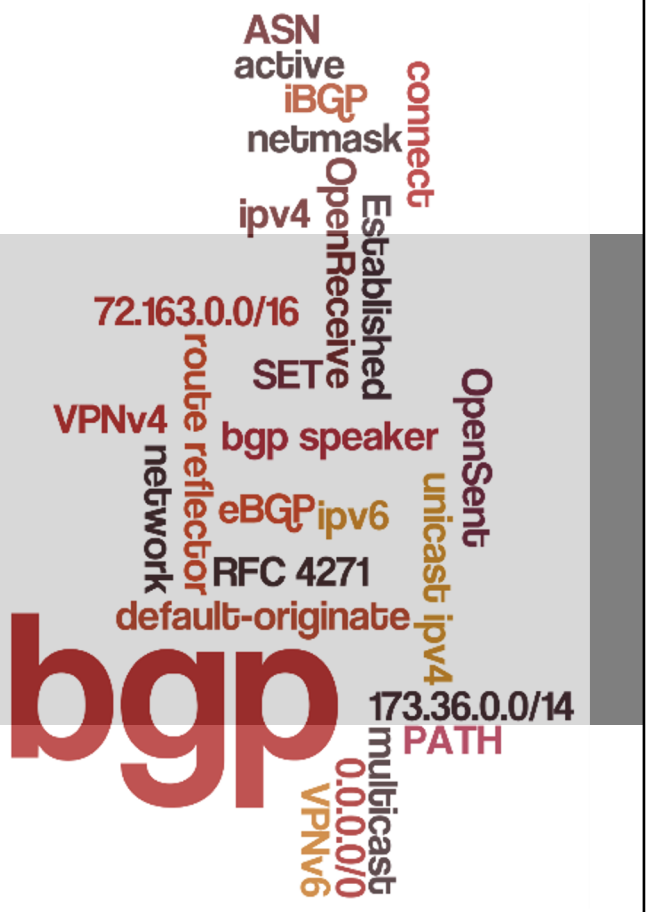
Jak zbudować profesjonalny styk z internetem cz. 2?



Łukasz Bromirski
lukasz.bromirski@gmail.com

PLNOG Webinaria, 10.X.2017

Plan na dzisiaj

- Higiena i dobre praktyki – jestem końcowym ASem
- Higiena i dobre praktyki – jestem tranzytowym ASem
- Skalowanie i większa kontrola polityk w większych ASach
  - AddPath
  - Optimal Route Reflection
- Q&A

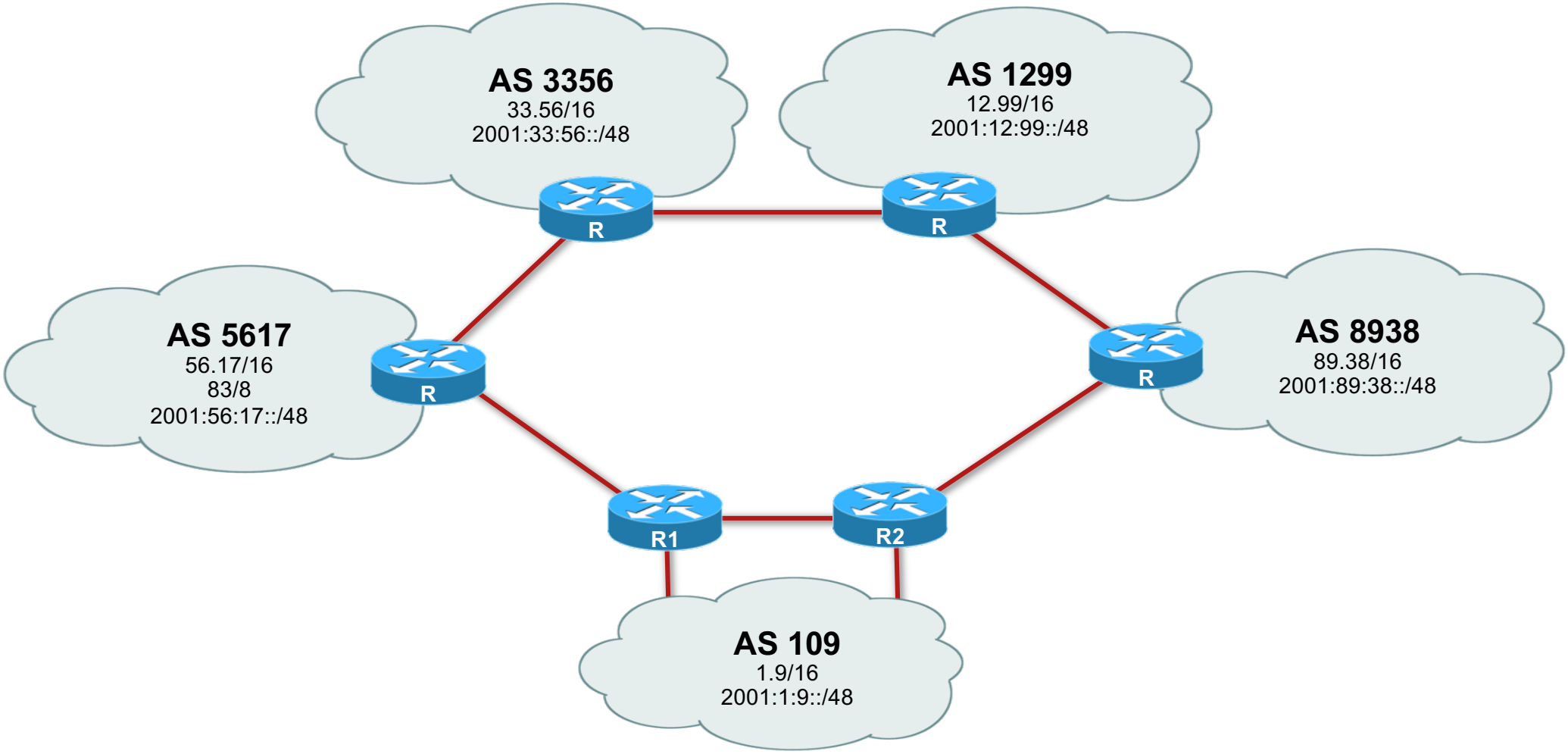
Higiena i dobre praktyki

Jestem końcowym ASem

Nie rozgłaszamy tranzytu!

- „Zabawne” wpadki
- Zdarzają się do dziś
- Teoretycznie operatorzy się przed tym bronią
- Ufaj i kontroluj

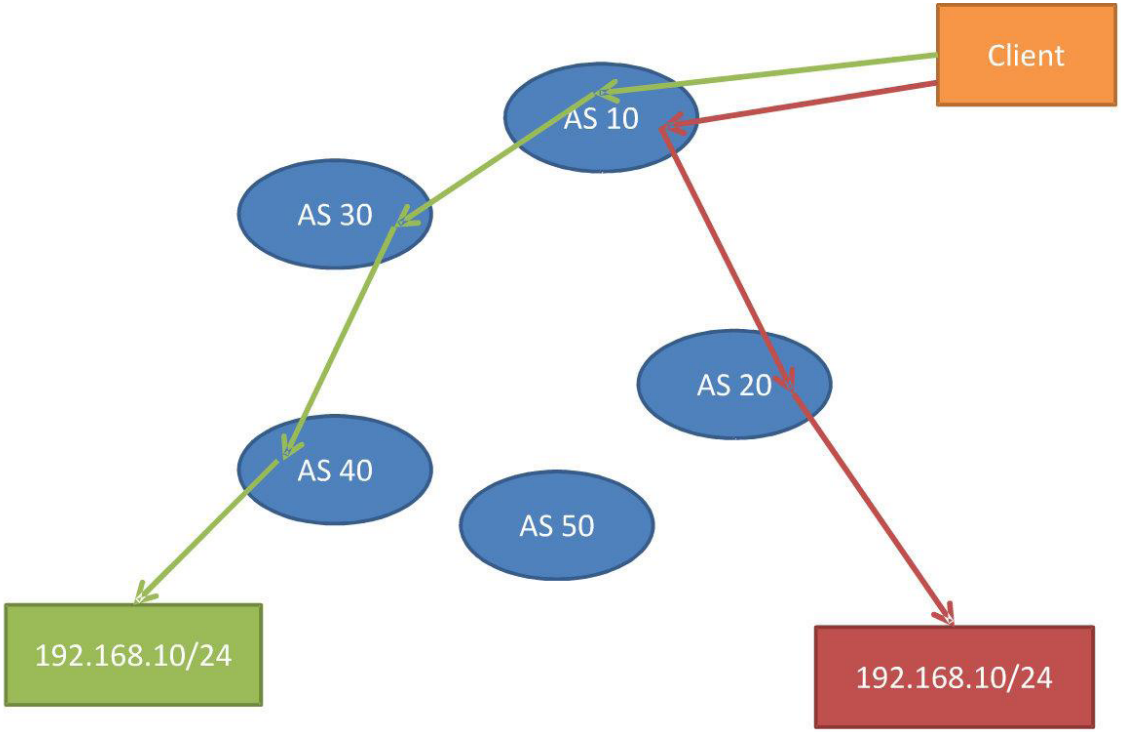
Jak ochronić się przed tranzytem?



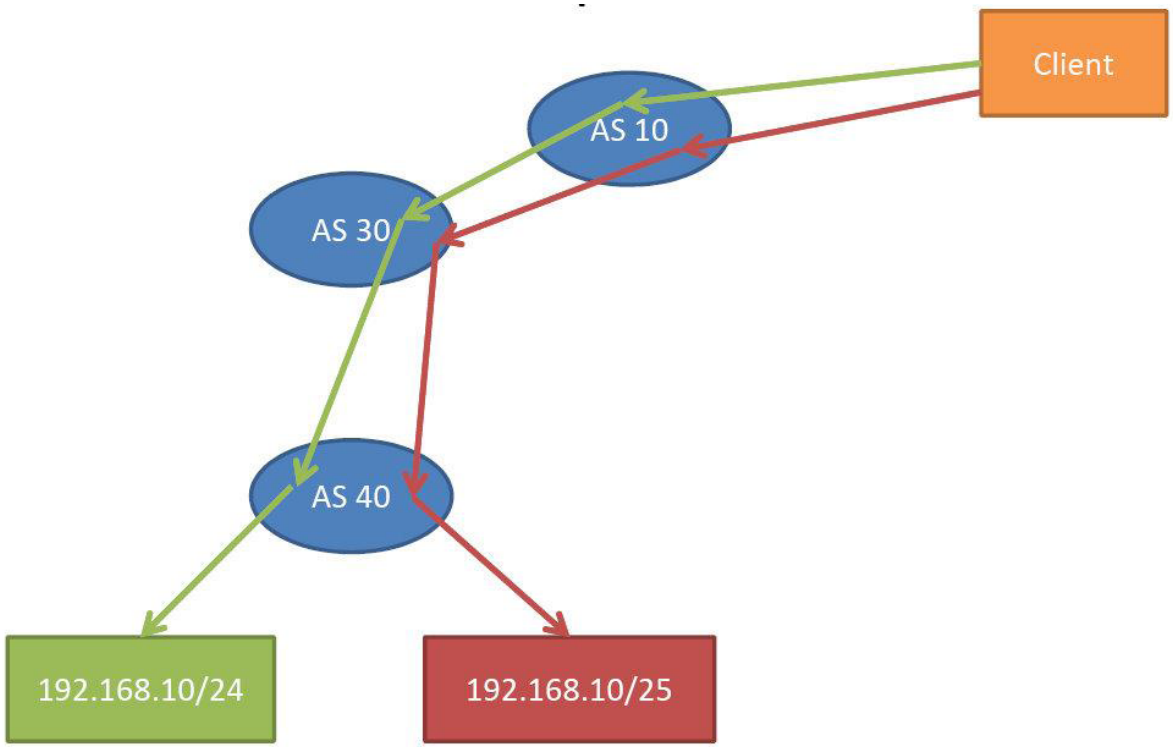
Bezpieczeństwo i higiena ponad wszystko

- unicast Reverse Path Filtering (uRPF)
- RPKI (SIDR)
- Właściwe wpisy w bazie RIR

RPKI: jaki problem próbujemy rozwiązać?

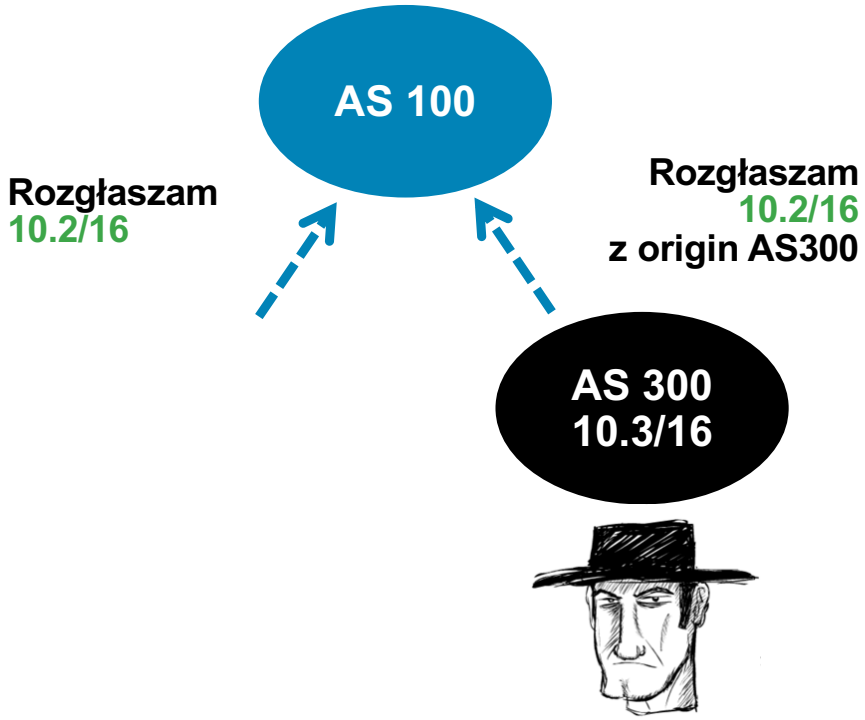


RPKI: jaki problem próbujemy rozwiązać?

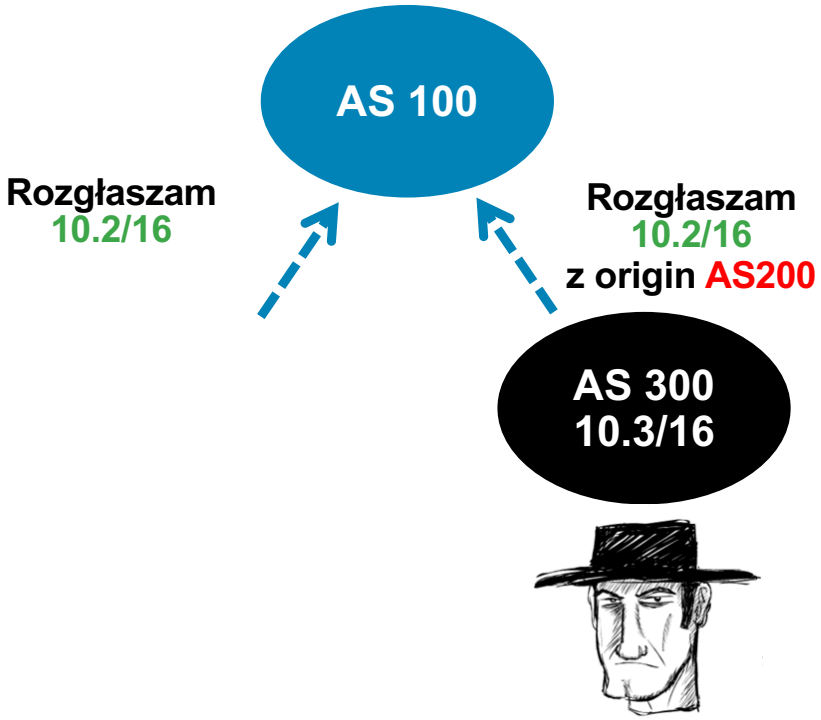


Dwa osobne problemy:

- 1°) Sprawdzenie (walidacja) pochodzenia trasy



- 2°) Walidacja ścieżki



RPKI: trzy elementy

Infrastruktura RPKI

- Repozytorium weryfikowalnych obiektów oparte o infrastrukturę klucza publicznego
- Podąża za zasobami zarządzanymi przez RIR (IPv4, IPv6 i ASN)

Walidacja Origin Prefiksu

- Weryfikujemy tylko Origin AS
- Rozwiązuje większość problemów
- Nie ma wpływu na router

Walidacja AS_PATH prefiksu

- Trwają prace standaryzacyjne nad BGPSEC

SIDR z perspektywy routera

- Po podłączeniu routera do walidatora (osobna sesja TCP lub BGP) prefiks może mieć trzy stany:

  valid – zweryfikowano origin i dopuszczalne długości maski pozytywnie

  invalid – j.w. - negatywnie

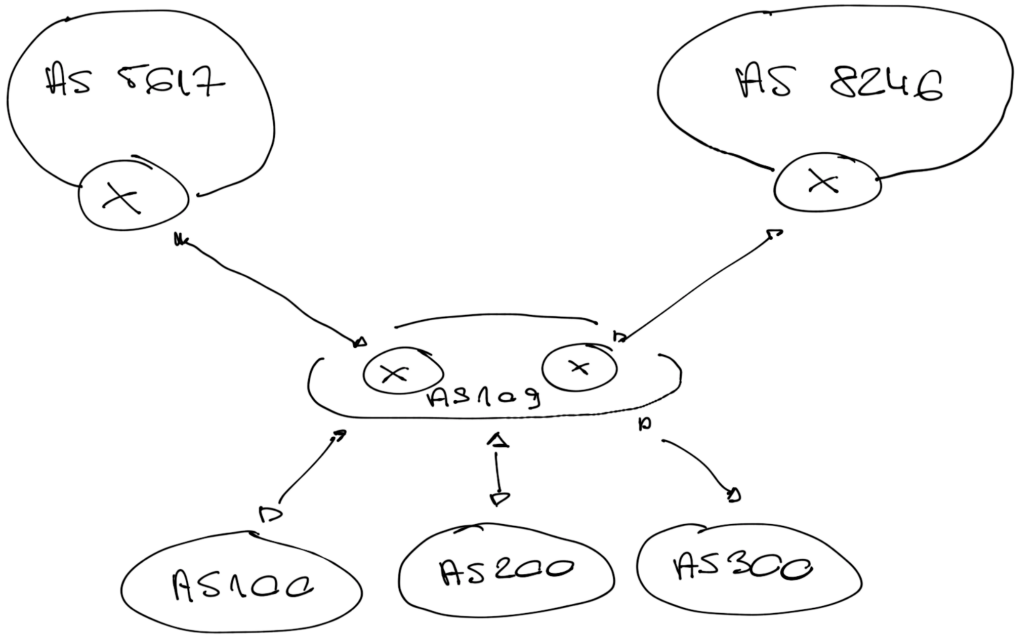
  not found – nie ma rekordu w bazie danych

Higiena i dobre praktyki

Jestem Asem tranzytowym

Jesteśmy tranzytem

- „Pełne” tablice?
- Selektywny wybór wyjść (upstreamów)?
- Tranzyt między klientami?
- Odpowiednie modyfikacje w rekordach RIR



* Funkcjonalność BGP Route Server – IOS-XE na ASR 1k oraz CSR1000v

Skalowanie i dokładniejsza kontrola

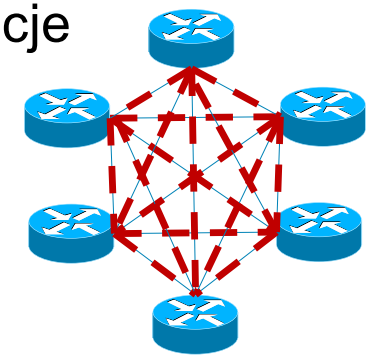
BGP Route Reflector i rozszerzenia do BGP

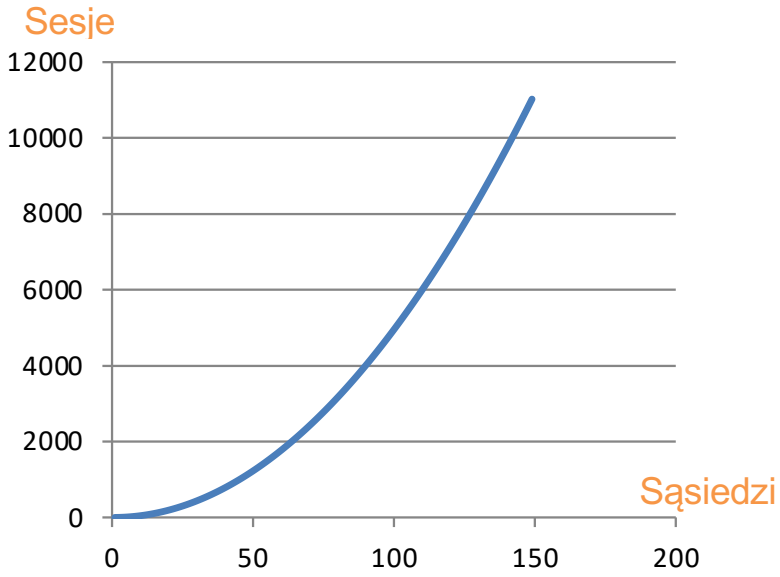
Pełna siatka iBGP – czy to się skaluje?

- iBGP wymaga połączeń w pełnej siatce
- Łącznie – n*(n-1)/2 sesji iBGP
- Dwa rozwiązania:

  Route Reflectory

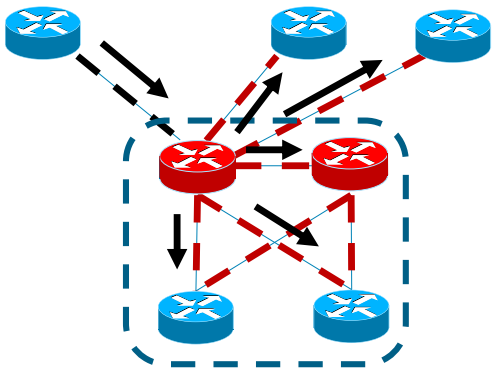
  Konfederacje





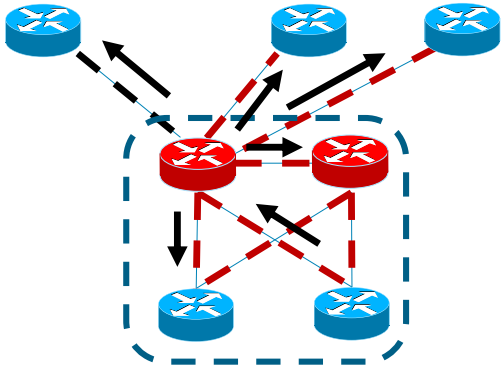
BGP Route Reflector – co odbijamy?

Prefiks od sąsiada eBGP



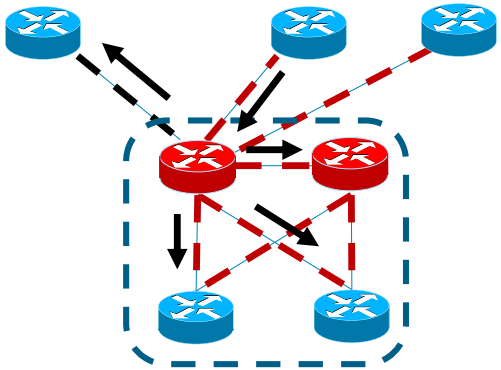
RR wysyłą prefiks do klientów i nie-klientów (oraz innych sąsiadów eBGP)

Prefiks od klienta RR



RR odbija prefiks do klientów i wysyłą go do nie-klientów (w tym innych sąsiadów eBGP)

Prefiks od nie-klienta



RR odbija prefiks do klientów (oraz innych sąsiadów eBGP)

To jak dużo BGP RR potrzebuję?

- Więcej RR = większa redundancja

  Ale i więcej konfiguracji, więcej sesji, więcej pamięci

- Ogólnie przyjęty i spotykany w sieciach model:

  2 RR dla każdej usługi (lub zestawu usług)

  Np.:

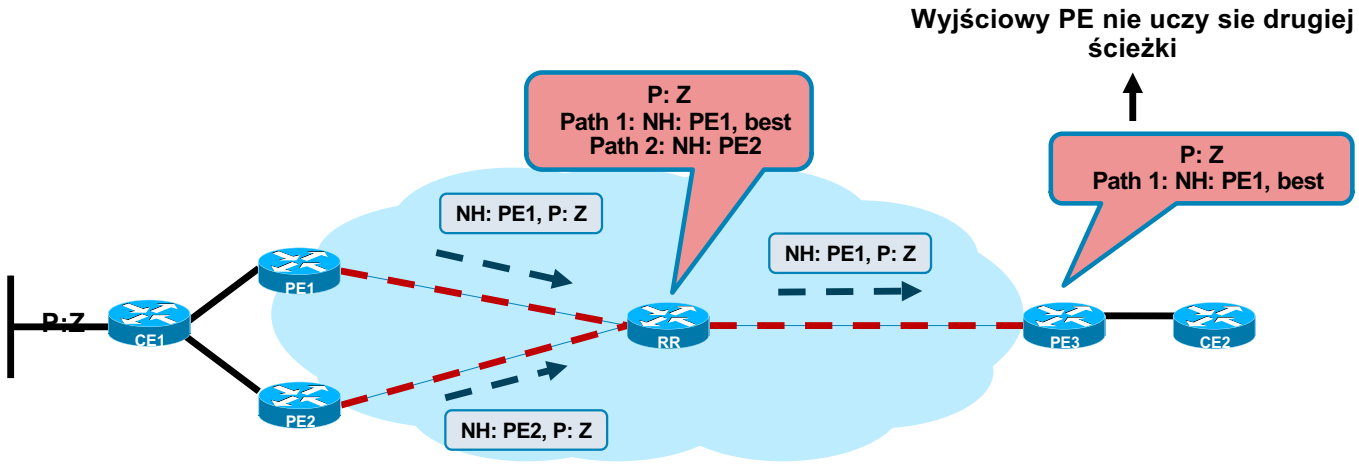
    2 RR dla IPv4

    2 RR dla VPNv4

    2 RR dla IPv6

    ...

Wybór ścieżki w sieci z RR



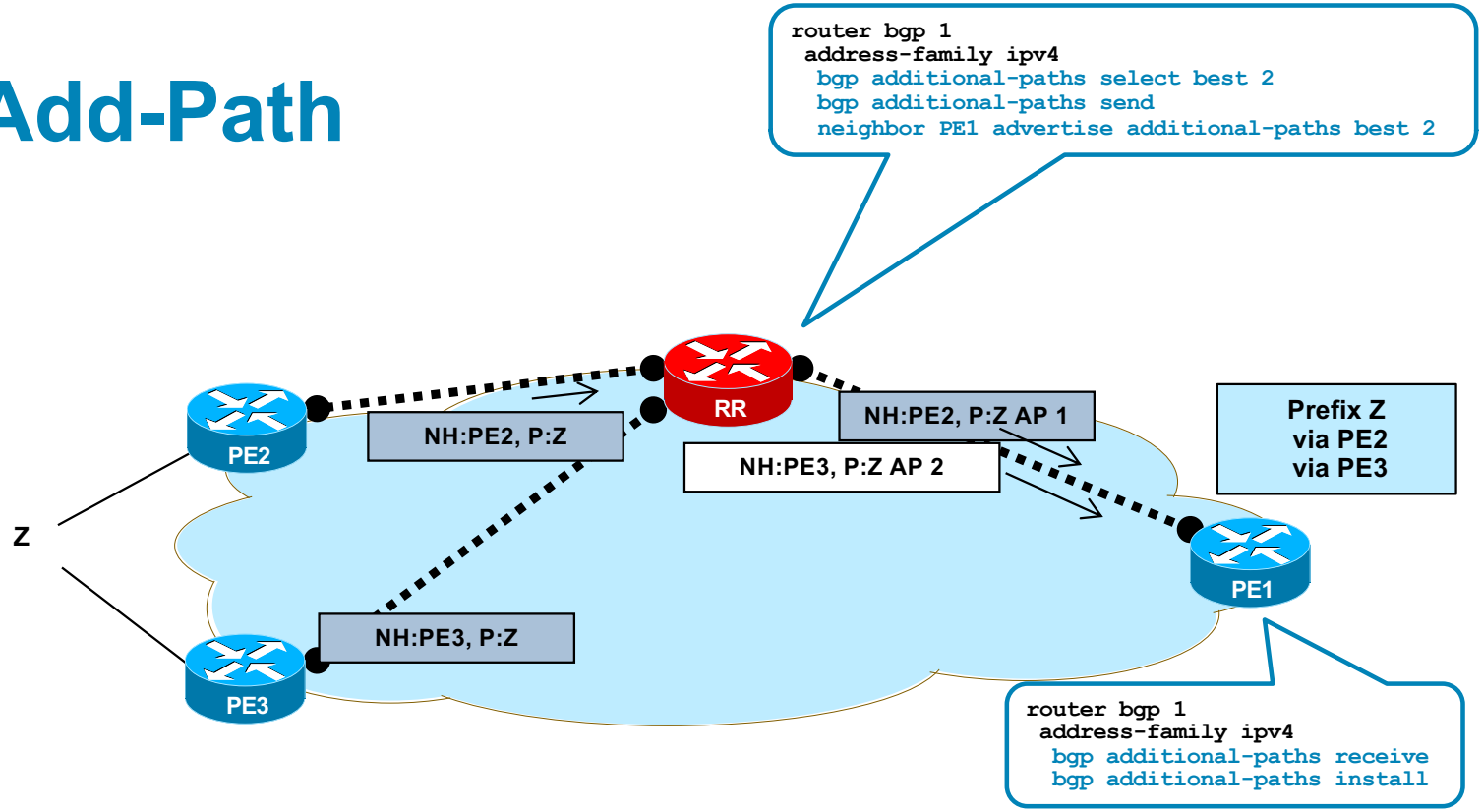
- Protokół BGP4 specyfikuje wybór i propagację jednej najlepszej ścieżki dla każdego prefiksu
- W przypadku użycia RRów, tylko najlepsza zostanie wysłana z RR do speakerów BRP

  „Multipath” nie rozwiązuje tego problemu
- Takie zachowanie prowadzi do wielu ograniczeń – w szczególności związanych z redundancją

Rozgłośmy wiele ścieżek w BGP!

- Unikalne RD w MPLS VPN
- BGP Best External
- BGP Add-Path
- BGP shadow RR / session
- BGP Optimal Route Reflection

BGP Add-Path



- Add-Path rozgłosi od 2 do X dodatkowych ścieżek
- Oczywiście router odbierający ścieżki (RR, PE) musi obsługiwać to rozszerzenie (capability) na etapie zestawiania sesji BGP

BGP Add-Path - rodzaje

Add-n-path: RR wykona obliczenia dla wszystkich tras i wyśle N najlepszych do PE

- Przykładowe zastosowanie: trasa główna + N tras zapasowych

Add-all-path: RR wykona obliczenie tylko dla pierwszej ścieżki, ale i tak wyśle komplet do PE

- Przykładowe zastosowanie: routing ‘hot potato’ mimo obecności RR, ECMP load-balancing w DC

Add-all-multipath+backup: RR wykona obliczenie dla najlepszej ścieżki, wyśle wszystkie trasy o tym samym koszcie i dodatkowo jedną zapasową do PE

- Przykładowe zastosowanie: ECMP load-balancing w DC

BGP Add-Path – konfigracja dla IOS-XR

Konfiguracja przykładowa

```
router bgp 1
 address-family vpnv4
  additional-paths install backup ! stare polecenie
  additional-paths advertise
  additional-paths receive
  additional-paths selection route-policy BGPAddPathX
```

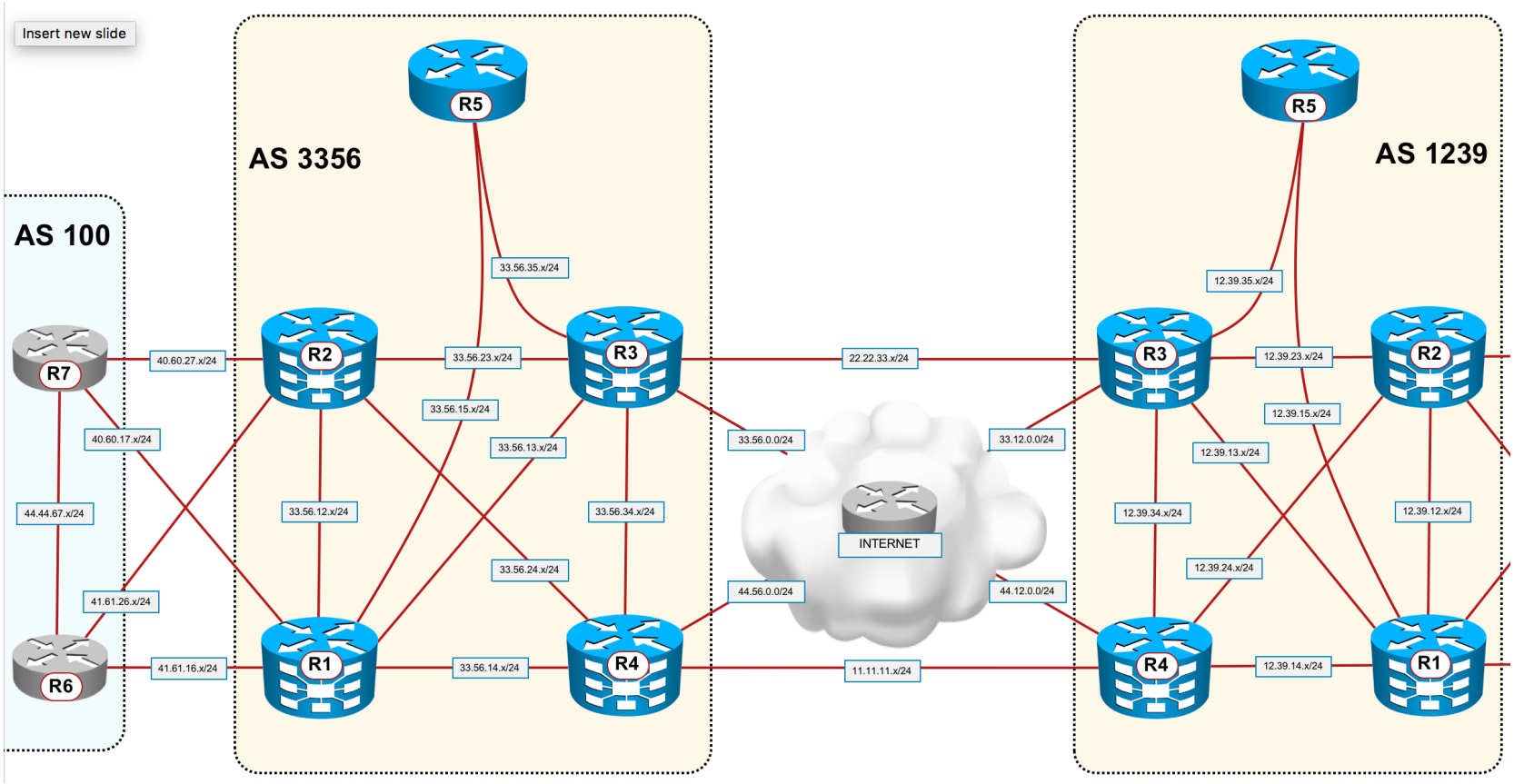
Przykładowe polityki RPL

```
route-policy BGPAddPath1
 if community matches-any (1:1) then
  set path-selection backup 1 install
 elseif destination in (10.1.0.0/16, 10.2.0.0/16)    then
 set path-selection backup 1 advertise install
endif

route-policy BGPAddPath2
 set path-selection all advertise

route-policy BGPAddPath3
 set path-selection multipath advertise
```

BGP RR i AddPath - demonstracja

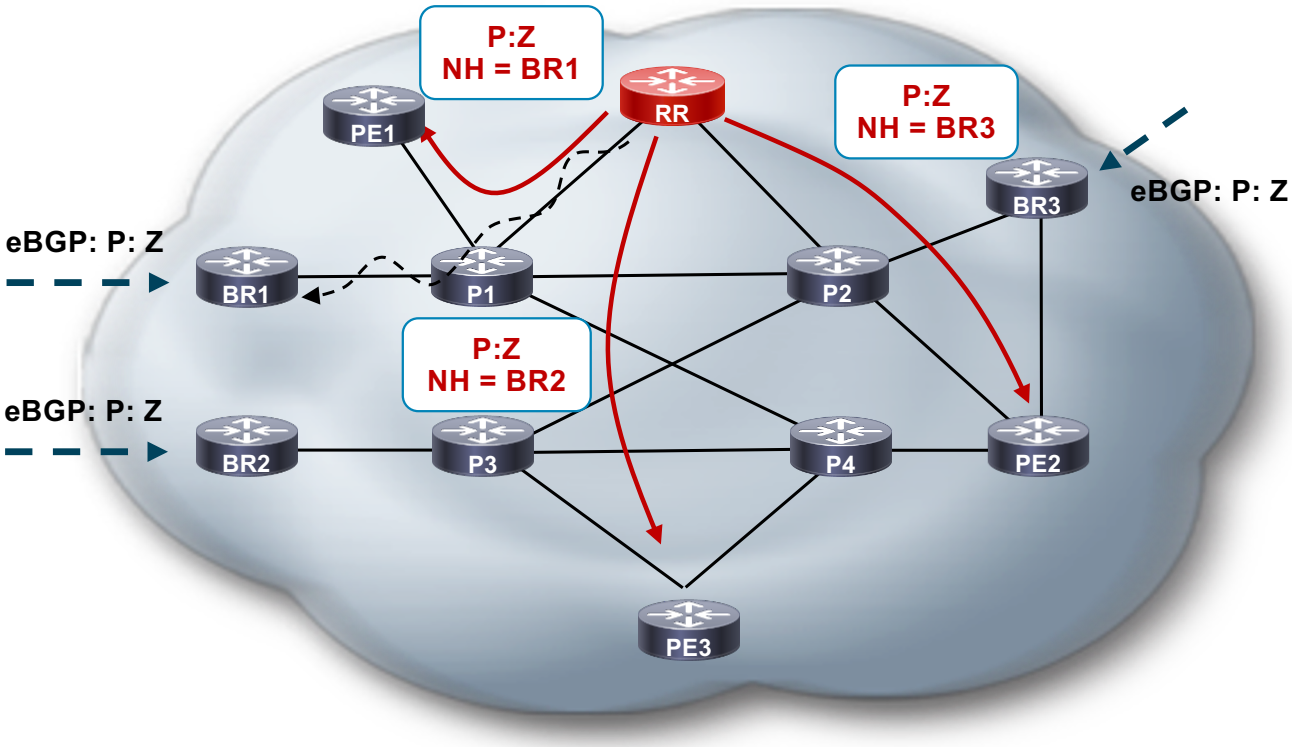


BGP Optimal Route Reflection

- W tradycyjnym routingu BGP o wyborze najlepszej ścieżki do prefiksu przez RR decyduje koszt IGP do rozgłaszającego go PE
  - RR preferuje PE bliższe kosztem PE dalszych od siebie (w rozumieniu kosztu IGP)
- Umieszczenie RR w jednym miejscu topologii pomaga, ale wraz z popularyzacją wirtualnych RR i narzędzi do ich automatycznego “ustawiania” problem rośnie
  - “one są wszędzie!”
- BGP ORR pomaga rozwiązać problem routingu wg. koncepcji “gorącego ziemniaka” wszędzie tam, gdzie topologia IGP nie podąża za umiejscowieniem wszystkich RR

BGP Optimal Route Reflection (ORR)

- RR przelicza ścieżki z perspektywy danego routera (grupy)



- Tylko RR wymaga kodu obsługującego ORR
- Wymagane wsparcie dla BGP LS oraz ze strony protokołu IGP (dystrybucja informacji LS do BGP)
- W Cisco IOS XR (obecnie) – włączony MPLS TE (nie ma potrzeby aktywowania MPLSa)

Chcę wiedzieć więcej!

Parę ważnych rzeczy do zapamiętania

- RIPE RPKI/SIDR: https://www.ripe.net/manage-ips-and-asns/resource-management/certification/router-configuration#Cisco
- Narzędzie IRRPT: https://github.com/6connect/irrpt
- Cisco dCloud – laby z BGP: https://dcloud2-lon.cisco.com/content/catalogue?search=bgp&screenCommand=openSearchScreen&isLoggingIn=true

Co dalej po dzisiejszym webinarium?

- Mamy dedykowany kanał poświęcony projektowaniu sieci

  Napisz do mnie na lukasz.bromirski@gmail.com po zaproszenie

- Repozytorium z topologią .virl oraz konfiguracją Cisco Routem z tej sesji:

  https://github.com/lukasz-bromirski/netdesign.zone/

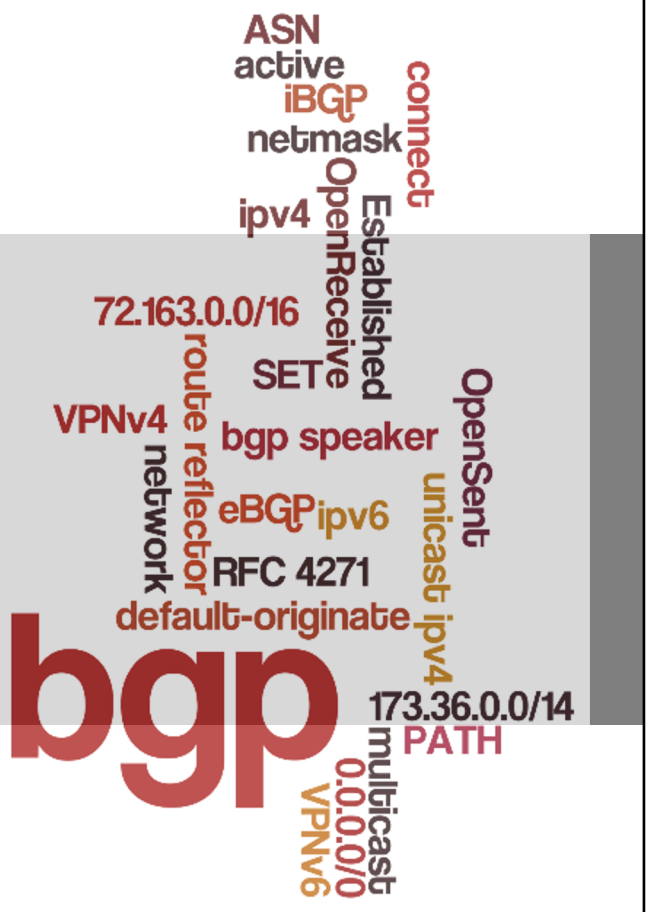
- Blog mój i Piotra Jabłońskiego o projektowaniu sieci i scenariuszach projektowych (w tym BGP!)

  http://netdesign.zone

- Forum https://ccie.pl – tutaj też odpowiadamy na pytania! ☺
- Zapisz się na kolejne webinaria: https://plnog.pl/webinaria.html

Q&A

Jak zbudować profesjonalny styk z internetem cz. 2?



Łukasz Bromirski
lukasz.bromirski@gmail.com

PLNOG Webinaria, 10.X.2017