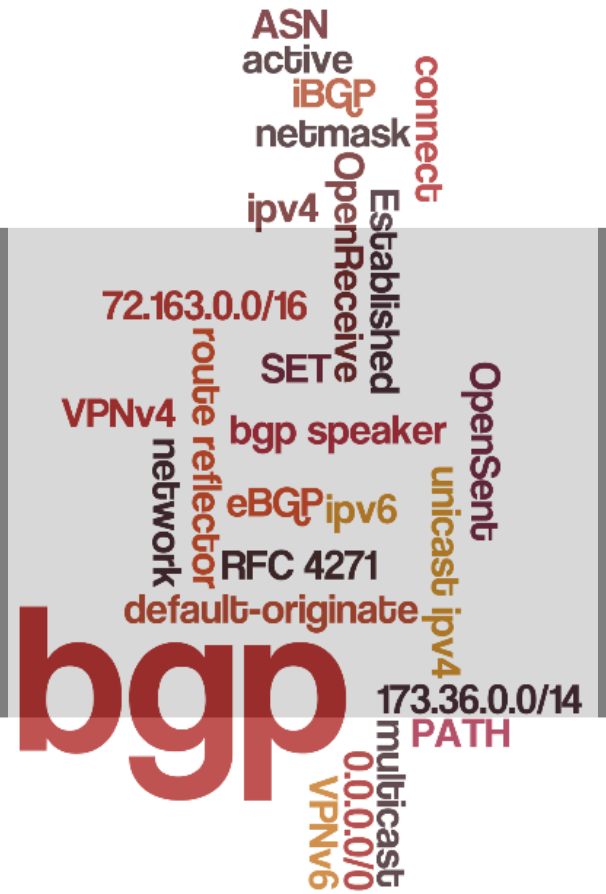
Jak zbudować profesjonalny styk z internetem?



Łukasz Bromirski
lukasz.bromirski@gmail.com

PLNOG Webinaria, 20.IX.2017

Plan na dzisiaj

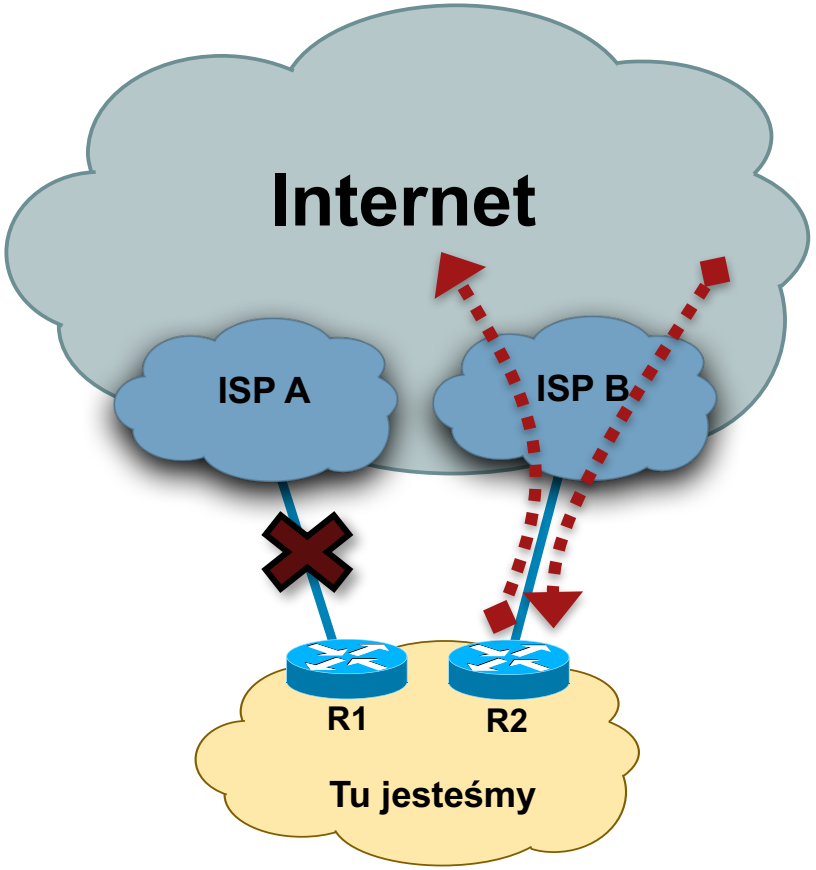
- Co próbujemy uzyskać?
- Co wziąć pod uwagę - platforma
- BGP: co musisz wiedzieć
- Różne scenariusze – omówienie
- Q&A

(profesjonalny) styk z internetem

Co próbujemy uzyskać?

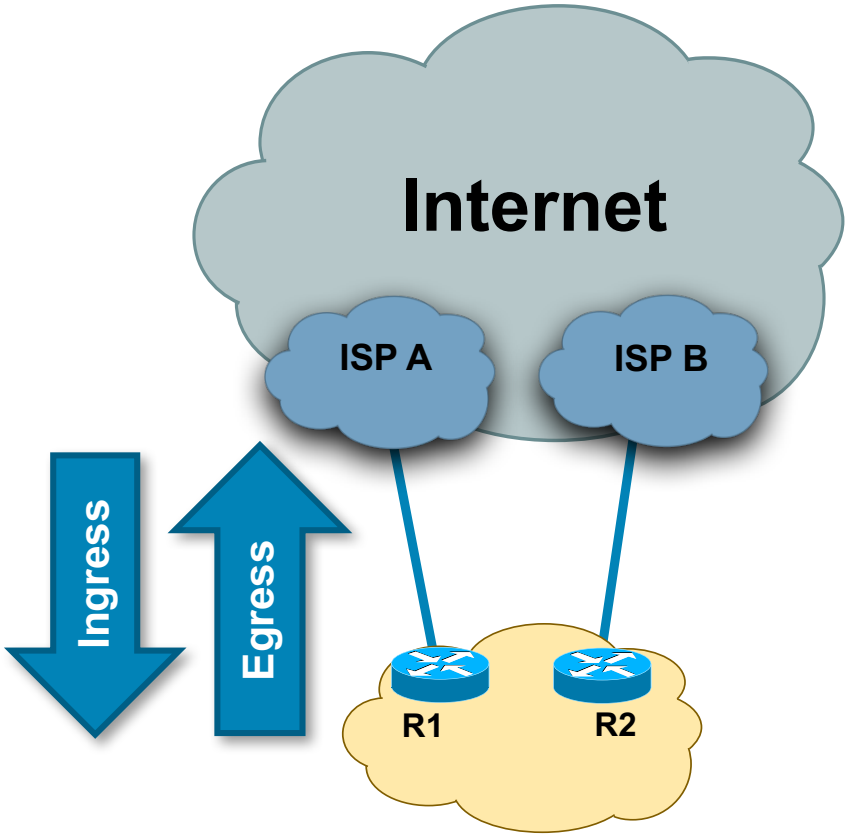
Chcemy być dostępni i niezależni

- Jedno łącze = nie ma problemu
- Dwa i więcej łącz:
  - Jeden router
  - Dwa i więcej routerów



Czy zawsze potrzebujemy BGP?

- Nie



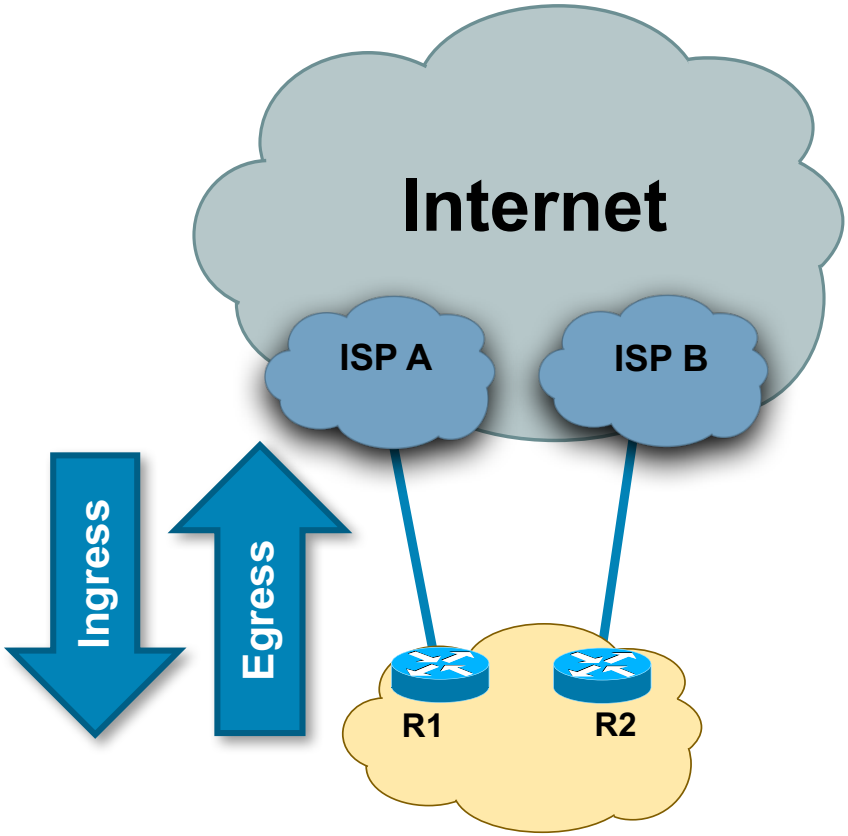
Czy zawsze potrzebujemy BGP?

- Jedno łącze = routing statyczny

  Pobranie out-of-band pełnej tablicy BGP daje pewne zalety (telemetria, bezpieczeństwo) ale żadnych z perspektywy routingu

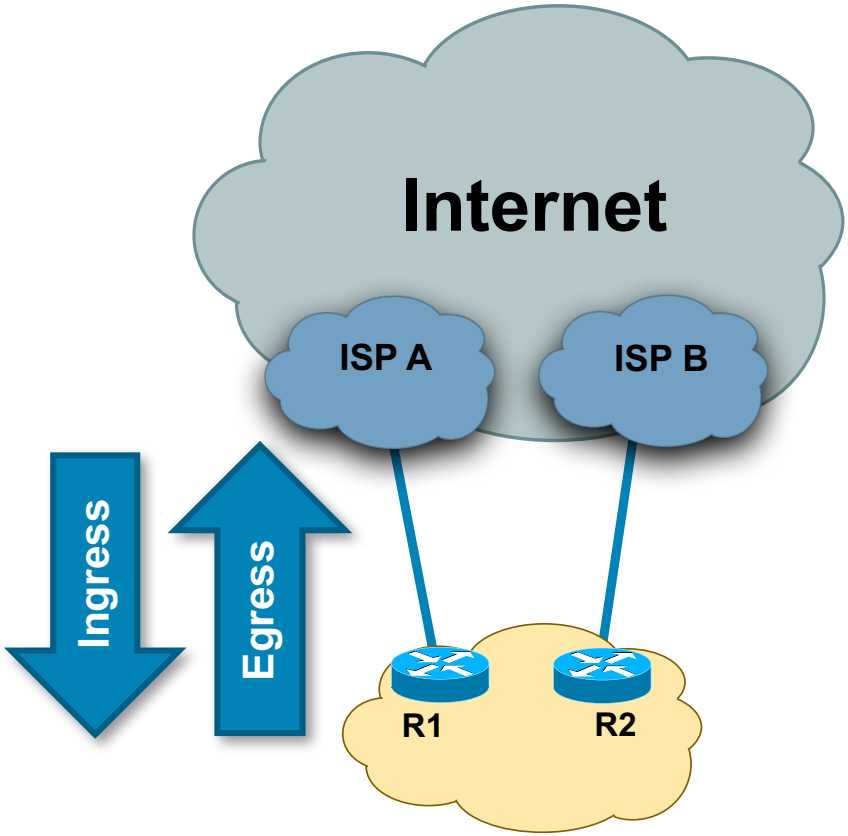
- Dwa łącza = routing statyczny też może zdać egzamin

  Np. dwie trasy domyślne (ang. default) dla ruchu wychodzącego i trasy do naszych prefiksów od operatorów



Ale BGP nie jest takie złe
(ani szczególnie droższe)

- Znakomita większość operatorów oferuje BGP nawet dla jednego łącza
- Implementacje BGP, w tym open source, są coraz lepsze
- BGP zawsze można realizować ”z boku” – o tym dokładniej później



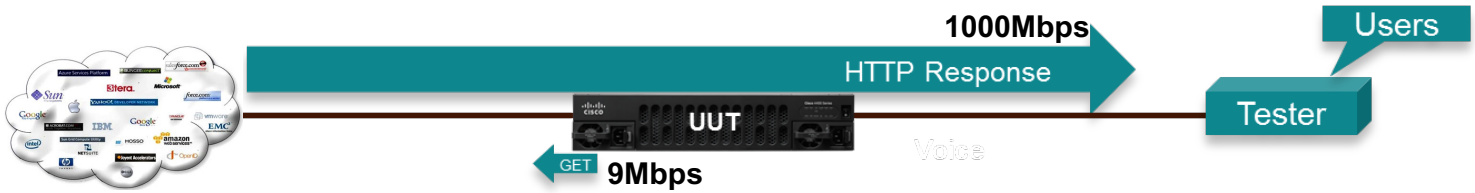
Jaki sprzęt? ”Mam łącze gigabitowe”

- Ruch bezstanowy

  1000+1000

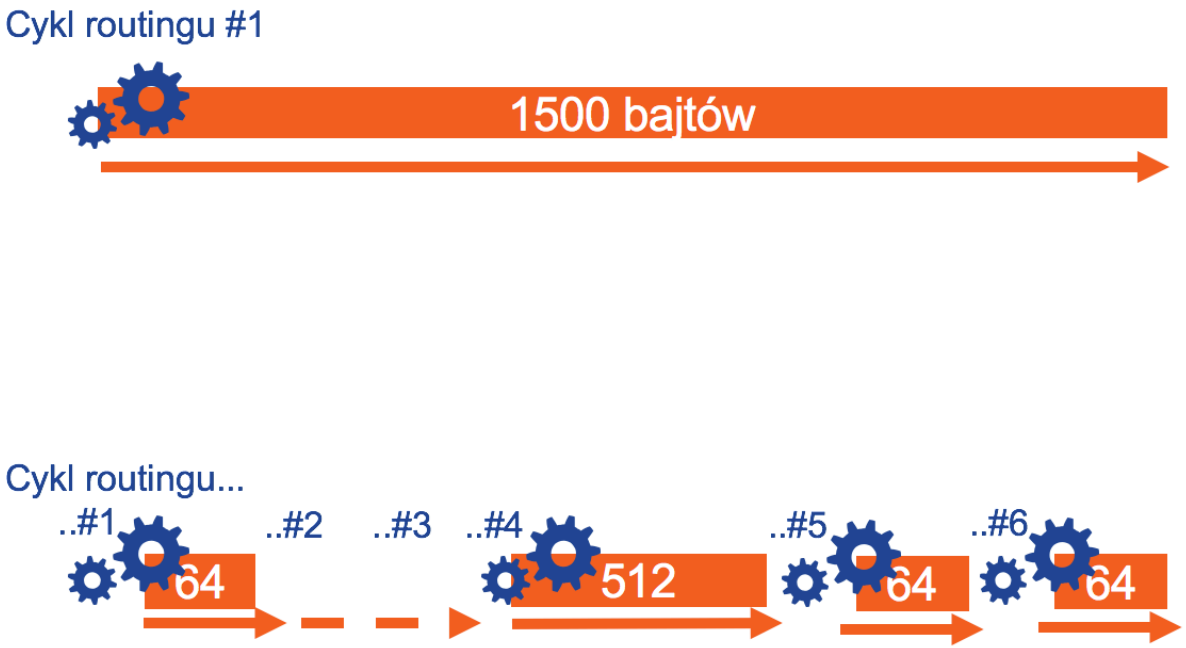
- Ruch ”typowych użytkowników”

  1000+9



Routing programowy vs sprzętowy

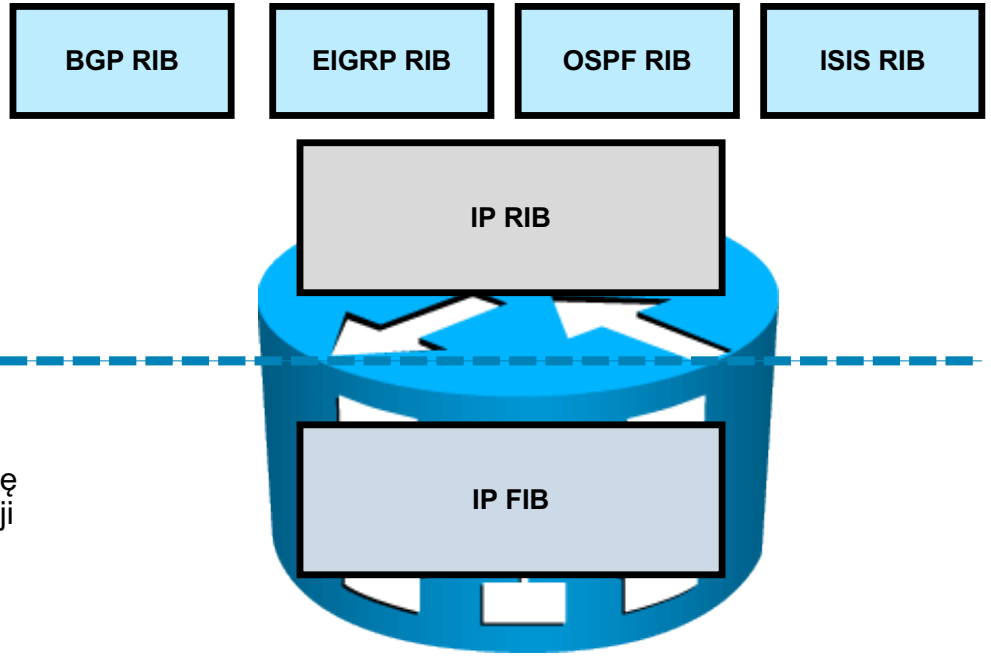
- Każdy CPU będzie musiał jakoś obsłużyć ruch per pakiet
- Pooling przerwań/wątków – to wszystko działa do pewnego momentu
- Gdzie dodatkowe funkcjonalności? QoS? VPN? ACL?



Skalowalność routera – RIB i FIB

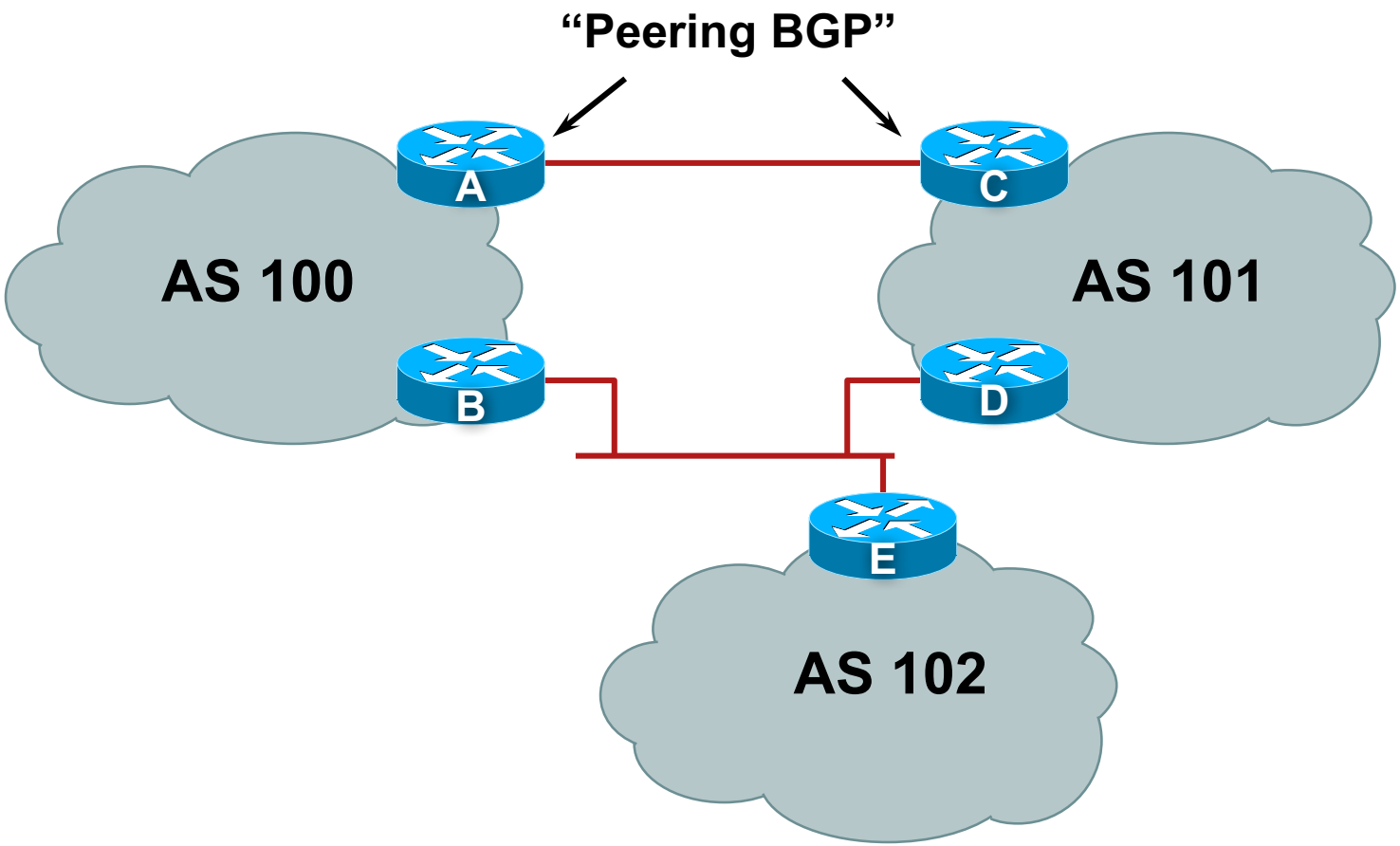
„Control Plane”
CPU + dużo RAMu
(1-2-4-16-32GB)
Miliardy operacji na sekundę
Miliony decyzji na sekundę

„Forwarding Plane”
Specjalizowane ASIC/FPGA
1-2-4mln pozycji na informację
Setki/tysiące miliardów decyzji
na sekundę



BGP:
Co muszę wiedzieć?

BGP – routing pomiędzy wyspami IGP



Numer Systemu Autonomicznego
(Autonomous System, ASN)

- Dwa zakresy

0-65535	(oryginalny zakres 16 bitowy)
65536-4294967295	(32 bitowy zakres - RFC4893)

- Zakresy:

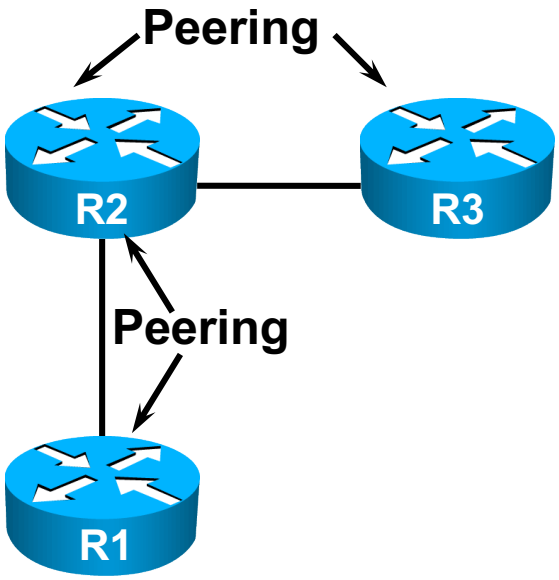
0 and 65535	(zakres zarezerwowany)
1-64495	(publicznie przydzielane wartości)
64496-64511	(dokumentacja - RFC5398)
64512-65534	(do użytku prywatnego)
23456	(wartość używana do zachowania kompatybilności)
65536-65551	(dokumentacja - RFC5398)
65552-4294967295	(publicznie przydzielane wartości)

- Zakres 32-bitowy zapisuje się w formacie “asplain” opisanym w RFC5396

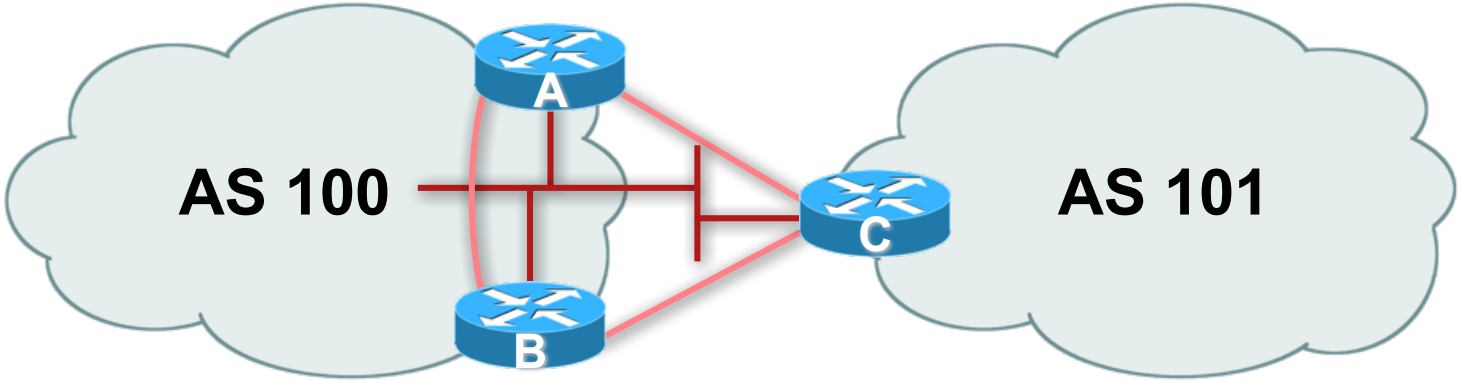
  “asdot” to zapis X.Y, uwaga na dziwactwa producentów (Ericsson!)

BGP z lotu ptaka

- BGP rozmawia z innymi procesami BGP
  - sąsiedzi (peer) nazywani są sąsiadami (neighbor)
  - używa połączenia TCP na port 179 (firewalle!)
- Wybiera najlepszą ścieżkę
  - Zgłasza do zainstalowania w tablicy RIB
  - Rozgłasza najlepsze ścieżki do sąsiadów
- Wiadomości UPDATE oprócz prefiksu zawierają atrybuty
  - Za pomocą polityk routingu w BGP można wpłynąć na atrakcyjność tras, a zatem – wybrać inne trasy jako najlepsze

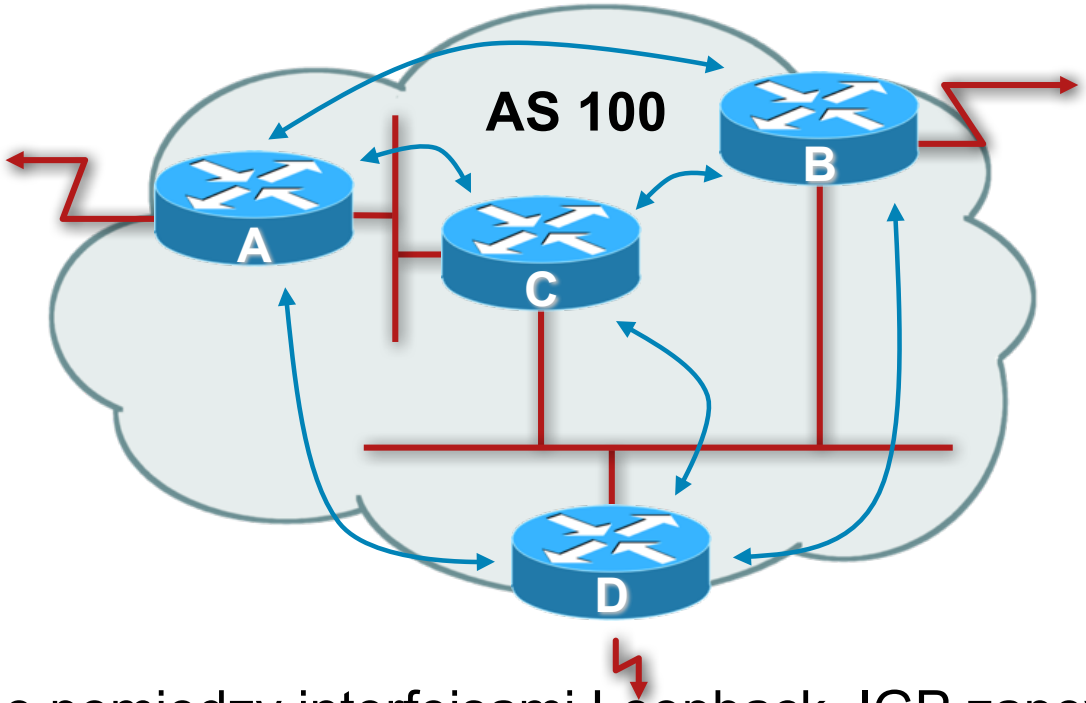


Dwa typy sesji BGP – External BGP



- Dwaj sąsiedzi w różnych ASNach
- Generalnie bezpośrednio połączeni (są wyjątki)
- Nie uruchamiaj IGP na interfejsach zewnętrznych!

Dwa typy sesji BGP – Internal BGP



- Sesje zwykle pomiędzy interfejsami Loopback, IGP zapewnia osiągalność
- Każdy sąsiad iBGP musi być połączony z każdym innym

  Route Reflectory lub Konfederacje – wyjątki dla skalowania ASNu

BGP – wybór najlepszej ścieżki

- Trasa jest nieważna, jeśli nie mamy trasy do jej next-hopa
- Najwyższa waga (weight – specyficzne dla Cisco)
- Najwyższa lokalna preferencja (LOCAL PREFERENCE)
- Trasa lokalnie wprowadzona do BGP (origin)
- Najkrótsza ścieżka AS_PATH
- Najniższy kod powstania trasy (origin) – IGP < EGP < incomplete
- Najniższa wartość MED
- Lepsza ścieżka eBGP niż iBGP
- Najniższy koszt IGP do next-hopa
- Trasa od najniższego router-id
- Minimalna długość pola listy klastrów
- Trasa od sąsiada o najniższym IP

BGP – wybór najlepszej ścieżki

- Trasa jest nieważna, jeśli nie mamy trasy do jej next-hopa
- Najwyższa waga (weight – specyficzne dla Cisco)
- **Najwyższa lokalna preferencja (LOCAL PREFERENCE)**
- Trasa lokalnie wprowadzona do BGP (origin)
- **Najkrótsza ścieżka AS_PATH**
- Najniższy kod powstania trasy (origin) – IGP < EGP < incomplete
- **Najniższa wartość MED**
- **Lepsza ścieżka eBGP niż iBGP**
- Najniższy koszt IGP do next-hopa
- Trasa od najniższego router-id
- Minimalna długość pola listy klastrów
- Trasa od sąsiada o najniższym IP

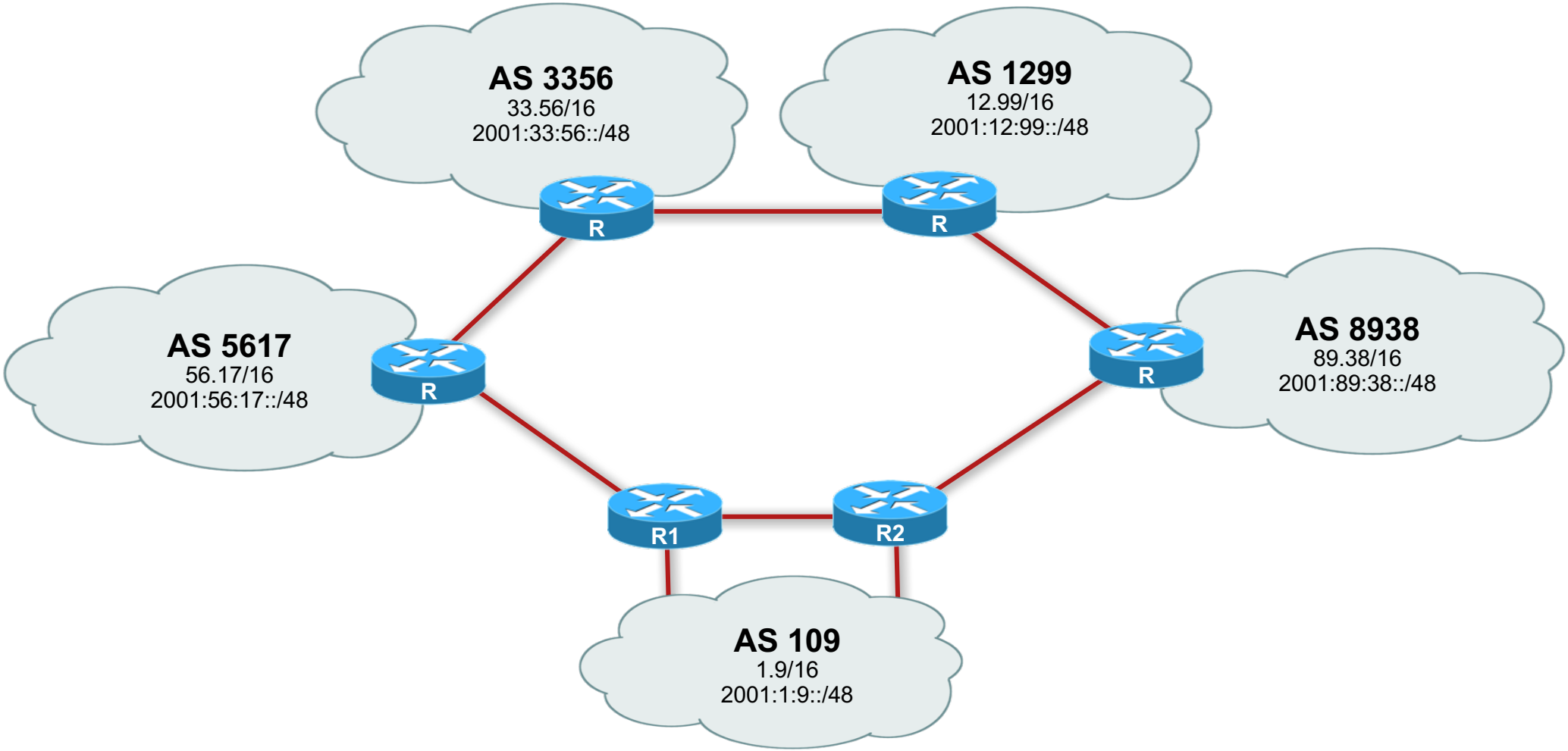
BGP – polityka routingu

- Wszystkie atrybuty BGP mogą posłużyć do nałożenia polityki
  - AS_PATH, wartości communities, local preference, MED są najczęściej używane
- Generalnie możemy:
  - sprawdzić wartość atrybutu lub atrybutów, na tej podstawie
  - ustawić/zmienić atrybuty i
  - zaakceptować lub odrzucić prefiks
- Narzędzia (Cisco IOS/XE/NX-OS/IOS-XR):
  - prefix-list, filter-list, route-map (z community)
  - RPL (IOS-XR)

Inżynieria ruchowa w BGP - ściąga

- Dla ruchu przychodzącego:
  - Dzielenie prefiksów
  - AS_PATH prepend
  - wartości communities
  - MED
- Dla ruchu wychodzącego:
  - LOCAL_PREF
  - Wartości community
  - AIGP

Nasza dzisiejsza topologia



BGP:
Efektywne wykorzystanie zasobów

“Każda pamięć ma ograniczoną pojemność”

- Jak zapewnić pełną wiedzę o routingu w całej sieci bez ogromnych inwestycji w duże platformy?

  routing nakładkowy (MPLS, Segment Routing)

  rozdzielenie RIB i FIB (“Selective Route Download”)

  optymalizacja (“wirtualna agregacja”* i LISP)
- Programowanie FIBu potrafi być wooooooolne (2-3 minuty? 10?)
- Zatem skoncentrujmy się nad rozdzieleniem RIB i FIB

- https://www.nanog.org/meetings/nanog47/presentations/Monday/Francis_VirtualAgg_N47_Mon.pdf i http://tools.ietf.org/html/rfc6769

SRD: Selective RIB download w praktyce

- Warto mieć ogólniejsze trasy w RIB i FIB – inaczej dojdzie do blackholingu lub wysłania ruchu przez trasę default

IOS-XR

```
route-policy block-into-rib
 if destination in (...) then
   drop
 else
   pass
 end-if

router bgp 1
 address-family ipv4 unicast
  table-policy block-into-rib
```

IOS/IOS-XE

```
router bgp 1
 address-family ipv4
  table-map block-into-rib filter

route-map block-into-rib deny 10
```

Chcę wiedzieć więcej!

Parę ważnych rzeczy do zapamiętania

- Bezpieczeństwo i dobre praktyki:

  RPKI
  https://www.ripe.net/manage-ips-and-asns/resource-management/certification/bgp-origin-validation

  MANRS
  https://www.routingmanifesto.org/manrs/

  bogons/blackholing i różnego rodzaju feedy (np. Cisco TALOS)

- Software Internet Router (SIR) – SDN na żywo:

  https://labs.spotify.com/2016/01/26/sdn-internet-router-part-1/

“Można obejrzeć”

- PLNOG #14: Nowości w BGP:
  https://www.youtube.com/watch?v=56H0VF1jV3E
- PLNOG #15: BGP Route Reflectory praktycznie:
  https://www.youtube.com/watch?v=vW21rcFbZVM
- PLNOG #18: CEF, fd.io i historia obsługi ruchu:
  https://www.youtube.com/watch?v=OenaX56LrwQ
- Cisco Live! 2016: Multihoming for Enterprise:
  https://www.ciscolive.com/online/connect/sessionDetail.ww?SESSION_ID=90875
- Warsztaty dla operatorów Philipa Smitha:
  http://www.bgp4all.com.au/dokuwiki/workshops/start

Co dalej po dzisiejszym webinarium?

- Mamy dedykowany kanał poświęcony projektowaniu sieci

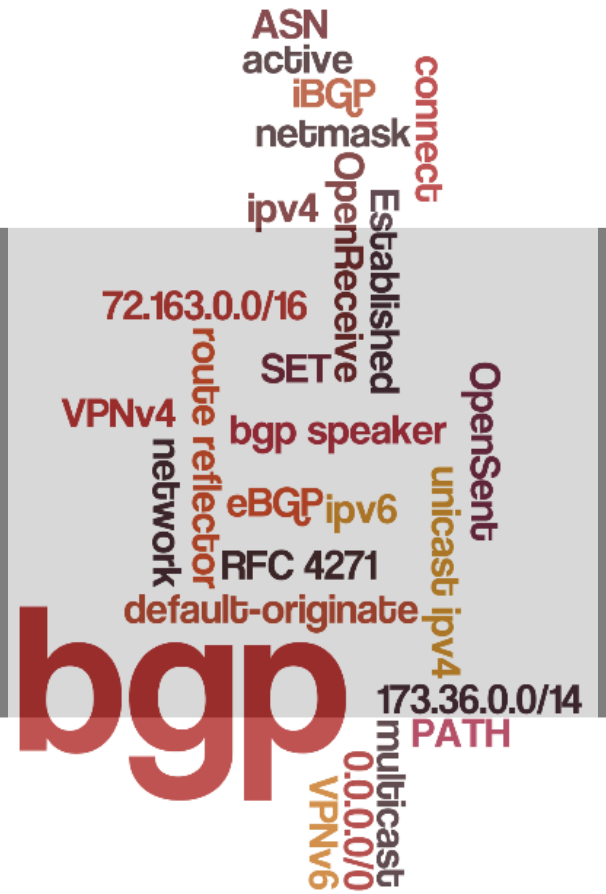
  Napisz do mnie na lukasz.bromirski@gmail.com po zaproszenie
- Repozytorium z topologią .virl oraz konfiguracją Cisco Routem z tej sesji:

  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
- Do zobaczenia na PLNOGu!

Q&A

Jak zbudować profesjonalny styk z internetem?



Łukasz Bromirski
lukasz.bromirski@gmail.com

PLNOG Webinaria, 20.IX.2017