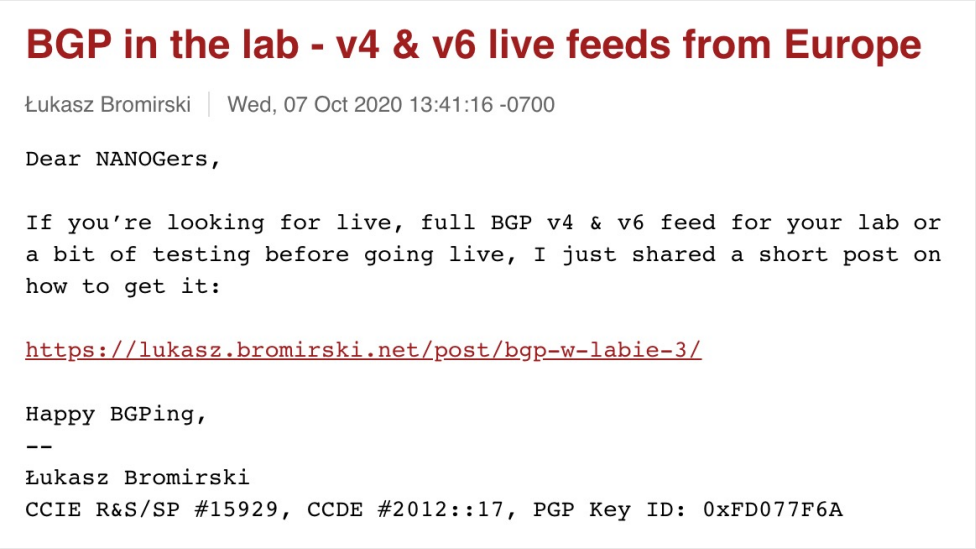
2 Retweets **14** Likes

Łukasz Bromirski @LukaszBromirski · Sep 2

Replying to @danieldibswe

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.

A ponieważ wszyscy i tak siedzieli na webexach...



Jak to wygląda w praktyce?

- Najprostsza możliwa konfiguracja

  Ja mam statyczne IP, ty możesz mieć statyczne albo dynamiczne

  Identyfikujesz się jako ASN 65001

  Nie akceptuje niczego, ale Ty możesz sobie robić co chcesz ☺

Konfiguracja po **mojej** stronie

Cisco IOS-XE

- Używamy funkcjonalności “dynamiczni sąsiedzi” w IOS-XE
- Całość działa w oparciu o Cisco CSR 1000v
- Mam parę feedów z Polski
- Nie wysyłaj mi żadnych prefiksów – i tak je odrzucę, ale po co mamy sobie utrudniać?

Definicja dynamicznych sąsiadów BGP

Definicja grupy sąsiadów, której użyjemy poniżej

Filtr RIB->FIB

Każdy sąsiad IPv6 otrzyma politykę BGP-FF-V6

```
!
router bgp 57355
 bgp router-id 192.0.2.101
 bgp asnotation dot
 bgp log-neighbor-changes
 bgp listen range ::/0 peer-group BGP-FF-V6
 bgp listen limit 100
 no bgp default ipv4-unicast
 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
  table-map TABLE-SRD filter
  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
!
```

Obserwacje po ponad roku działania

Typy użytkowników i wartości maksymalne

- Trzy rodzaje użytkowników:
  - “Dobrzy znajomi”: sesje po IPv4 i IPv6 utrzymujące się miesiącami
  - “Testerzy”: wpadają, zestawiają sesję, resetują ją parenaście razy i znikają w otchłaniach Internetu (“niefajne”)
  - “Powracający”: zestawią coś czasem ale zrywają i znikają czasem na całe tygodnie
- Maksymalne wartości historyczne:
  - 118 sąsiadów po IPv4
  - 38 sąsiadów po IPv6

```
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[redacted]
                            65001      93  206932 118986159    0    0 2d04h           0
*2001:67C[redacted]
                            65001    1121   91539 118986139    0    0 21:42:11        0
2001:1A68[redacted]
                            57355 9172901   26752 118986206    0    0 2w2d        43378
2001:1A68[redacted]
                            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[redacted]4        57355 9571237   26752 110892405    0    0 2w2d       814746
46.2[redacted]4        57355 9564005   26742 110892405    0    0 2w2d       814746
*78.[redacted]4        65001   83609 10929705 110892275    0    0 8w2d            0
*78.[redacted]4        65001     436 2111019 110892283    0    0 1w3d            0
*87.[redacted]4        65001    1114  353830 110892105    0    0 21:33:14        0
*89.[redacted]4        65001      91  602125 110892381    0    0 2d04h           0
*99.[redacted]4        65001     315 1831015 110892088    0    0 1w1d            0
*103[redacted]
                4       65001 6087284 11168403 110892307    0    0 8w5d            0
*178[redacted]4        65001    3836 8816869 110892399    0    0 6w5d            0
*193[redacted]
                4       65001    2536 11439478 110892373    0    0 9w0d            0
*194[redacted]4        65001    1964 8747495 110892283    0    0 6w4d            0
*203[redacted]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#
```

Pozdrowienia dla...

- Cisco MIG Lab w RTP ☺ (Frederic Cuiller & Nicolas Fevrier)
- Ericsson Lab in Sweden ☺ (czemu tylko IPv4?)
- Stowarzyszenie Miłośników Internetu INDS, Częstochowa (yeah!)
- Wszystkich “testerów” z sieci Orange, T-Mobile i UPC (i upartego kolegę używającego karty Playa ;))

Typowe problemy

- Nie czytam instrukcji (BGP local-as) – ASN 65179 często używany (?)

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

- “Wyślę Ci to, co sam mam”

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

[...]
                                        Outbound    Inbound

  Local Policy Denied Prefixes:         --------    -------

    Total:                                     0      91726
```

- Jeśli masz problem, nie wiesz o co chodzi – po prostu napisz:

lukasz @ bromirski.net

Co z tym można zrobić?

Pomysł #1: Walidować prefiksy z RPKI

RPKI?

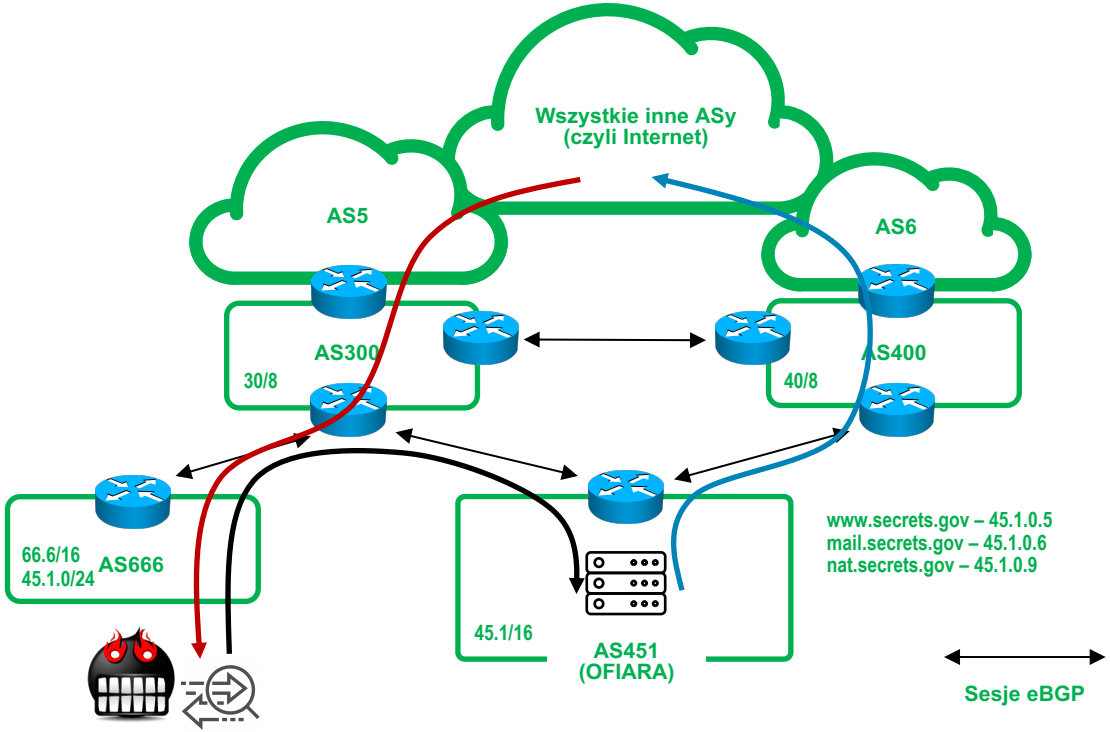
ROA
172.25.0.0/16-24
12343

- Prefiksy BGP mogą zostać „porwane” przez rozgłoszenie dokładniejszego prefiksu lub zabawę z AS_PATH
- SIDR pozwala wymienić (tylko control plane) cache obiektów zawierających ROA – podpisane cyfrowo prefiksy wraz z Origin AS
- Obiekty ROA utrzymywane są przez RIRów i mogą zostać pobrane do Twojego własnego serwera
- BGP na routerze może wtedy porównać dwie tablice – routingu BGP oraz cache ROA żeby oznaczyć każdy z prefiksów jako znajdujący się w jednym ze stanów:

  VALID – obecny jako obiekt ROA i zgodny z prefiksem w tablicy BGP
  INVALID – obecny jako obiekt ROA, ale prefiks BGP ma nieprawidłową zawartość
  NOT FOUND – nie ma obiektu ROA dla danego prefiksu BGP

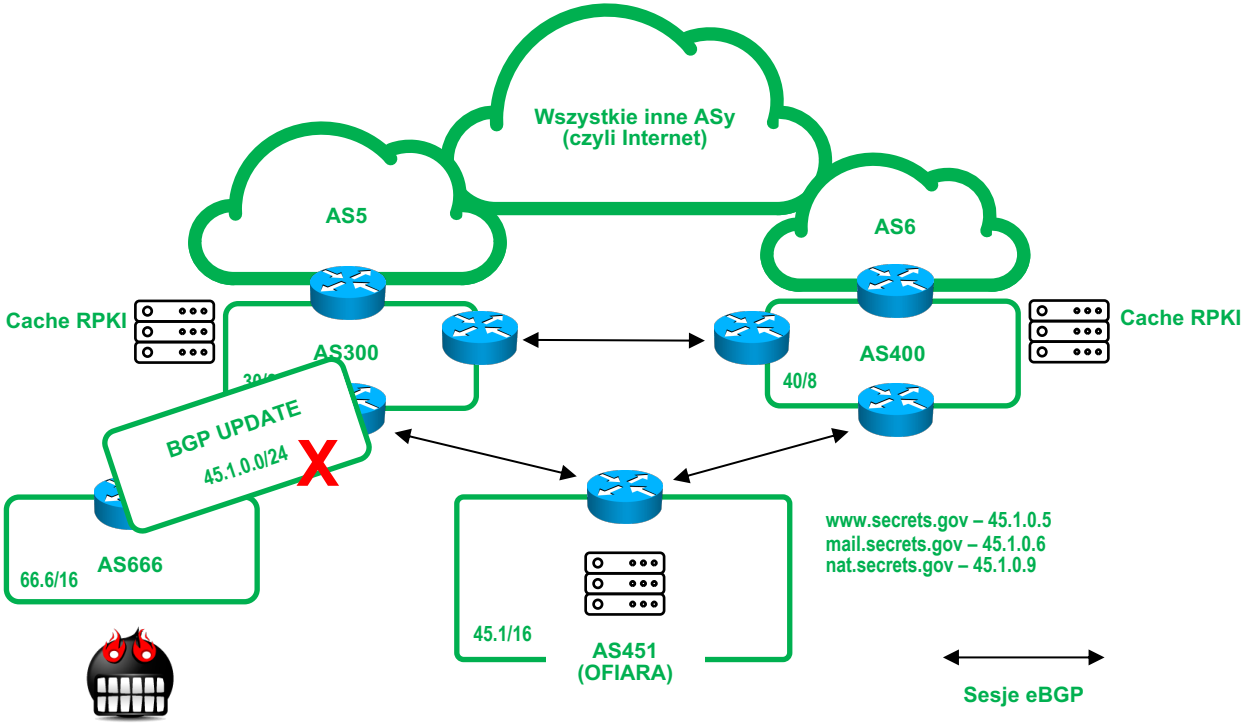
- Możesz wtedy podjąć decyzję, co chcesz zrobić z prefiksami na podstawie statusu walidacji – np. odrzucić w ogóle prefiksy INVALID

“Porwane”? “w BGP”? (”tygrys? w Afryce?”*)



* https://www.youtube.com/watch?v=oLdk2C25Z14

“Ale nie z RPKI!”



RPKI? Ale w domu?

- Po pobraniu pełnej tablicy BGP możesz przeprowadzać walidację u siebie na brzegu sieci – przydatne w szczególności do zabaw w labie
- Parę dostępnych demonów RPKI

  polecam NLNetLabs Routinator 3000*
- Walidacja i odrzucanie prefiksów INVALID ma szczególne znaczenie dla operatorów tranzytowych

  ...domowy użytkownik i tak JAKOŚ pewnie będzie chciał wysłać ruch, najczęściej przez trasę domyślną a zatem walidacja i tak wielkiej różnicy zwykle nie zrobi...

  ...chyba, że masz wiele łącz od wielu operatorów i tylko część podsyła Ci “porwane” prefiksy

* https://nlnetlabs.nl/projects/rpki/routinator/

Jak to wygląda naprawdę?

- Konfiguracja Routinator 3000 (@ULNetLabs):

```
/usr/local/etc/routinator $ more routinator.conf
repository-dir = "/var/routinator"
tal-dir = "/var/routinator/tals"
refresh = 3600
retry = 600
expire = 7200
history-size = 10
# This is an array of strings, each string a socket address of the form
# "address:port" with IPv6 address in square brackets.
rtr-listen = [”TWÓJ_WŁASNY_IP_DO_KTÓREGO_PODŁĄCZYSZ_ROUTERY:3323"]

log-level = "warn"
log = "default"
syslog-facility = "daemon"

tal-labels = [
    ["afrinic.tal", "AfriNIC RPKI Root"],
    ["apnic.tal", "APNIC RPKI Root"],
    ["arin.tal", "ARIN"],
    ["lacnic.tal", "LACNIC RPKI Root"],
    ["ripe.tal", "RIPE NCC RPKI Root"],
]
```

Elastyczność wykorzystania informacji

- Proces BGP na poziomie globalnym ma wiele równoległych ustawień:
  - Pobierz tablicę RPKI
  - Nie bierz pod uwagę prefiksów ‘invalid’
- Dodatkowo za pomocą polityki BGP można sterować atrybutami sprawdzając status weryfikacji RPKI – bardziej elastyczne, ale oczywiście potencjalnie groźne

Zamiast walidacji – polityki routingu

Cisco IOS XR

- Tożsamy z “bgp bestpath origin-as use validity”:

```
route-policy BGP-RPKI
  if validation-state is valid then
    set local-preference 150
    pass
  elseif validation-state is not-found then
    set local-preference 90
  else
    drop
  endif
end-policy
```

Jak to wygląda naprawdę?

Cisco IOS-XE

- Po skonfigurowaniu może wyglądać tak (dla IPv6):

```
rtr-edge#sh bgp ipv6 unicast
[...]
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric LocPrf Weight Path
N*>   2001::/32       2001:db8::666:1                     0 57355 15694 6939 i
N*>   2001:4:112::/48 2001:db8::666:1
                                                          0 57355 15694 112 i
V*>   2001:200::/32   2001:db8::666:1
                                                          0 57355 15694 1299 2500 i
V*>   2001:200:900::/40
                      2001:db8::666:1
I*    2001:388:CF85::/48
I*    2001:5A0:3902::/48
I*    2001:5A0:3903::/48
```

Po włączeniu wykorzystania RPKI

Cisco IOS XR

```
router bgp 65001
 address-family ipv4 unicast
  bgp origin-as validation enable
 address-family ipv6 unicast
  bgp origin-as validation enable
```

```
RP/0/RP0/CPU0:feeder#show bgp origin-as validity
[...]

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled
    Network          Next Hop            Metric LocPrf Weight Path
V*> 1.0.0.0/24       192.0.2.101                            0 57355 15694 13335 i
V*> 1.0.4.0/22       192.0.2.101                            0 57355 15694 6939 4826 38803 i
V*> 1.0.4.0/24       192.0.2.101                            0 57355 15694 6939 4826 38803 i
N*> 1.0.64.0/18      192.0.2.101                            0 57355 15694 1299 2497 7670 18144 i
N*> 1.0.128.0/17     192.0.2.101                            0 57355 15694 6939 38040 23969 i
N*> 1.0.128.0/18     192.0.2.101                            0 57355 15694 6939 38040 23969 i
I*> 1.6.172.0/24     192.0.2.101                            0 57355 15694 6453 9583 i
I*> 1.6.219.0/24     192.0.2.101                            0 57355 15694 6453 9583 137130 i
I*> 1.6.229.0/24     192.0.2.101                            0 57355 15694 6453 4755 i
```

Po włączeniu wykorzystania RPKI

Cisco IOS XR

```
router bgp 65001
 address-family ipv4 unicast
  bgp origin-as validation enable
 address-family ipv6 unicast
  bgp origin-as validation enable
```

```
RP/0/RP0/CPU0:feeder#show bgp origin-as
[...]

Origin-AS validation codes: V valid, I i
    Network          Next Hop
V*> 1.0.0.0/24       192.0.2.101
V*> 1.0.4.0/22       192.0.2.101
V*> 1.0.4.0/24       192.0.2.101
N*> 1.0.64.0/18      192.0.2.101
N*> 1.0.128.0/17     192.0.2.101
N*> 1.0.128.0/18     192.0.2.101
I*> 1.6.172.0/24     192.0.2.101
I*> 1.6.219.0/24     192.0.2.101
I*> 1.6.229.0/24     192.0.2.101
```

```
RP/0/RP0/CPU0:feeder#sh bgp ipv4 unicast 1.37.85.0/24
BGP routing table entry for 1.37.85.0/24
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker             848463      848463
Last Modified: Sep 19 16:23:38.228 for 00:02:09
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  57355 15694 6453 4775 132199
    192.0.2.101 from 192.0.2.101 (192.0.2.101)
      Origin IGP, localpref 100, valid, external, best, group-best
      Received Path ID 0, Local Path ID 1, version 848463
      Origin-AS validity: invalid
```

```
0 57355 15694 6939 38040 23969 i
0 57355 15694 6453 9583 i
0 57355 15694 6453 9583 137130 i
0 57355 15694 6453 4755 i
```

Włączymy walidacje prefiksów

Cisco IOS XR

```
RP/0/RP0/CPU0:feeder(config)#router bgp 65001
RP/0/RP0/CPU0:feeder(config-bgp)#address-family ipv4 unicast
RP/0/RP0/CPU0:feeder(config-bgp-af)#bgp bestpath origin-as use validity
RP/0/RP0/CPU0:feeder(config-bgp-af)#exit
RP/0/RP0/CPU0:feeder(config-bgp)#address-family ipv6 unicast
RP/0/RP0/CPU0:feeder(config-bgp-af)#bgp bestpath origin-as use validity
RP/0/RP0/CPU0:feeder(config-bgp-af)#commit
```

```
RP/0/RP0/CPU0:feeder#sh bgp ipv4 unicast 1.37.85.0/24 detail
Sun Sep 19 16:32:35.330 CEST
BGP routing table entry for 1.37.85.0/24
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker            1694224     1694224
    Flags: 0x00000001+0x00000000;
Last Modified: Sep 19 16:30:08.228 for 00:02:27
Paths: (1 available, no best path)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Flags: 0x1000000000000003, import: 0x20
  Not advertised to any peer
  57355 15694 6453 4775 132199
    192.0.2.101 from 192.0.2.101 (192.0.2.101), if-handle 0x00000000
      Origin IGP, localpref 100, valid, external
      Received Path ID 0, Local Path ID 0, version 0
      Origin-AS validity: invalid
```

Podpisz swoje prefiksy... ale tylko swoje... ;)

Type of Certificate Authority

You can choose between asking the RIPE NCC to host your RPKI Certificate Authority (Hosted RPKI) or running your own Certificate Authority (Delegated RPKI).

Select “Hosted” if you would like the RIPE NCC to host your Certificate Authority, keys, ROAs, manifests etc. and publish the information in our repository. You will only need to maintain your ROAs in our dashboard. This is the recommended option if you are not an RPKI expert.

Select “Delegated” to run your own Certificate Authority and and to host your own keys, ROAs, manifests etc. You will need to run additional software to proceed.

- ○ Hosted
- ○ Delegated

By clicking on 'I accept' below you confirm that that you have read, understood and agree to the RIPE NCC Certification Service Terms and Conditions.

I accept. Create my Certificate Authority

Resource Certification (RPKI) for Provider Independent End Users and Legacy End Users

The RIPE NCC offers Resource Certification (RPKI) as a free, additional service to holders of Provider Independent (PI) and Legacy End User resources. This page offers an overview of the RPKI requirements, contact information and a link to the wizard to guide you through the process.

A resource certificate can be requested and managed by the Internet number resource holder directly or by the Sponsoring LIR, depending on who maintains the **organisation** object in the RIPE Database.

You will need to meet the following requirements:

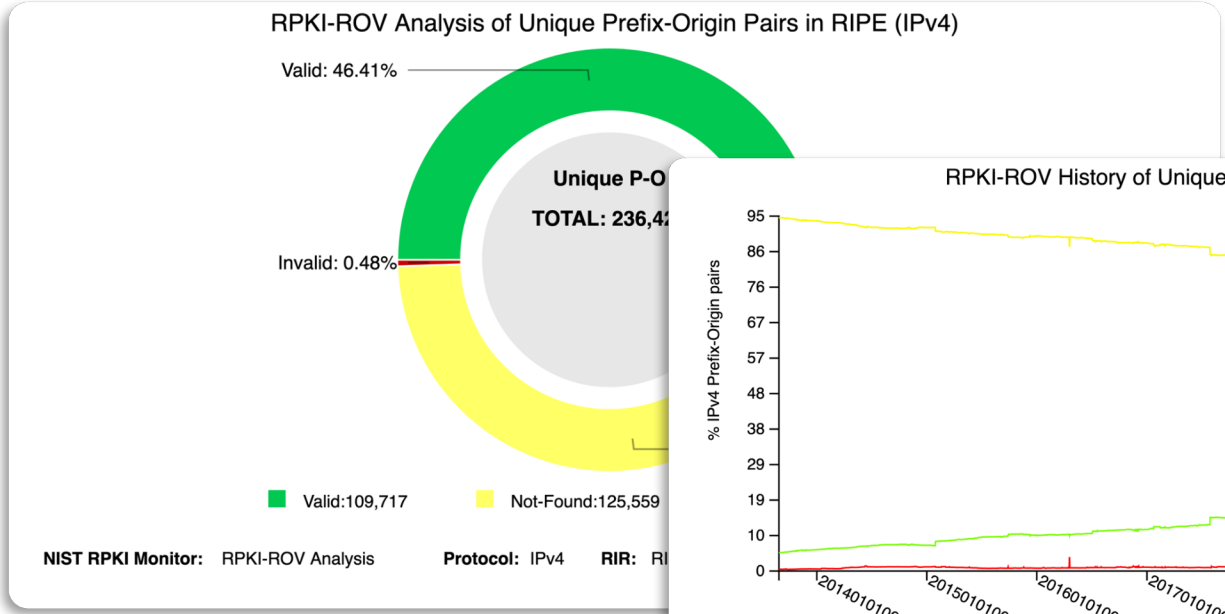
1. An “End User Assignment Agreement” (for PI) or “Sponsoring LIR Agreement for Legacy Internet Resources (for Legacy) has been submitted and verified by the RIPE NCC
2. You can authenticate against one of the maintainers of the RIPE Database **organisation** object to which your resource objects refer

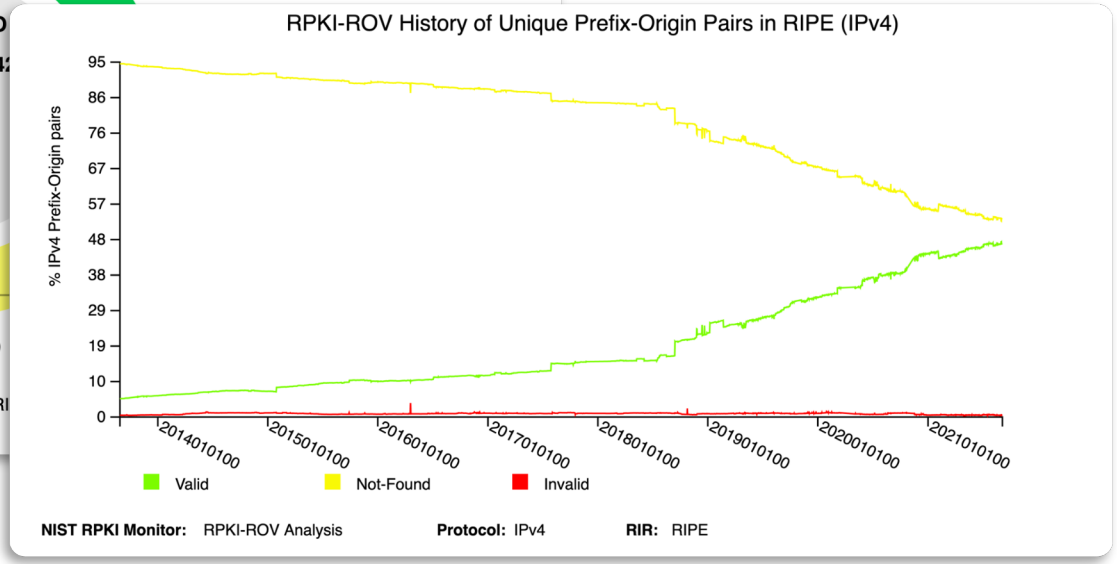
If you do not meet all of the above requirements, the **Sponsoring LIR** should contact the RIPE NCC at lir-help@ripe.net in order to resolve this.

We have created a wizard that will guide you through the process. You will need to sign in with your RIPE NCC Access account. If you don't have a RIPE NCC Access account, you can create one. Once you have successfully completed the wizard, you will be able to log into the LIR Portal, request a certificate, and manage Route Origin Authorisations.

Start the wizard to set up Resource Certification for PI and Legacy End User resources

Statystyki BGP RPKI

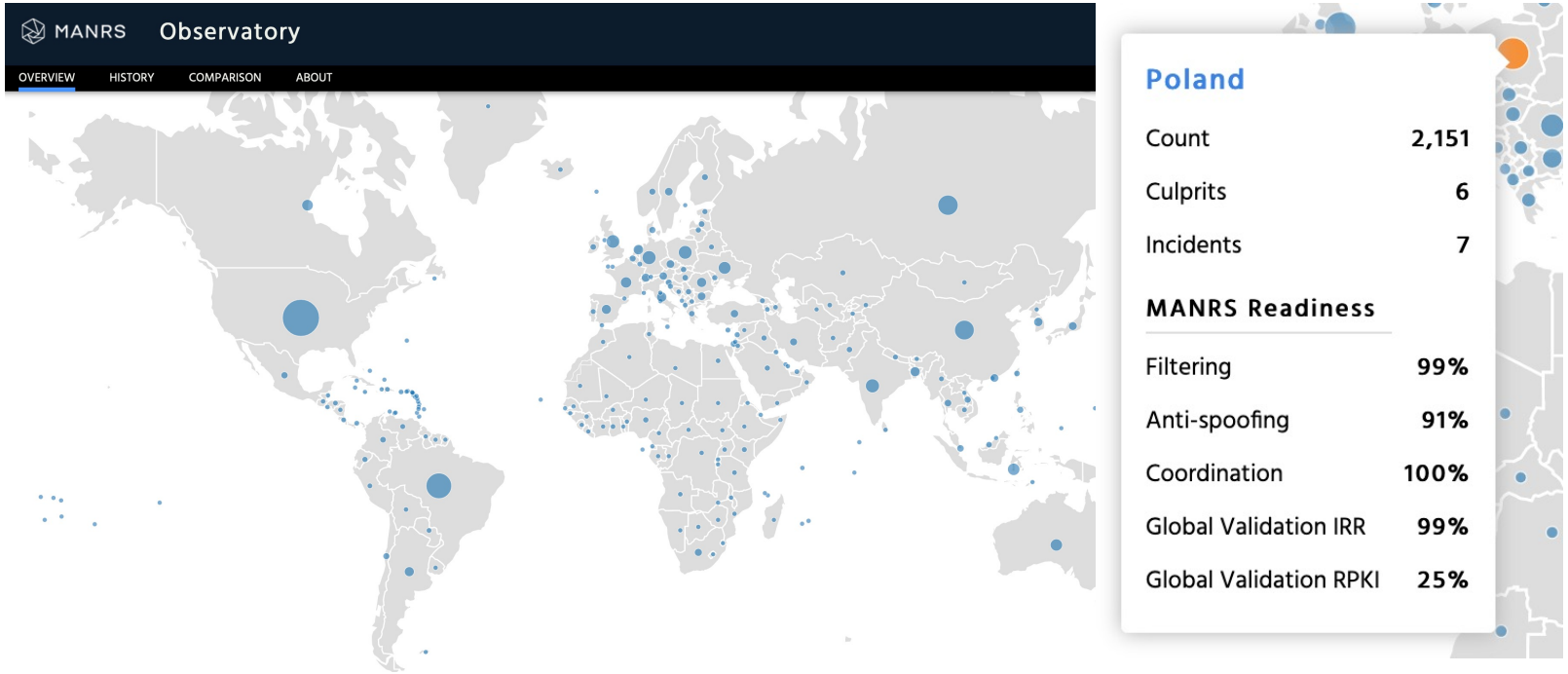




Sources:
https://rpki-monitor.antd.nist.gov/ROV/20210905.00/R/All/4

Statystyki dla Polski

- https://observatory.manrs.org/#/overview



Co z tym można zrobić?

Pomysł #2: Wzbogacić raporty

Po co BGP do statystyk?

- Pełen feed BGP (IPv4 i IPv6) pozwala wypełnić poprawnie pola źródłowego i docelowego ASNa dla ruchu na Twoim routerze
- Jeśli interesuje Cię “z kim” wymieniasz najwięcej ruchu, warto skorzystać z tego sposobu
- Więcej tu: https://lukasz.bromirski.net/pl/post/flexible-netflow-w-sluzbie-statystyk/
- Jeśli nie masz możliwości kupienia kolektora NetFlow/IPFIX/etc, jest parę dobrych, darmowych rozwiązań w sam raz “dla domu”:

  https://github.com/manuelkasper/AS-Stats

  https://github.com/nidebr/as-stats-gui

Co z tym można zrobić?

Pomysł #3: Inżynieria ruchowa

Co warto wiedzieć o takim dłubaniu

- Nie bierzemy pod uwagę ”rozwiązań” SD-WAN
- Nie bierzemy pod uwagę wariacji na temat SCTP i MPTCP:

  https://www.openmptcprouter.com/
- Patrzymy na świat w warstwie 3 i 4, warstwa aplikacyjna jest dla nas drugorzędna (odważne założenie)
- Tak, to tylko zabawa, ale pożyteczna i dająca rezultaty
- Pomysły można łączyć ;)

Inżynieria ruchu wychodzącego
Wiele dostępnych operatorów (1/3)

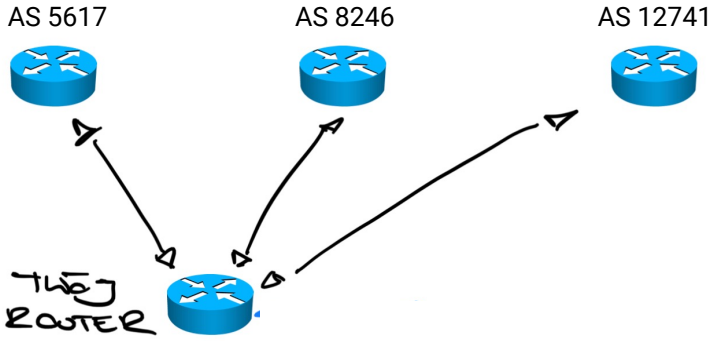
Cisco IOS-XE

- Przykład:

  AS 5617 – Orange

  AS 8246 – T-Mobile

  AS 12741 - Netia



- Najprościej – trzy trasy domyślne do każdego z operatorów

```
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
```

- Parę trików z CEFem:

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

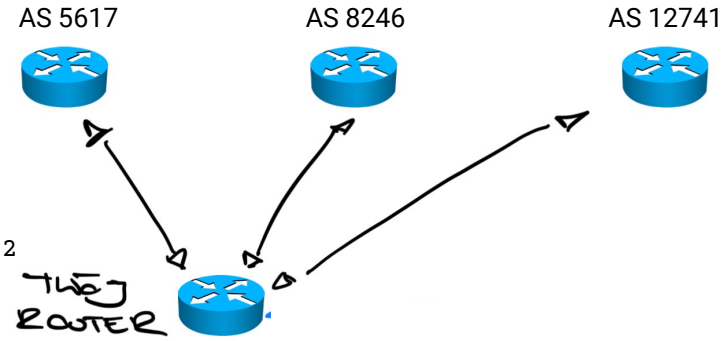
Inżynieria ruchu wychodzącego

Wiele dostępnych operatorów (2/3)

Cisco IOS-XE

- Dodaj śledzenie sąsiada żeby uniknąć problemów z osiągalnością trasy jeśli nie dojdzie do awarii całego interfejsu:

```
ip sla 101
 icmp-echo 100.64.0.2 source-interface GigabitEthernet0/0/0
  tag Track 101: Orange Gi0/0/0
ip sla schedule 101 life forever start-time now
ip sla 102
 icmp-echo 169.254.0.2 source-interface GigabitEthernet0/0/1
  tag Track 102: T-Mobile Gi0/0/1
ip sla schedule 102 life forever start-time now
ip sla 103
 icmp-echo 195.81.100.2 source-interface GigabitEthernet0/0/2
  tag Track 103: Netia Gi0/0/2
ip sla schedule 103 life forever start-time now
!
track 101 ip sla 101 state
track 102 ip sla 102 state
track 103 ip sla 103 state
```



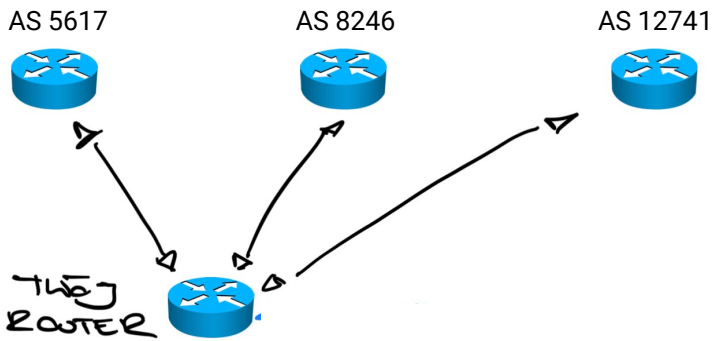
- Dołączenie śledzenia do tras statycznych:

```
ip route 0.0.0.0 0.0.0.0 100.64.0.2 track 101
ip route 0.0.0.0 0.0.0.0 169.254.0.2 track 102
ip route 0.0.0.0 0.0.0.0 195.81.100.2 track 103
```

Inżynieria ruchu wychodzącego
Wiele dostępnych operatorów (3/3)

- Jak pójść dalej? Można nieco zautomatyzować dopisywanie dokładniejszych prefiksów do każdego z operatorów:

```
$ 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
```



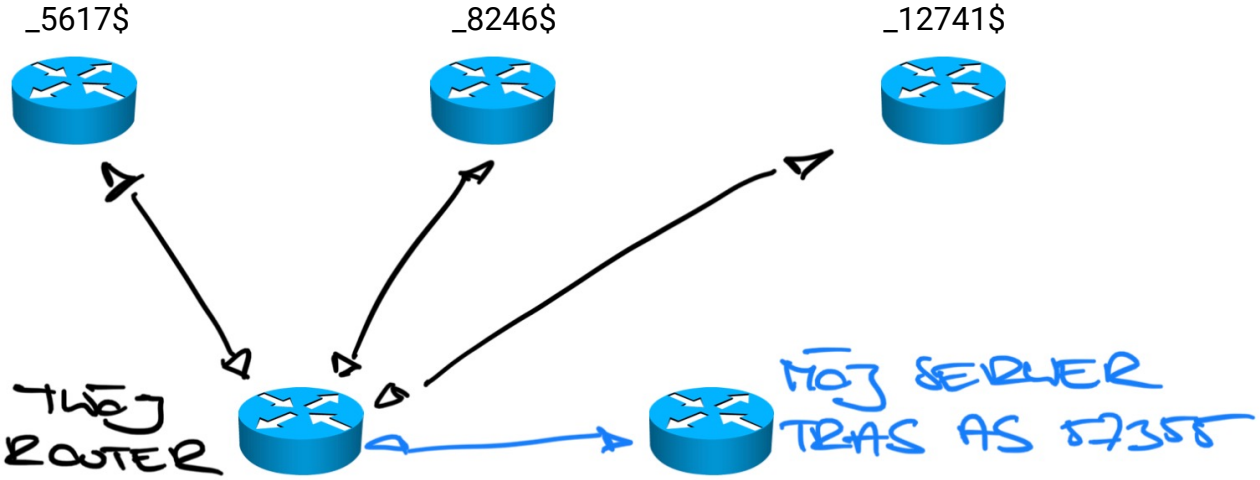

```
$ whois -h whois.radb.net -- "-K -i origin AS5617" | grep route | awk '{ print $2 }'
194.204.128.0/18
195.116.0.0/16
195.117.0.0/16
195.205.0.0/16
212.244.0.0/16
212.160.0.0/16
213.25.0.0/16
[...]
```

* https://github.com/snar/bgpq3

Inżynieria ruchu wychodzącego

Można też posłużyć się właśnie BGP

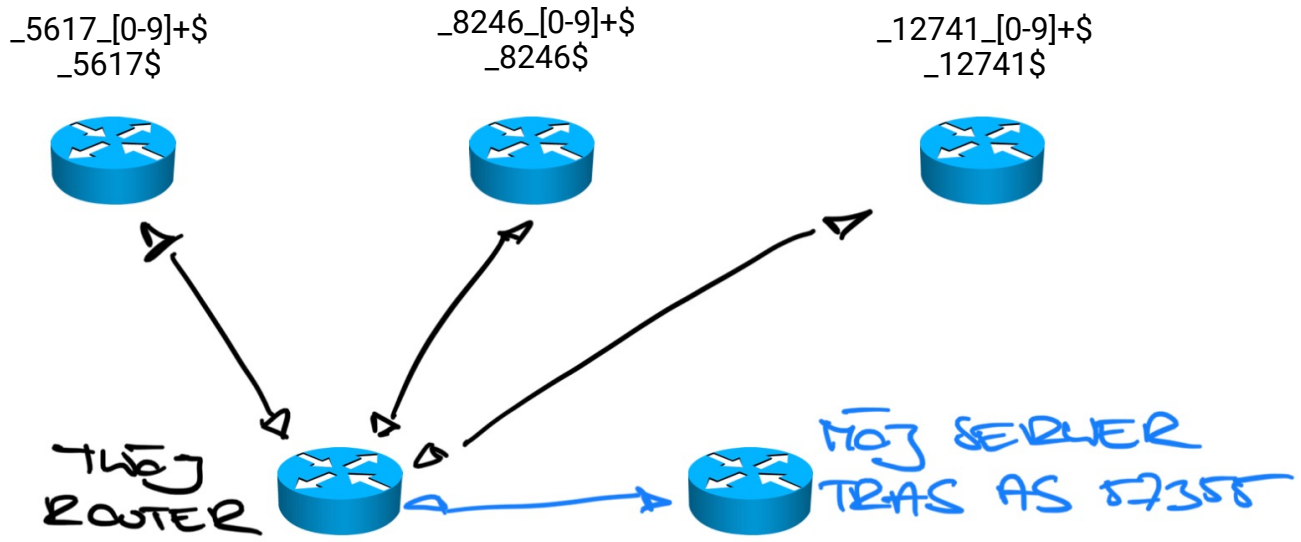
- ”Żywa” sesja BGP, nawet zestawiona zupełnie w oderwaniu od Twoich lokalnych operatorów, dostarcza Ci na bieżąco prefiksów rozgłaszanych przez każdego z nich
- Prosto jest zatem kierować ruch do prefiksów danego operatora... przez łącze do tego operatora



Inżynieria ruchu wychodzącego

Można też posłużyć się właśnie BGP

- Oczywiście pomysł można rozszerzyć – dlaczego tylko prefiksy należące do operatora upstreamowego, skoro można np. rozszerzyć je o jego sąsiadów… albo sąsiadów jego sąsiadów?



Inżynieria ruchu wychodzącego

Jak to skonfigurować w praktyce?

Cisco IOS-XE

- Zestawiasz sobie sesję BGP z mojego projektu
- Odbierając prefiksy odpowiednio modyfikujesz BGP next-hop na podstawie AS_PATH

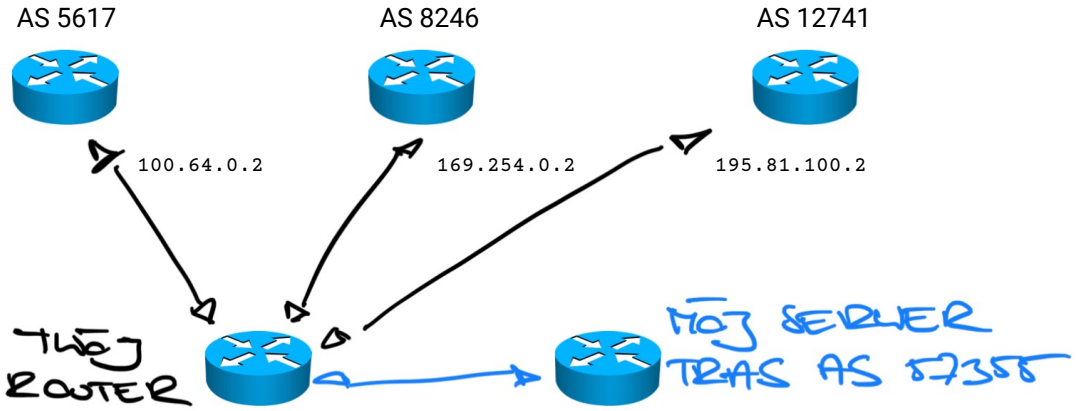
```
route-map BGP-TE permit 10
 description TE towards 5617
 match as-path 101
 set ip next-hop 100.64.0.2

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 _12741$

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



- Ruch automatycznie podąży optymalną ścieżką i samodzielnie się przełączy w razie awarii lub zmian w globalnej tablicy BGP

Inżynieria ruchu wychodzącego

Ale ale ale.... nie mam next-hopa, bo mam interfejs PPPoA/PPPoE/DHCP

Cisco IOS-XE

- Definiujemy trasę pośrednią skierowaną na lokalny, prywatny adres, wskazując statycznie lub dynamicznie obsługiwane interfejsy, np. tak:

```
ip route 192.0.2.100 255.255.255.255 100.64.0.2 name ORANGE
ip route 192.0.2.101 255.255.255.255 169.254.0.2 name TMOBILE
ip route 192.0.2.102 255.255.255.255 195.81.100.2 name JJS
ip route 192.0.2.103 255.255.255.255 195.81.1.2 name PPPOE
```

- Atrybut modyfikowany w BGP ustawiasz na trasę pośrednią a nie bezpośrednio IP po drugiej stronie (na platformach Cisco z uwagi na CEF brak wpływu na wydajność routingu)

```
route-map BGP-TE permit 40
 description TE towards 8737 (przykład)
 match as-path 104
 set ip next-hop 192.0.2.103

rtr-edge#sh ip cef 13.32.1.0/24 detail
13.32.1.0/24, epoch 2, flags [rib only nolabel, rib defined all labels]
  0 packets, 0 bytes switched to the prefix
  NetFlow: Origin AS 16509, Peer AS 0, Mask Bits 0
  recursive via 192.0.2.103
    recursive via 195.81.1.2
      attached to GigabitEthernet0/0/4
```

Inne miejsca w które warto zajrzeć
(i pomyśleć)

Więcej o BGP, routingu i różnych innych... 1/2

Oderwij karteczkę

- BGP dla bezpieczeńswa – prosty przewodnik po Blackholingu i FlowSpec: https://lukasz.bromirski.net/docs/prezos/certee2017/BGP_Security_101.pdf
- Skalowanie usług za pomocą IP Anycast: http://lukasz.bromirski.net/docs/prezos/plnog2011/ip_anycast.pdf
- BGP w labie część #3 – feedy IPv4 i IPv6 za darmo (“operatorzy go nienawidzą!”): https://lukasz.bromirski.net/post/bgp-w-labie-3/
- Jak zbudowana jest moja sieć domowa – BGP, anycast, serwery DNS, etc: https://lukasz.bromirski.net/post/moja-siec-domowa-2/
- Blog Daniela Stockera o serwowaniu prefiksów: https://puffy.nolink.ch/posts/fullbgp_at_home/
- NLNetLabs Routinator 3000: https://nlnetlabs.nl/projects/rpki/routinator/

Więcej o BGP, routingu i różnych innych... 2/2

Oderwij karteczkę

- Blog i narzędzia Tomka Mikołajka
  https://showroute.pl/blog/
- Blog “Inna sieć”:
  https://innasiec.pl/
- Blog Piotra Wojciechowskiego:
  https://blog.it-playground.eu/

PLNOG
where ipeople meet

Zabawy z lokalnym i globalnym routingiem...

(dla zabawy, bezpieczeństwa, lepszej inżynierii ruchu i for profit też)

Łukasz Bromirski
lukasz@bromirski.net
https://lukasz.bromirski.net
@LukaszBromirski

Q&A