Warsztaty: Protokół BGP

Podstawy i najlepsze praktyki

Andrzej Szulc
andrszul@cisco.com

Agenda

- Protokół BGP – wstęp teoretyczny
- Podstawowa konfiguracja protokołu BGP
- Lab #1
- Konfiguracja konfederacji i route-reflectorów
- Lab #2
- Praktyka konfiguracji styku z Internetem
- Lab #3

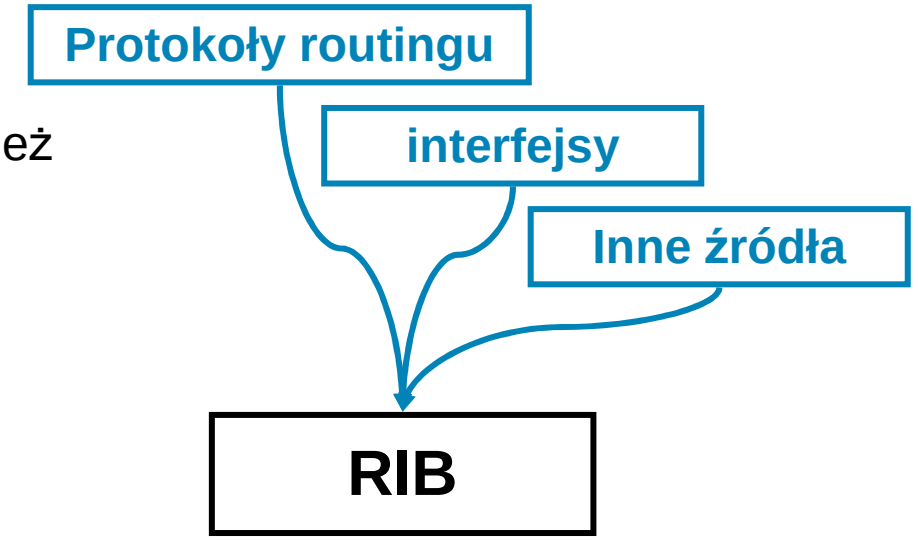
Routing w Cisco IOS

Routing IP
O czym mówimy?

- Routing IP to decyzja wykonana na podstawie adresu docelowego pakietu IP
- Router podejmuje tą decyzję na podstawie tablicy FIB – Forwarding Information Base
- Procesy realizujace routing dynamiczny utrzymują zwykle swoją tablicę – RIB – Routing Information Base – z której najlepsze wpisy eksportowane są do FIB

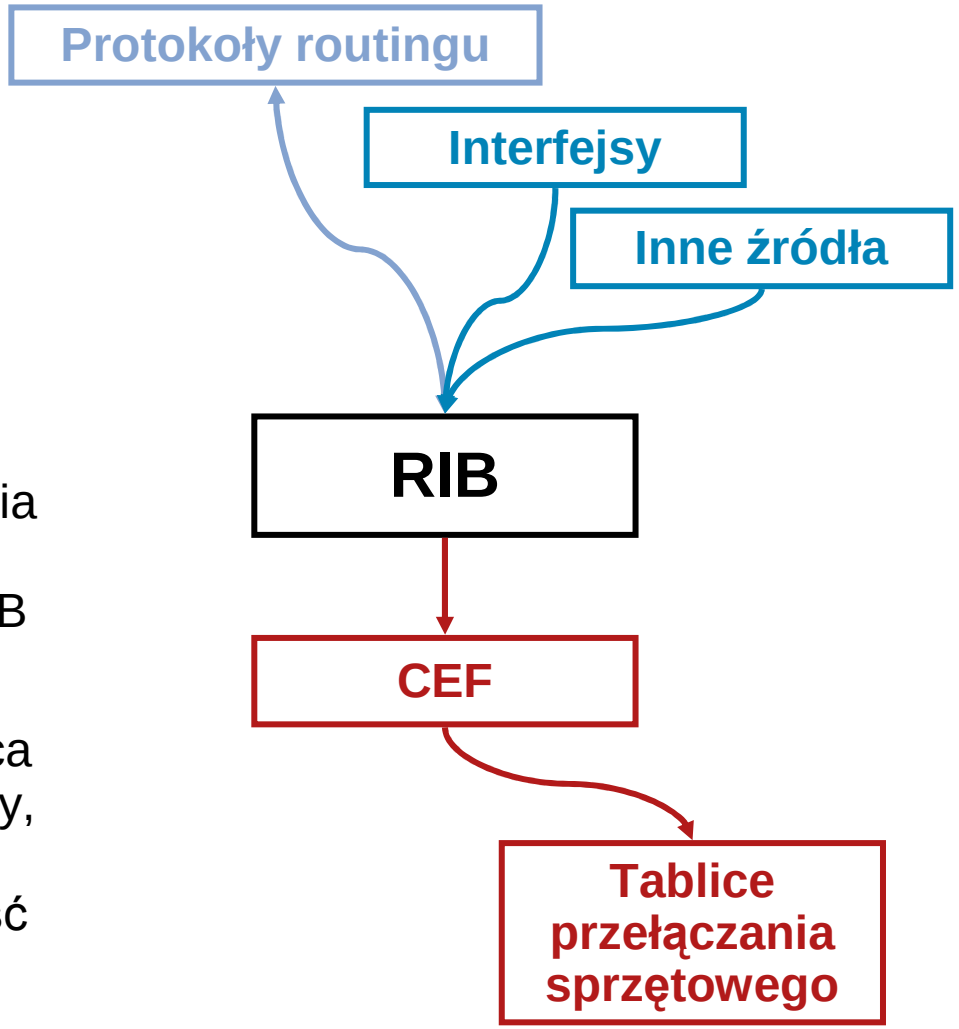
Budowa tablicy routingu

- Routing Information Base (RIB)
- Protokoły routingu
  - –instaluje trasy w RIB
  - –routing statyczny jest również „protokołem routingu”
- Interfejsy
  - –instaluje trasy w RIB
- Inne źródła
  - –instaluje trasy w RIB



Budowa tablicy routingu

- Protokoły routingu
  - –Pobierają trasy z RIB w celu redystrybucji
- Cisco Express Forwarding (CEF)
  - –Pobiera trasy z RIB
  - –Uzupełnia tablice przełączania sprzętowego bazując na informacjach pobranych z RIB
  - –RAM (RIB) w przełącznikach zawiela o wiele więcej miejsca niż układy sprzętowe (TCAMy, przechowujące FIB) – dla ruchu unicast IPv4 pojemność to 8-12k prefiksów

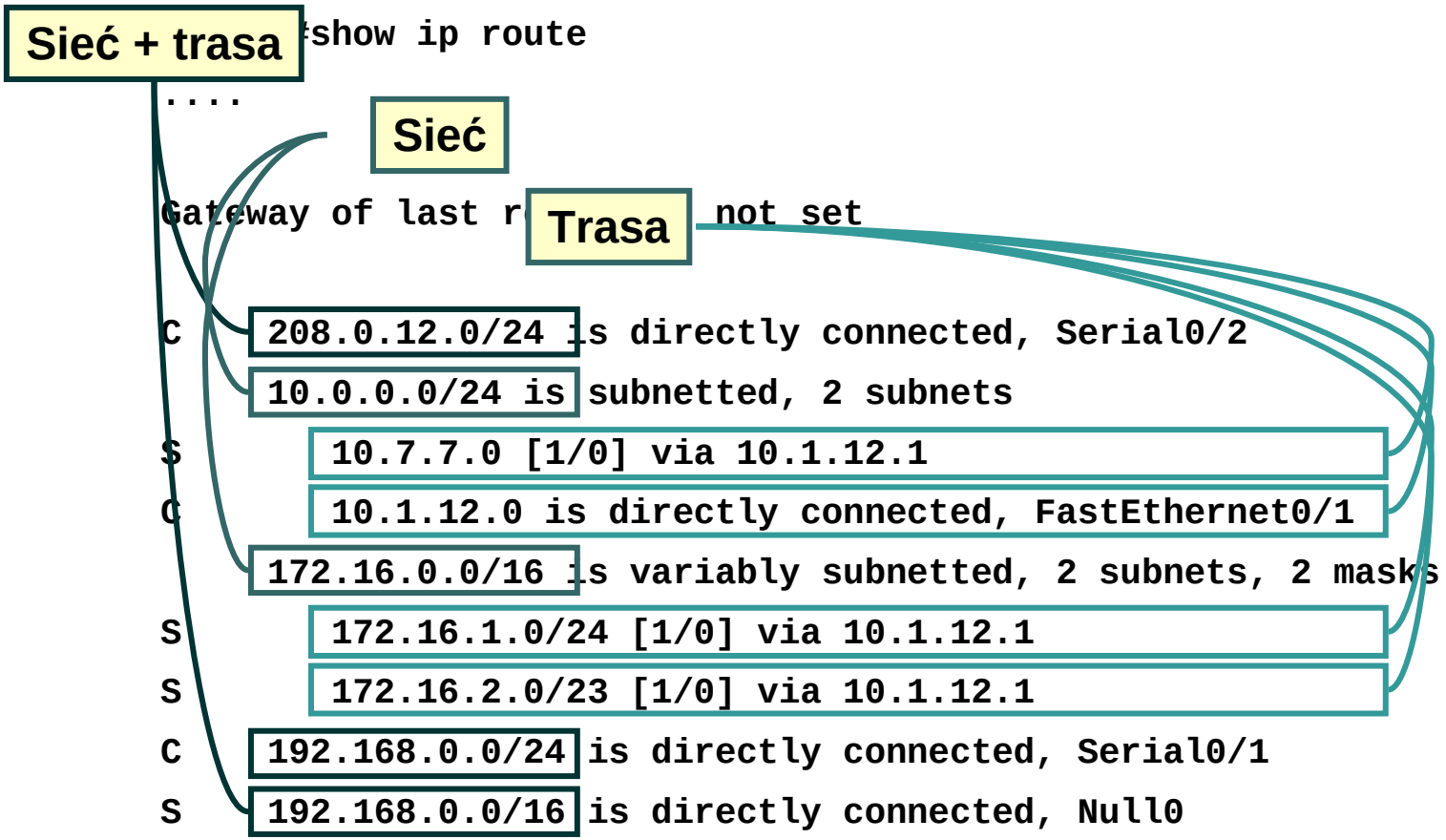


Budowa tablicy routingu

```
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route
```

```
C    208.0.12.0/24 is directly connected, Serial0/2
     10.0.0.0/24 is subnetted, 2 subnets
S       10.7.7.0 [1/0] via 10.1.12.1
C       10.1.12.0 is directly connected, FastEthernet0/1
     172.16.0.0/16 is variably subnetted, 2 subnets, 2 masks
S       172.16.1.0/24 [1/0] via 10.1.12.1
S       172.16.2.0/23 [1/0] via 10.1.12.1
C    192.168.0.0/24 is directly connected, Serial0/1
S    192.168.0.0/16 is directly connected, Null0
```

Budowa tablicy routingu



Budowa tablicy routingu

Główne sieci z podsieciami wyświetlane są jako jedna sieć z wieloma trasami

Pojedyncze trasy z maską natywną wyświetlane są jako pojedynczy wpis

```
C    192.168.12.0/24 is directly connected, Serial0/2
     10.0.0.0/24 is subnetted, 2 subnets
S       10.7.7.0 [1/0] via 10.1.12.1
C       10.1.12.0 is directly connected, FastEthernet0/1
     172.16.0.0/16 is variably subnetted, 2 subnets, 2 masks
S       172.16.1.0/24 [1/0] via 10.1.12.1
S       172.16.2.0/23 [1/0] via 10.1.12.1
C    192.168.0.0/24 is directly connected, Serial0/1
S    192.168.0.0/16 is directly connected, Null0
```

Maski natywne tras i ich supersieci wyświetlane są jako różne sieci

Budowa tablicy routingu

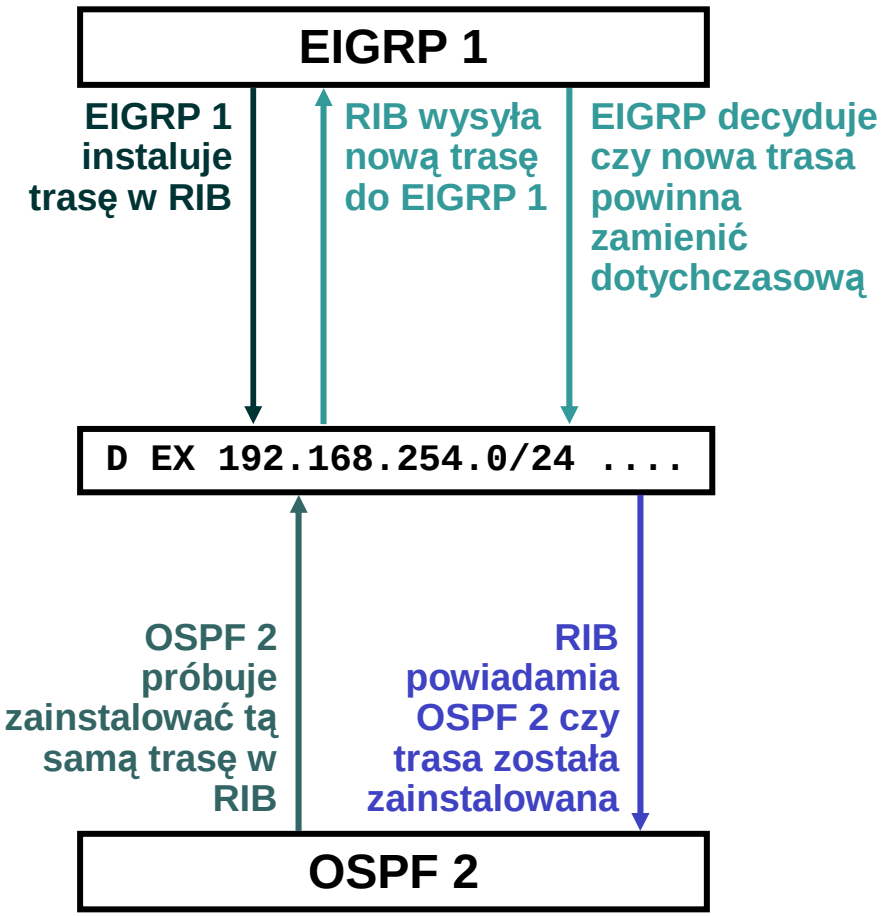
- Dystans administracyjny używany jest do określenia która trasa z pośród wielu możliwych tras jest instalowana w tablicy routingu
- Licznik wyświetla czas od ostatniej modyfikacji trasy
  - –Rekalkulacja EIGRP włączając utratę trasy alternatywnej resetuje ten licznik
  - –Przeliczenie drzewa SPF w OSPF resetuje ten licznik
  - –Przeliczenie drzewa SPF w IS-IS resetuje ten licznik

Dystans administracyjny

Protokół	Dystans administracyjny
Connected	0
Trasa statyczna	1
EIGRP summary	5
eBGP	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
ODR	160
External EIGRP	170
iBGP	200

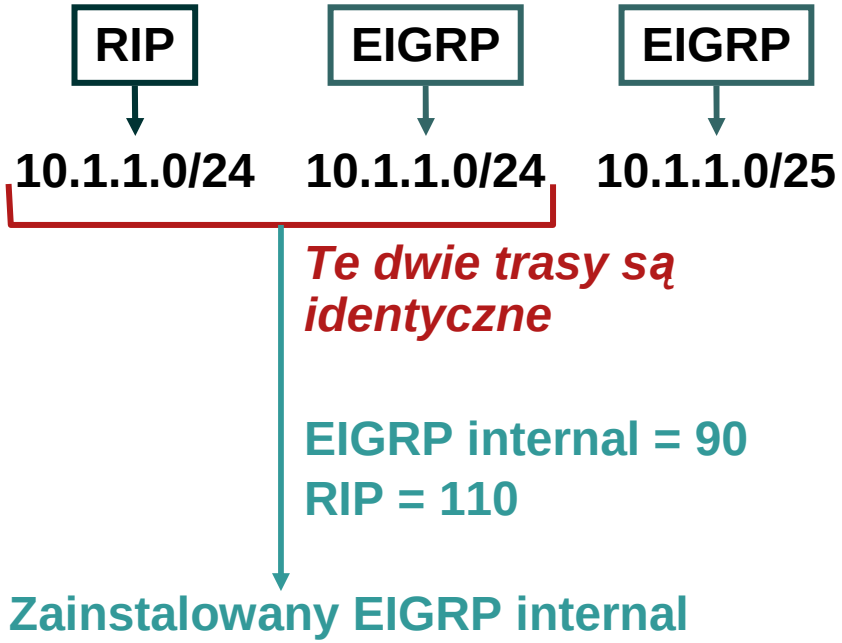
Wybór trasy

- Jak RIB decyduje z którego źródła trasa jest najlepsza?
- Tak naprawdę nie robi tego!
- Każda trasa oznaczona jest procesem z którego pochodzi
- Gdy proces routingu próbuje zainstalować trasę nakładającą się z już istniejącą w RIB, RIB pozwala zdecydować właścicielowi wpisu
- Generalnie decyzja podejmowana jest na podstawie dystansu administracyjnego obu procesów



Wybór trasy

- Jak dystans administracyjny określa która trasa powinna być zainstalowana?
- Tylko identyczne trasy są porównywane
  - –....identyczne prefiksy z różną długością maski nie są tą samą trasą
- W tablicy instalowana jest trasa z procesu o najniższym dystansie administracyjnym



Protokół BGP

Border Gateway Protocol – version 4

Protokół BGP
Wstęp

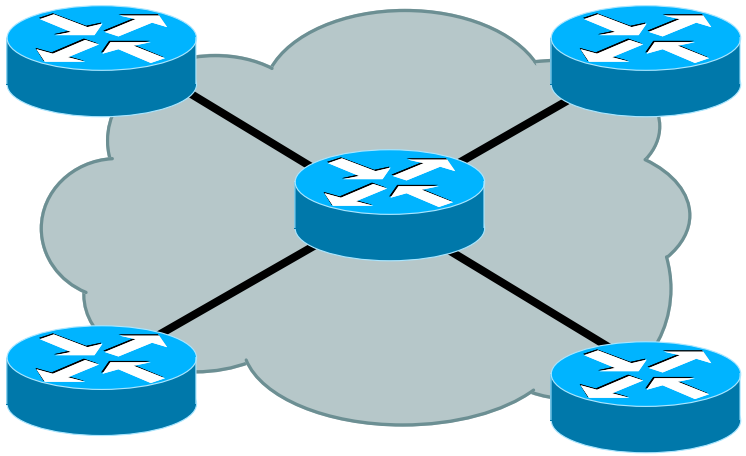
- BGP jest protokołem typu dystans-wektor
- Posługuje się bogatą metryką (atrybuty prefiksu)
- Posługuje się protokołem TCP jako transportowym – zakłada zatem istnienie innego protokołu (IGP) do zapewnienia transportu wewnątrzdomenowego
- Za pomocą pakietów keepalive utrzymuje sąsiedztwa
- Uaktualnienia rozsyłane są tylko w przypadku zmian:
  - co 5 sekund do sąsiadów wewnętrznych
  - co 30 sekund do sąsiadów zewnętrznych

Protokół BGP
Wstęp

- Protokół routingu używany do wymiany informacji o osiągalności sieci (prefiksów) pomiędzy systemami autonomicznymi
  - klasy EGP – Exterior Gateway Protocol
- Do niedawna RFC1771 + rozszerzenia, od stycznia 2006 obowiązuje RFC4271:
  - http://www.ietf.org/rfc/rfc4271.txt
  - RFC4276 opisuje raport implementacyjny RFC4271
  - RFC4277 dodatkowo opisuje doświadczenia z różnymi implementacjami

Protokół BGP

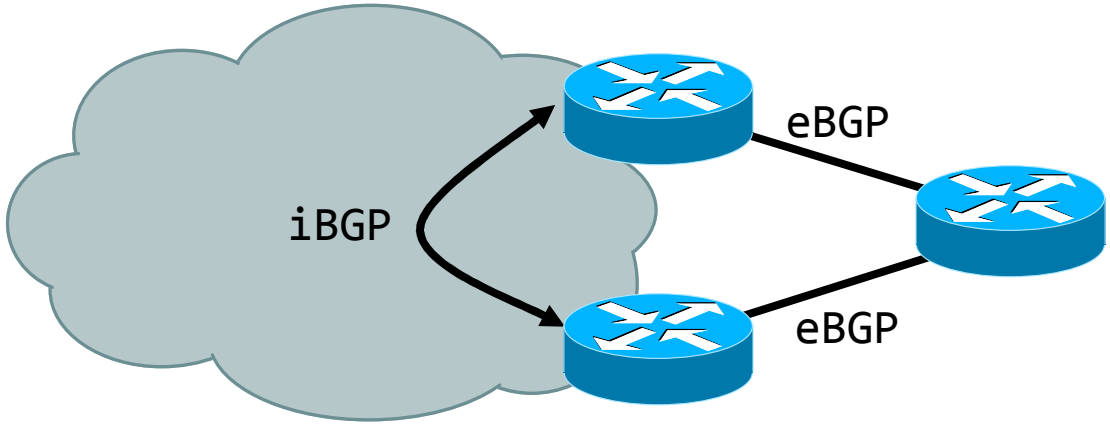
System autonomiczny (AS – Autonomous System)



AS100

- Zwykle jedna firma/organizacja, zarządzana przez jedną grupę ludzi
- Jedna polityka routingu wewnętrznego i zewnętrznego
- ASN 0 i 65535 zarezerwowane, 23456 również (do przejścia na 32-bitową notację), natomiast 1-64511 zarządzane centralnie (publiczne)
- 64512-65534 do prywatnego użytku
- Obecnie zarejestrowano trochę ponad 39935 ASN, z czego około 29000 jest widocznych w globalnych tablicach routingu (12000 rozgłasza tylko jeden prefiks)

Protokół BGP
Internal BGP a External BGP



- iBGP – sesje pomiędzy routerami w tym samym AS
  - –wszystkie routery wewnątrz AS muszą nawiązać sesje każdy z każdym (można to obejść przez konfederacje/klastry)
- eBGP – sesje pomiędzy routerami w różnych AS
  - –domyślnie połączenie bezpośrednie, należy wprost wskazać że połączenie jest multihop

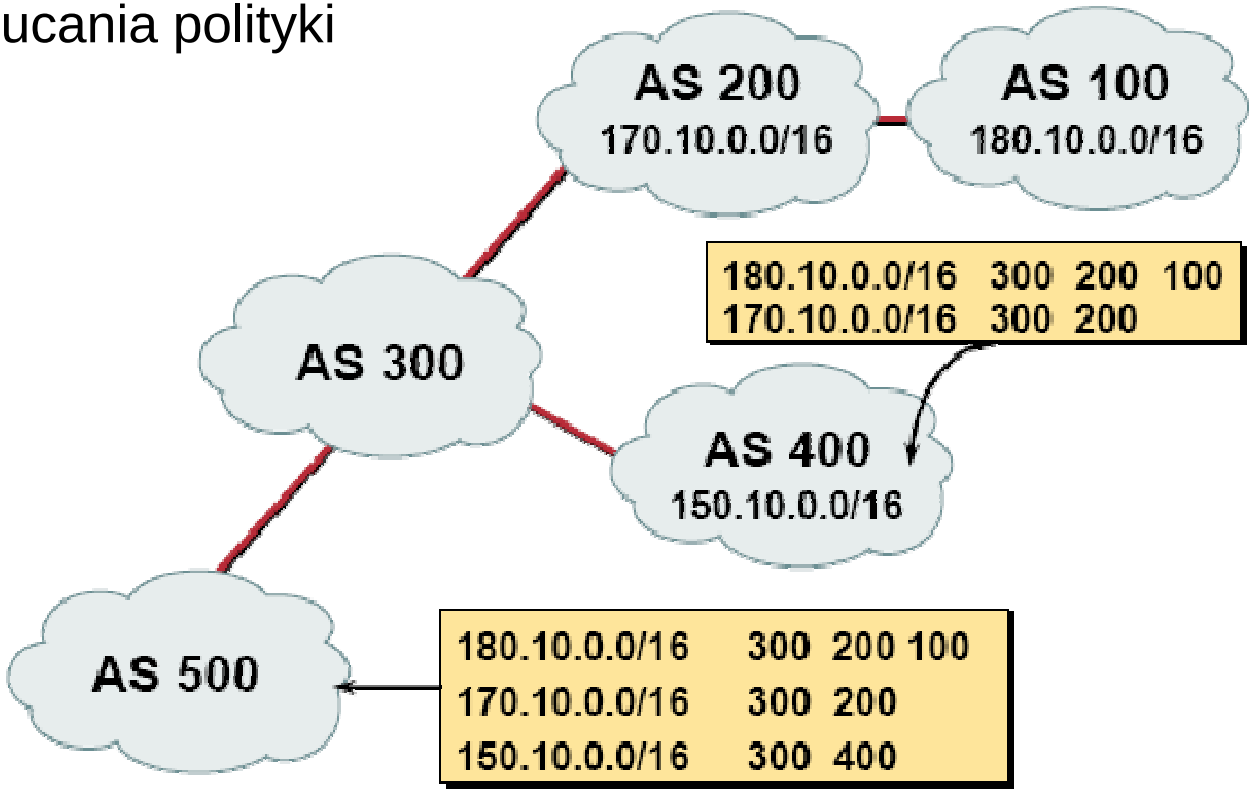
Protokół BGP

Atrybuty pozwalające wpływać na wybór trasy

- –1: ORIGIN
- –2: AS-PATH
- –3: NEXT-HOP
- –4: MED
- –5: LOCAL_PREF
- –6: ATOMIC_AGGREGATE
- –7: AGGREGATOR
- –8: COMMUNITY
- –9: ORIGINATOR_ID
- –10: CLUSTER_LIST
- –14: MP_REACH_NLRI
- –15: MP_UNREACH_NLRI
- –

Protokół BGP
Atrybut – AS-Path

- Trasa, jaką prefiks przeszedł
- Mechanizm zapobiegający pętlom
- Narzędzie narzucania polityki



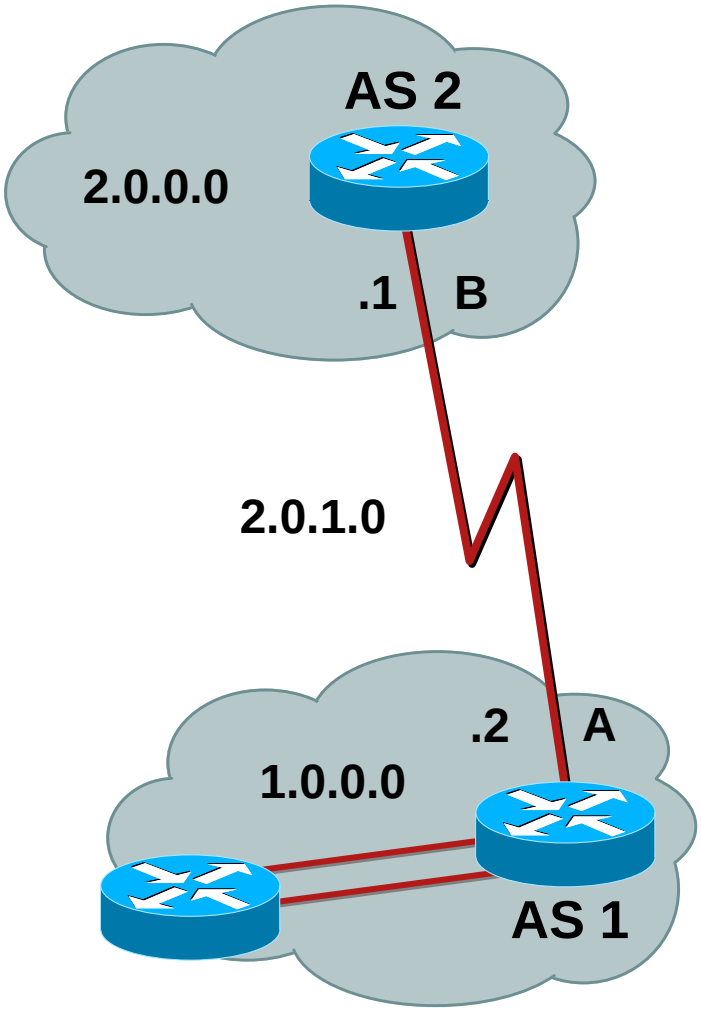
Protokół BGP
Atrybut – next-hop

- Wskaźnik na następny węzeł w drodze do sieci docelowej

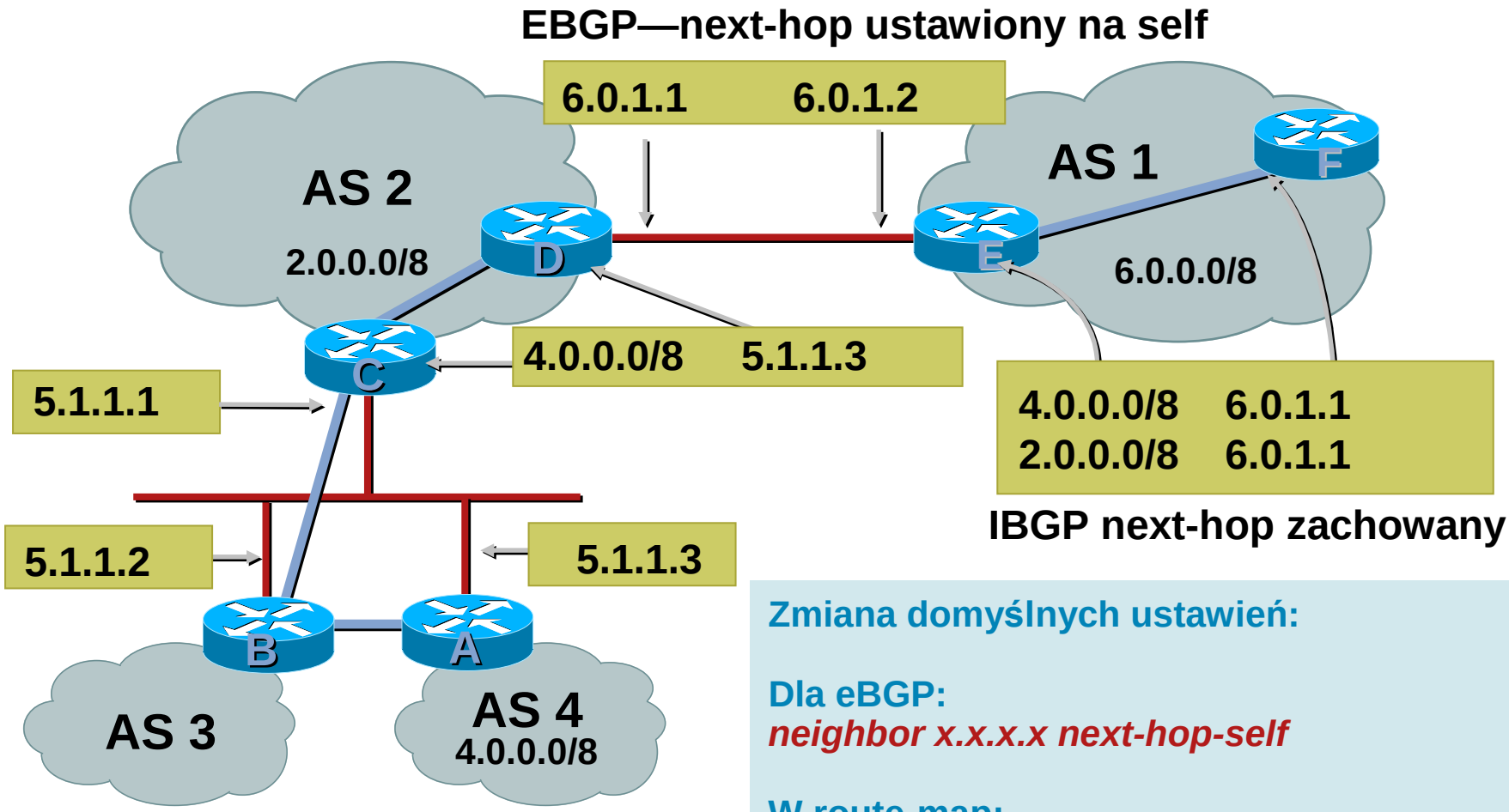
  Dla eBGP – adres sąsiada

  Dla iBGP – pole next-hop rozgłoszone przez eBGP

- IGP powinny przenosić pola next-hop (routery wykonują rekursywny lookup w tablicy FIB)



Protokół BGP
Atrybut – next-hop

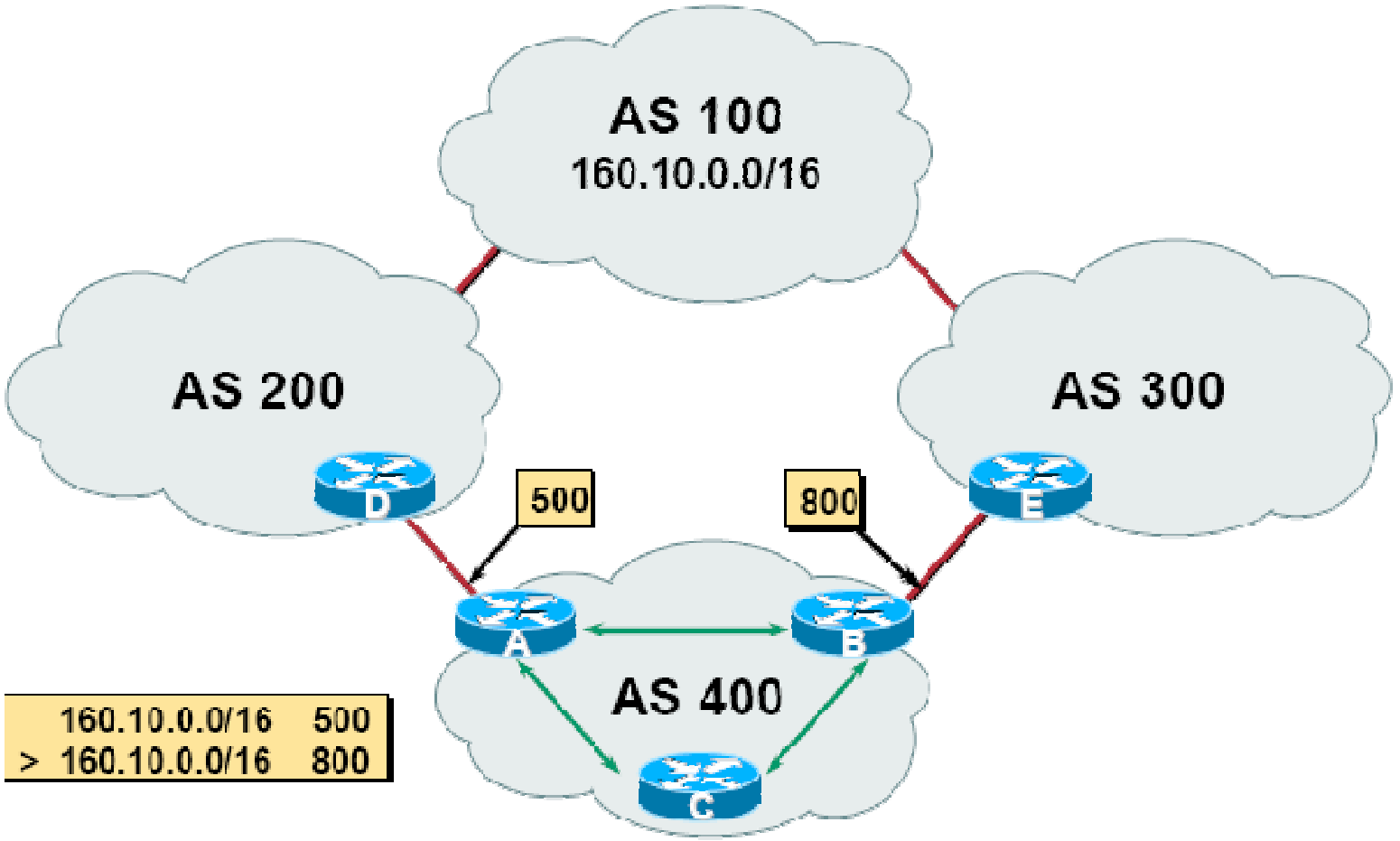


Zmiana domyślnych ustawień:

Dla eBGP:
neighbor x.x.x.x next-hop-self

W route-map:
set ip next-hop { A.B.C.D | peeraddress}

Protokół BGP
Atrybut – local-preference



Protokół BGP
Atrybut – community

- Opisane w RFC1997, mogą b
  pomiędzy ASami (transitive) i
  - –32 bitowa wartość dodatnia
- Dla wygody zapisywana jako
  rozdzielone dwukropkiem
  - –Standardowo <NUMER-AS>:XX
  - –Na przykład: 64999:666
- Bardzo przydatne i często sp
  peeringach do sterowania pol
  oznaczonego prefiksu

```
================================================
Communities used in AS5617
================================================

5617:103 - prefix is not announced to foreign peers

We do not accept all RFC1997 communities (no-export etc.)

communities for specific link are 5617:ab0x :
x=0 for "do not advertise"; x=1,2,3 for "prepend 1,2,3 times"

for AS1299 Telia           5617:110x
for AS5511 OpenTransit     5617:120x

for AS8501  POL34:         5617:210x
for AS16283 Lodman:        5617:220x
for AS8664  ICM:           5617:240x
not used                   5617:250x
for AS8890  Kampus Ochota: 5617:260x
for AS8364  Pozman:        5617:270x
for AS8508  Silweb:        5617:280x
for AS8970  WASK:          5617:290x

================================================

5617:997 - BLACKHOLING  COMMUNITY

Blackhole community for \32 routes available for peers
advertising only one AS

================================================

================================================

ENTRY POINT INDICATORS IN AS5617

================================================

5617:3385: Bialystok
5617:3352: Bydgoszcz
5617:3358: Gdansk
5617:3332: Katowice
5617:3341: Kielce
5617:3312: Krakow
5617:3342: Lodz
5617:3381: Lublin
5617:3389: Olsztyn
5617:3361: Poznan
5617:3317: Rzeszow
5617:3391: Szczecin
5617:3322: Warszawa
5617:3371: Wroclaw
```

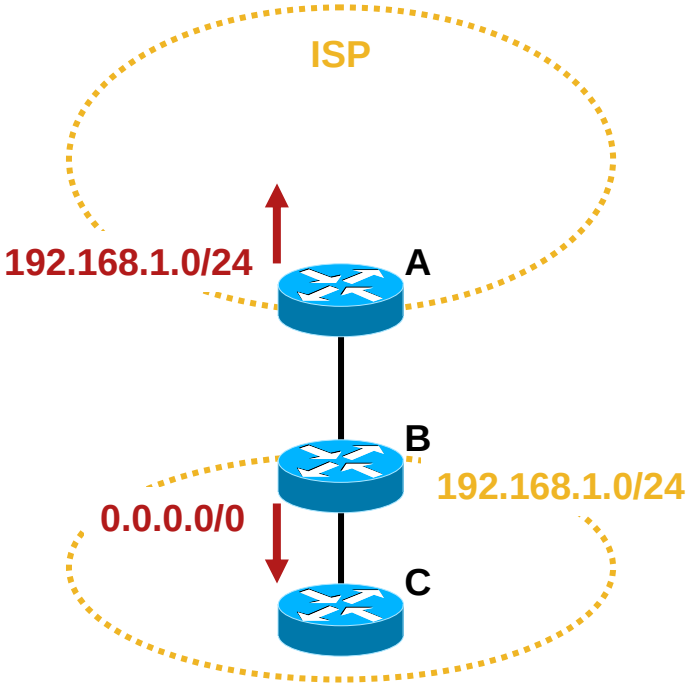
Protokół BGP
Jak BGP wybiera najlepszą trasę? (wersja skrócona)

- Najwyższa waga (lokalna dla routera)
- Najwyższy local preference (w ramach AS)
- Najkrótsza ścieżka AS-Path
- Najniższy kod pochodzenia (Origin)
  - –IGP < EGP < incomplete
- Najniższa wartość MED (Multi-Exit Discriminator)
- Lepiej trasa z eBGP niż z iBGP
- Najpierw trasa z niższym kosztem wg. IGP do next-hop
- Najniższy router-id routera BGP
- Najniższy adres peer’a

–(pełna lista w RFC4271)

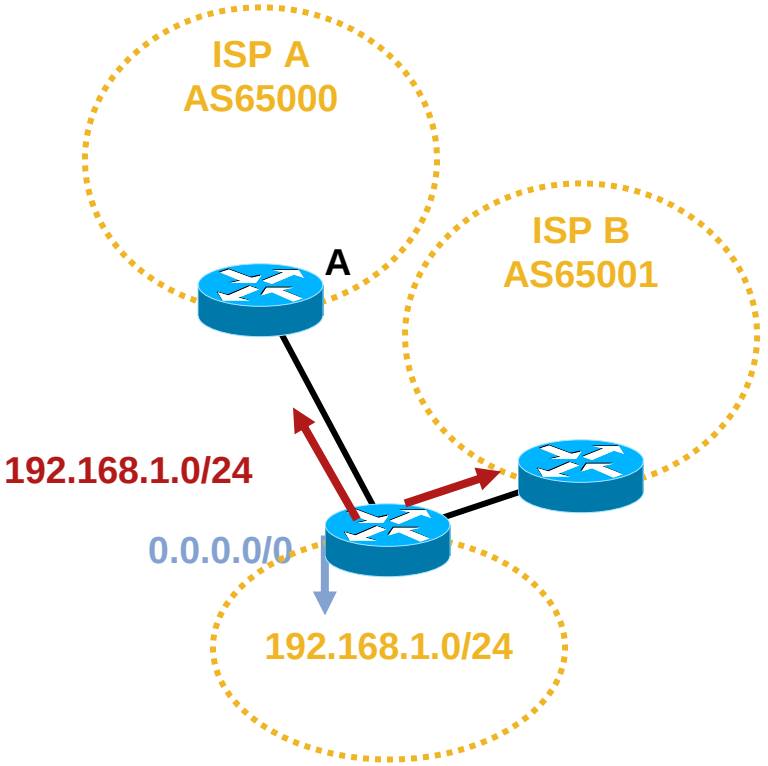
Czy na pewno muszę mieć BGP?

- Pojedyncze połączenie do Internetu - NIE
  - –w sieci używasz domyślnej trasy (statycznie lub protokół routingu)
  - –Twój ISP zapewnia widoczność i osiągalność przydzielonej Ci adresacji IP
- Nawet jeśli łącza są dwa lub trzy do tego samego ISP, BGP nie jest potrzebne
  - –wiele tras domyślnych na różne IP po stronie Twojej i ISP



Czy na pewno muszę mieć BGP?

- Jeśli jesteś połączony do dwóch różnych ISP – BGP jest pożądane – rozgłasza do obu swoją pulę adresów, Internet widzi ją przez obu ISP
- Nie oznacza to, że musisz pobierać wszystkie światowe prefiksy
  - –...pozwala to jednak ‘świadomie’ kształtować własną politykę routingu dużo dokładniej



Podstawowa konfiguracja protokołu BGP

Prezentacja

Konfiguracja protokołu BGP
Podstawy

- Na jednym routerze może pracować tylko jeden proces BGP w konkretnym numerze AS
- Aby usunąć źle zdefiniowany numer AS należy wykasować konfigurację protokołu BGP i zacząć konfigurację od nowa (no router x)

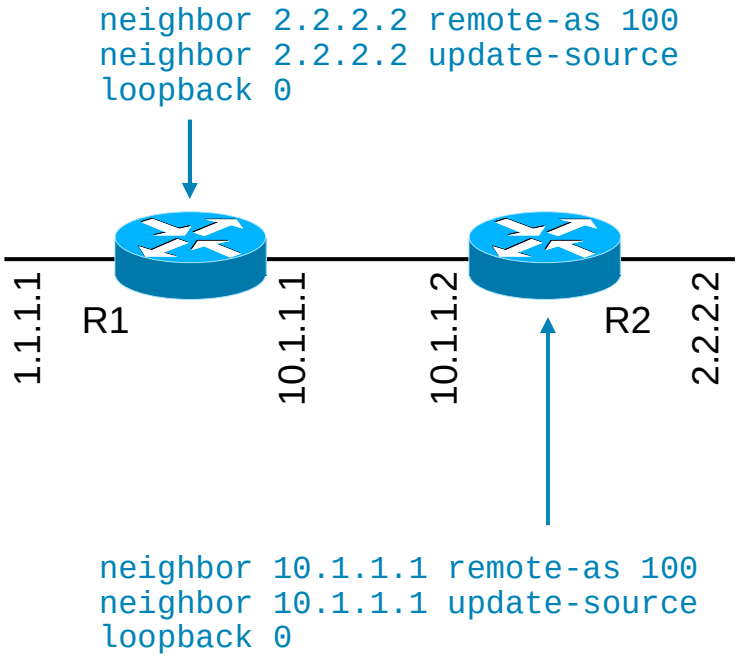
```
router(config)# router bgp 100
```

Konfiguracja protokołu BGP
Uwaga konfiguracyjna

- BGP używa protokołu TCP, sesja zestawiana jest na port 179
- Sąsiedzi wymieniają wiadomości OPEN zawierające podstawowe informacje:
  - –Router ID
  - –AS #
  - –Capabilities
- Adresy IP:
  - –Docelowy adres sąsiada wskazuje się przez neighbor x.x.x.x
  - –Źródłowy IP to domyślnie adres interfejsu w stronę sąsiada
  - –Źródłowy IP można wskazać wprost neighbor x.x.x.x update-source interface

Konfiguracja protokołu BGP
Uwaga konfiguracyjna

- Obie strony muszą zgodzić się co do adresów IP
- Jeśli R1 i R2 nie zgodzą się co do adresów BGP rozłączy sesje TCP



Konfiguracja protokołu BGP
Uwaga konfiguracyjna

- R2 próbuje otworzyć sesję do R1

```
BGP: 10.1.1.1 open active, local address 2.2.2.2
```

- R1 odrzuca sesję z powodu złego adresu źródłowego
- debug ip bgp na R1 pokazuje:

```
BGP: 2.2.2.2 passive open to 10.1.1.1

BGP: 2.2.2.2 passive open failed - 10.1.1.1 is
 not update-source Loopback0's address (1.1.1.1)
```

Konfiguracja protokołu BGP
Podstawy

- Sesje pomiędzy routerami BGP konfiguruje się wskazując sąsiadów
- Sąsiad w tym samym AS = iBGP, sąsiad w innym = eBGP

```
router bgp 100
  neighbor 1.1.1.1 remote-as 100
  neighbor 1.1.1.1 description Sesja iBGP
  neighbor 3.3.3.3 remote-as 300
  neighbor 3.3.3.3 description Sesja eBGP
```

Konfiguracja protokołu BGP
Podstawy

- Sesje pomiędzy sąsiadami iBGP domyślnie mają TTL równy 254, pomiędzy eBGP – 1
- Sesja może zaczynać się z dowolnego, wskazanego wprost interfejsu

```
router bgp 100
  neighbor 3.3.3.3 remote-as 300
  neighbor 3.3.3.3 description Sesja eBGP
  neighbor 3.3.3.3 ebgp-multihop 24
  neighbor 3.3.3.3 update-source loopback666
```

Konfiguracja protokołu BGP
Podstawy

- Sesje można tymczasowo zamknąć, nie usuwając jej z konfiguracji

```
router bgp 100
  neighbor 3.3.3.3 shutdown
```

Konfiguracja protokołu BGP
Podstawy

- Sesje można uwierzytelnić tagiem MD5

```
router bgp 100
  neighbor 3.3.3.3 password BARDZO.trudne.H4SL0
```

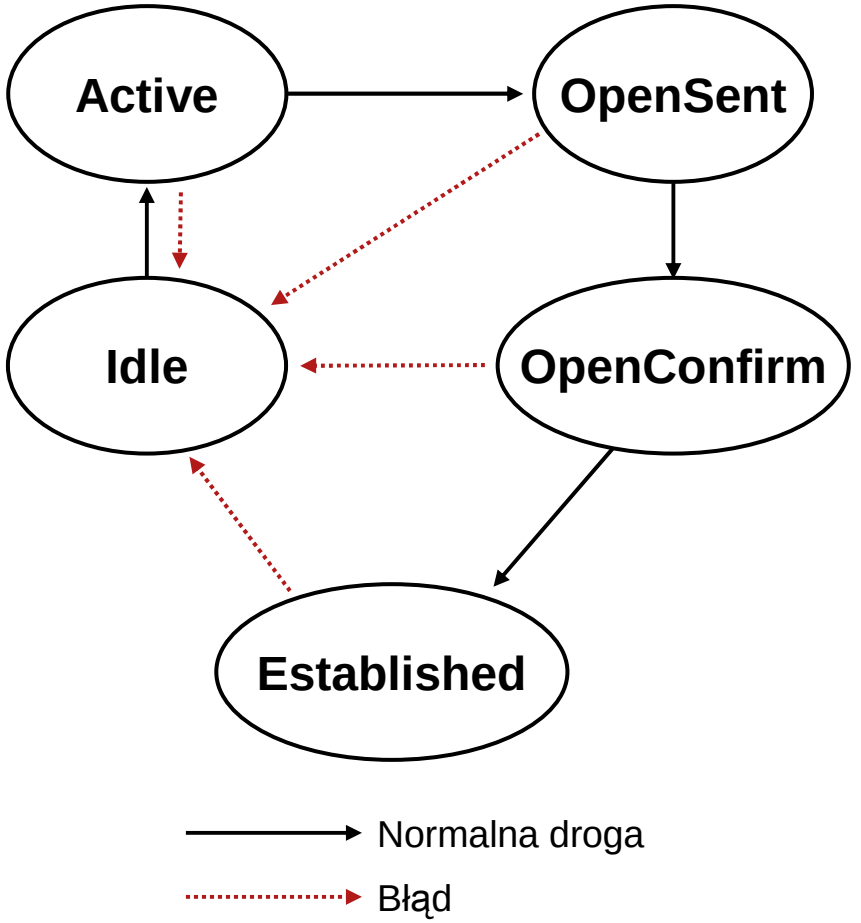
Konfiguracja protokołu BGP
Podstawy

- Sesje BGP używają stosu TCP routera do zestawienia sesji i wymiany informacji
- Domyślna wartość MSS to 576 bajtów – w większości wypadków to mało w porównaniu do MSS interfejsów

```
ip tcp path-mtu-discovery

! lub

router bgp 100
  neighbor 3.3.3.3 transport path-mtu-discovery
```

Konfiguracja protokołu BGP
Maszyna stanów BGP

- Jeśli wszystko przebiega w porządku, przejdź do następnego stanu
- Jeśli nie – przejdź do stanu „Idle”



Konfiguracja protokołu BGP
Rozgłaszanie sieci

- BGP rozgłosi sieć jeśli:

  posiada ją już w tablicy routingu (przy włączonej synchronizacji)

  zostanie ona wprost zdefiniowana poleceniem network

  zostanie rozdystrybuowana do BGP

```
router bgp 100
  no synchronization
  network 16.0.0.0
  network 192.168.0.0 mask 255.255.255.0
  redistribute ospf 100 match internal external 1
```

Konfiguracja protokołu BGP
Monitoring pracy

- Stan sesji BGP

```
router# sh ip bgp summary

BGP router identifier 192.168.0.1, local AS number 7581
BGP table version is 151059260, main routing table version 151059260
265964 network entries using 31915680 bytes of memory
546795 path entries using 28433340 bytes of memory
96324/48382 BGP path/bestpath attribute entries using 11944176 bytes of memory
85196 BGP AS-PATH entries using 2340426 bytes of memory
4214 BGP community entries using 242414 bytes of memory
7 BGP extended community entries using 236 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
32832 BGP filter-list cache entries using 393984 bytes of memory
Bitfield cache entries: current 3 (at peak 6) using 92 bytes of memory
BGP using 75270348 total bytes of memory
BGP activity 6666084/6400118 prefixes, 62119090/61572295 paths, scan interval 60

Neighbor        V    AS  MsgRcvd  MsgSent    TblVer  InQ OutQ Up/Down  State/PfxRcd
80.50.4.221     4  5617 22921212   278356 151059253    0    0 20w4d       260884
195.47.239.14   4 64999   594255   594954 151059260    0    0 2w2d            48
217.153.217.33  4  8246 50753653   610315 151059253    0    0 2w2d        263698
```

Konfiguracja protokołu BGP
Monitoring pracy

- Stan sesji do konkretnego sąsiada BGP

```
router# sh ip bgp neighbors 217.153.217.33
BGP neighbor is 217.153.217.33,  remote AS 8246, external link
 Description: Peering do GTSu
  BGP version 4, remote router ID 217.153.216.2
  BGP state = Established, up for 2w2d
  Last read 00:00:01, last write 00:00:12, hold time is 180, keepalive interval is 60s
  Neighbor capabilities:
    Route refresh: advertised and received(old & new)
    Address family IPv4 Unicast: advertised and received
  Message statistics:
    InQ depth is 0
    OutQ depth is 0
                         Sent       Rcvd
    Opens:                 29         29
    Notifications:         20          2
    Updates:               32   50143720
    Keepalives:        610239     610066
    Route Refresh:          0          0
    Total:             610320   50753819
  Default minimum time between advertisement runs is 30 seconds
```

Konfiguracja protokołu BGP
Monitoring pracy

- Stan sesji do konkretnego sąsiada BGP (c.d.)

```
For address family: IPv4 Unicast
 BGP table version 151059583, neighbor version 151059546/0
 Output queue size: 0
 Index 1, Offset 0, Mask 0x2
 1 update-group member
 Inbound soft reconfiguration allowed
 Outbound path policy configured
 Outgoing update AS path filter list is 1
                                Sent       Rcvd
 Prefix activity:               ----       ----
   Prefixes Current:               1     263696 (Consumes 13712192 bytes)
   Prefixes Total:                 1   12270511
   Implicit Withdraw:              1   11747023
   Explicit Withdraw:              0     259792
   Used as bestpath:             n/a     107900
   Used as multipath:            n/a          0

                                 Outbound    Inbound
 Local Policy Denied Prefixes:   --------    -------
   filter-list:                   5695440          0
   Suppressed duplicate:                0     263727
   AS_PATH loop:                      n/a          2
   Total:                         5695440     263729
 Number of NLRIs in the update sent: max 1, min 1
```

Konfiguracja protokołu BGP
Monitoring pracy

- Stan sesji do konkretnego sąsiada BGP (c.d.)

```
Connections established 29; dropped 28
  Last reset 2w2d, due to BGP Notification sent, hold time expired
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled, Mininum incoming TTL 0, Outgoing TTL 1
Local host: 217.153.217.34, Local port: 15674
Foreign host: 217.153.217.33, Foreign port: 179
Connection tableid (VRF): 0

Enqueued packets for retransmit: 0, input: 0  mis-ordered: 0 (0 bytes)

Event Timers (current time is 0x88C5A1314):
Timer          Starts    Wakeups            Next
[...]

iss: 3132171889  snduna: 3132632338  sndnxt: 3132632338     sndwnd:  15320
irs:  985225841  rcvnxt: 1302257327  rcvwnd:      15213  delrcvwnd:   1171

SRTT: 300 ms, RTTO: 303 ms, RTV: 3 ms, KRTT: 0 ms
minRTT: 0 ms, maxRTT: 616 ms, ACK hold: 200 ms
Status Flags: active open
Option Flags: nagle
IP Precedence value : 6

Datagrams (max data segment is 1460 bytes):
Rcvd: 290491 (out of order: 199), with data: 267312, total data bytes: 317031485
Sent: 263862 (retransmit: 878, fastretransmit: 0, partialack: 0, Second Congestion: 0
```

Konfiguracja protokołu BGP
Monitoring pracy

- Podgląd tablicy RIB protokołu BGP

```
router# show ip bgp
BGP table version is 151059911, local router ID is 192.168.0.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 1.0.0.0          192.0.2.1                0    200      0 64999 i
*> 2.0.0.0          192.0.2.1                0    200      0 64999 i
*> 3.0.0.0          217.153.217.33                         0 8246 6453 2914 80 i
*                   80.50.4.221                            0 5617 5511 2914 80 i
*> 4.0.0.0/9        217.153.217.33                         0 8246 1299 3356 i
*                   80.50.4.221                            0 5617 5511 3356 i
```

Konfiguracja protokołu BGP
Monitoring pracy

- Detale konkretnego prefiksu

```
router# sh ip bgp 4.21.103.1
BGP routing table entry for 4.21.103.0/24, version 132113781
Paths: (2 available, best #2, table Default-IP-Routing-Table)

  8246 3320 3549 46133, (received & used)
    217.153.217.33 from 217.153.217.33 (217.153.216.2)
      Origin IGP, localpref 100, valid, external
      Community: 3320:1276 3320:2010 3320:9020 3549:30840 8246:667

  5617 5511 3549 46133
    80.50.4.221 from 80.50.4.221 (213.25.2.126)
      Origin IGP, localpref 100, valid, external, best
      Community: 5511:666 5511:710 5511:5511
```

Konfiguracja protokołu BGP
Co zostało rozgłoszone?

- Rozgłoszone prefiksy do sąsiada

```
router# sh ip bgp neigh 1.1.1.1 advertised-routes
```

- Prefiksy otrzymane od sąsiada

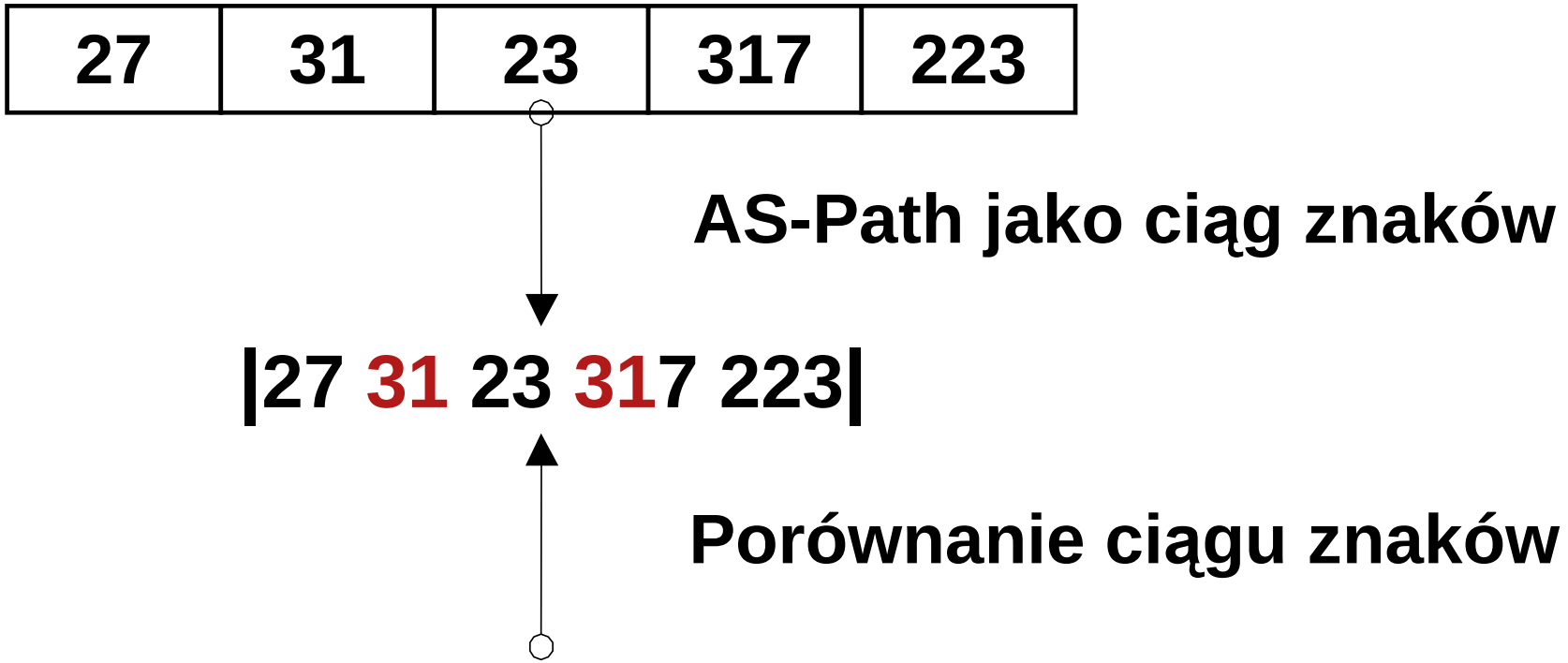
```
router# sh ip bgp neigh 1.1.1.1 routes
```

Konfiguracja protokołu BGP
Filtrowanie prefiksów

- Prefiks może zostać odfiltrowany lub zaakceptowany za pomocą:
  - as-path
  - prefix-listy
  - route-map

Konfiguracja protokołu BGP

Filtrowanie prefiksów – as-path



Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Ciąg znaków w wyrażeniu regularnym pasuje do dowolnego podciągu znaków w AS-Path:

Ile razy pasuje ciąg 31?

|213 317 2316 31|

Odpowiedź:

|213 317 2316 31|

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Wyrażenie *expr1*|*expr2* pasuje jeśli któreś z podwyrażeń znajduje się w AS-Path

Ile razy pasuje ciąg 21|31?

|213 317 2316 31|

Odpowiedź:

|213 317 2316 31|

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Zakres znaków pasuje do każdego wystąpienia znaku z zakresu: [1234] lub [1-4]

  Kropka (.) pasuje do dowolnego znaku

Ile razy pasuje ciąg [1-3].[34]?

|213 317 2316 31|

Odpowiedź:

|213 317 2316 31|

|213 317 2316 31|

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Inne symbole znane z wyrażeń regularnych:
  - ^ początek ciągu
  - $ koniec ciągu
  - _ dowolny znak separacji (początek, koniec, spacja)

Ile razy pasuje ciąg ^21, 31$, _31_?

|213 317 218 31 731|

Odpowiedź:

|213 317 218 31 731|

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Można użyć nawiasów do grupowania wyrażeń regularnych

Ile razy pasuje ciąg (213|218)_31?

|213 317 1218 316 31|

Odpowiedź:

|213 317 1218 316 31|

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path

- Dodatkowo
  - * Pasuje do zera lub większej ilości atomów
  - ? Pasuje do zera lub jednego atomu
  - \+ Pasuje do jednego lub większej ilości atomów

Jak sprawdzić czy AS-Path nie zawiera “23 45” i “23 78 45” w jednym teście?

Odpowiedź:

_23(_78)?_45_

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path – parę przykładów

- _100_ - przez AS100
- ^100$ - bezpośrednio podłączony do AS100
- _100$ - rozgłoszony przez AS100
- ^100_. - sieci za AS100
- ^ [0-9]+$ - AS-Path z jednego AS
- ^$ - sieci z lokalnego AS
- .* - dowolna AS-Path

Konfiguracja protokołu BGP
Filtrowanie prefiksów – as-path - konfiguracja

- Definicja AS-Path i przypisanie jej do sąsiada:

```
ip as-path access-list 1 {permit|deny} regexp

router bgp 100
   neighbor 1.1.1.1 filter-list 1 {in|out}
```

- Przejrzenie tablicy RIB BGP z użyciem konkretnego wyrażenia regularnego:

```
show ip bgp regexp regexp

show ip bgp filter-list numer-as-path-listy
```

Konfiguracja protokołu BGP
Filtrowanie prefiksów – prefix-listy

- Prefix-listy mają nazwy i numery sekwencyjne
- Wpis bez ‘le’ i ‘ge’ pasuje tylko dokładnie do wskazanej sieci
- Wpis z ‘le’ lub ‘ge’ pasuje jeśli prefiks mieści się w przestrzeni adresowej (‘ge’) lub jest ‘krótszy lub równy’ (‘le’)

```
ip prefix-list list-name [seq seq] {permit|deny}
network/len [ge value] [le value]
```

Konfiguracja protokołu BGP
Filtrowanie prefiksów – prefix-listy - przykład

Która z poniższych tras przejdzie przez filtr?

`ip prefix-list MyList permit 192.168.0.0/16 le 20?`

✓ `192.168.0.0/16` ✓ `192.168.17.0/20` X `192.168.2.0/24`

`ip prefix-list MyList permit 192.168.0.0/16 ge 18?`

X `192.168.0.0/16` ✓ `192.168.17.0/20` ✓ `192.168.2.0/24`

Konfiguracja protokołu BGP
Filtrowanie prefiksów – prefix-listy - konfiguracja

- Prefix-lista może dotyczyć prefiksów otrzymywanych od sąsiada (in) lub wysyłanych do sąsiada (out)

```
neighbor {ip-address|peer-group-name}
  prefix-list prefix-listname {in|out}
```

Podstawowa konfiguracja protokołu BGP

Warsztaty

Mapowanie podów

- 10.62.0.3:6**XX**YY
  - –**XX** – numer grupy (podu)
  - –**YY** – numer routera w grupie
- Przykład:
  - –Grupa 6, router 2: 60602
  - –Grupa 9, router 6: 60906

- R1 – 10.62.0.3:6XX01
- R2 – 10.62.0.3:6XX02
- R3 – 10.62.0.3:6XX03
- R4 – 10.62.0.3:6XX04
- R5 – 10.62.0.3:6XX05
- R6 – 10.62.0.3:6XX06
- R7 – 10.62.0.3:6XX07
- R8 – 10.62.0.3:6XX08
- R9 – 10.62.0.3:6XX09

Skalowanie BGP: konfederacje i RR

Prezentacja

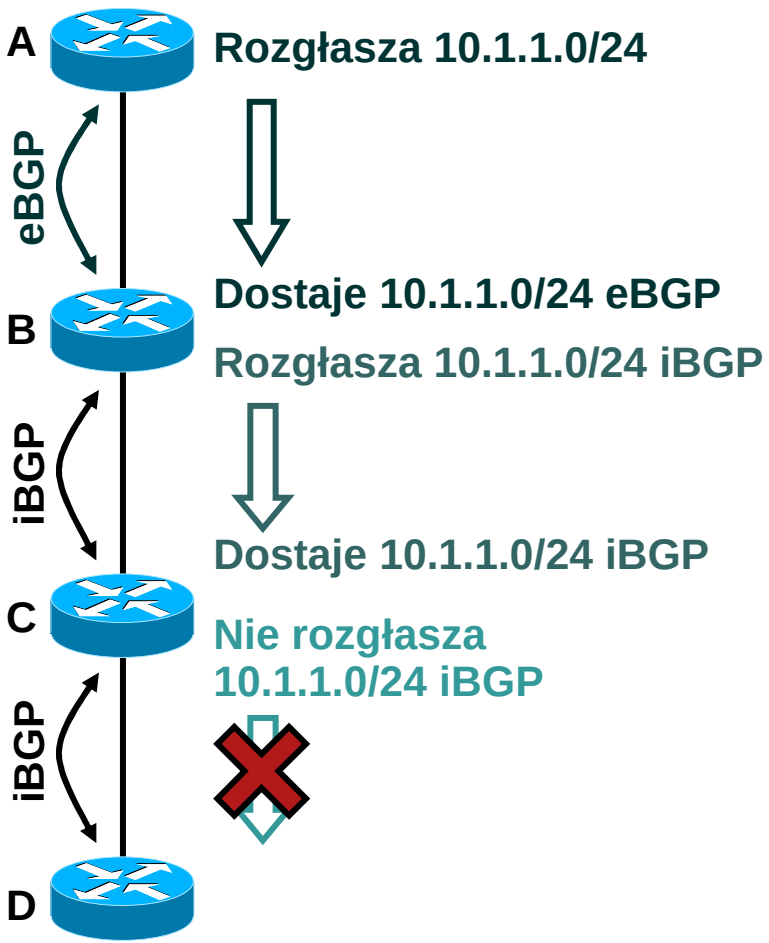
Po co pełna siatka połączeń iBGP?

- If a particular AS has multiple BGP speakers and is providing transit service for other AS's, then care must be taken to ensure a consistent view of routing within the AS. A consistent view of the interior routes of the AS is provided by the IGP used within the AS. For the purpose of this document, it is assumed that a consistent view of the routes exterior to the AS is provided by having all BGP speakers within the AS maintain IBGP sessions with each other.

RFC4271, http://tools.ietf.org/html/rfc4271

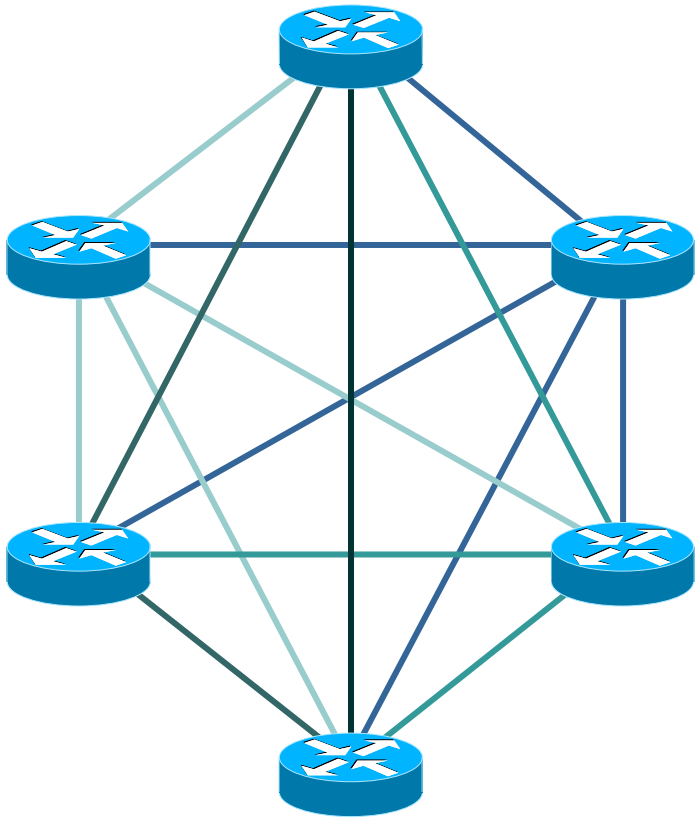
Pełna siatka iBGP

- Router ucząc się trasy od sąsiada iBGP nie rozgłosi tej samej trasy do innego sąsiada iBGP
- **Dlaczego?**
- Ponieważ BGP opiera się na AS Path by uniknąć pętli
- Sąsiedzi iBGP znajdują się w tym samym ASie, więc nie dodają niczego do ścieżki AS
- Nie ma żadnej możliwości oceny czy trasa została rozgłoszona przez wiele routerów iBGP i czy nie jest pętlą



Pełna siatka iBGP

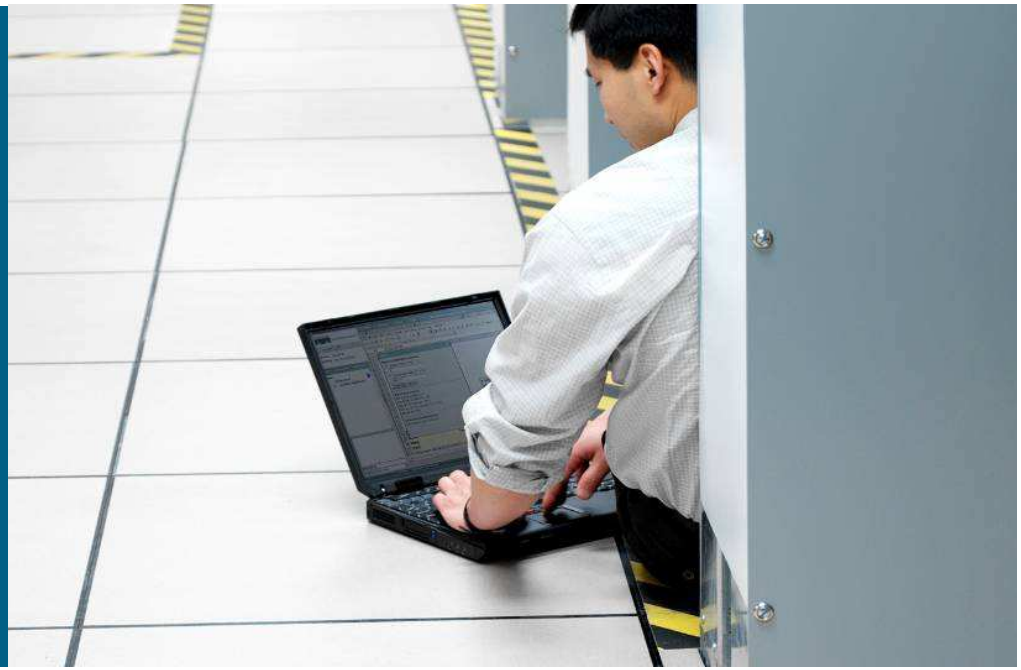
- Jak skalowalne jest stworzenie pełnej siatki połączeń iBGP?
  - 2 speakers == 1 peer
  - 3 speakers == 3 peers
  - 4 speakers == 6 peers
  - 5 speakers == 10 peers
  - 6 speakers == 15 peers
- n(n-1)/2
- ...no, nie bardzo, ale jak ktoś lubi...



Podstawy odbijania tras...

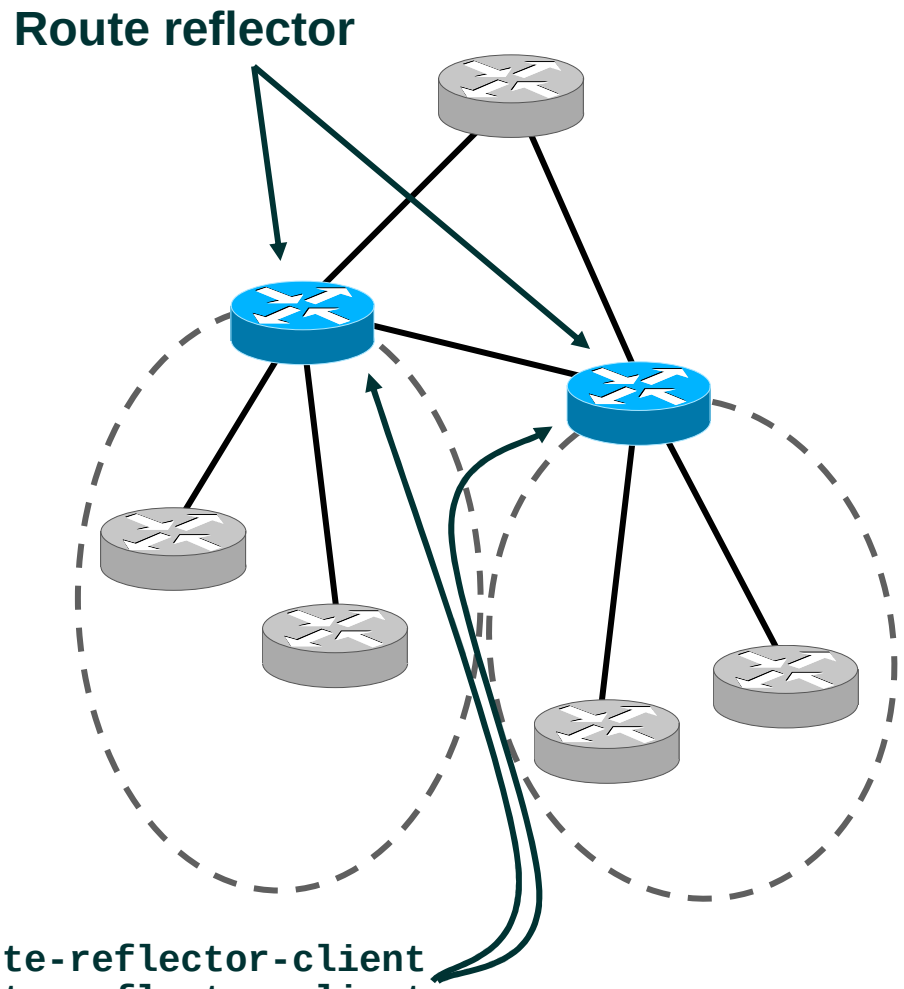
- Jeśli pełna siatka się nie skaluje, jakie mamy opcje?
- Route Reflectory
- Konfederacje

Route Reflector



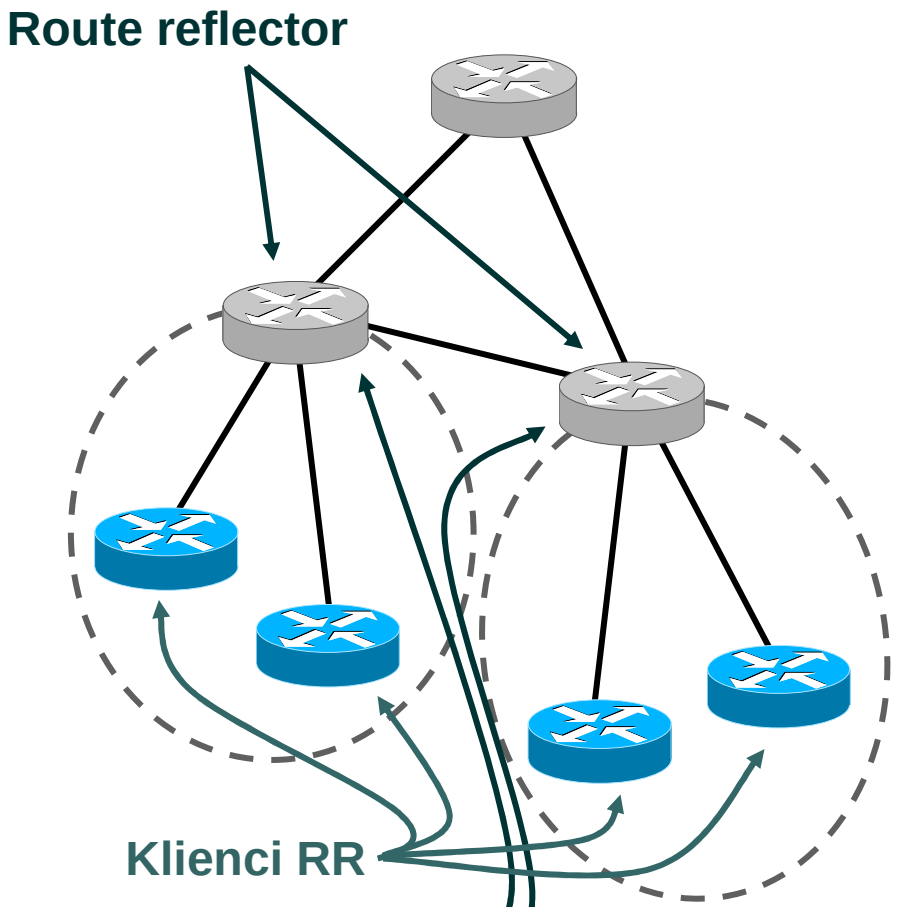
Route Reflector

- Router iBGP odbijający trasy nauczone od jednego routera iBGP do innych routerów iBGP
- Konfiguracja RR odbywa się przez wskazanie relacji klienckiej z określonymi sąsiadami



Route Reflector

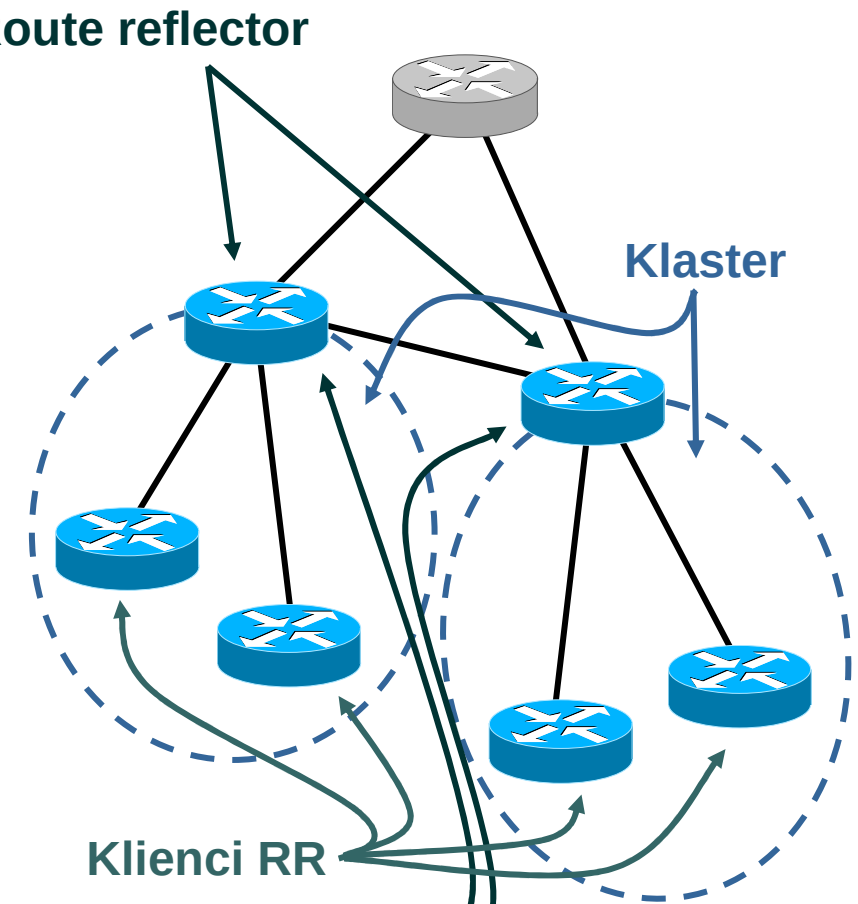
- Klient RR jest normalnym routerem iBGP i normalnym sąsiadem
- Nie ma żadnej specjalnej konfiguracji




```
neighbor <X> route-reflector-client
neighbor <X> route-reflector-client
```

Route Reflector

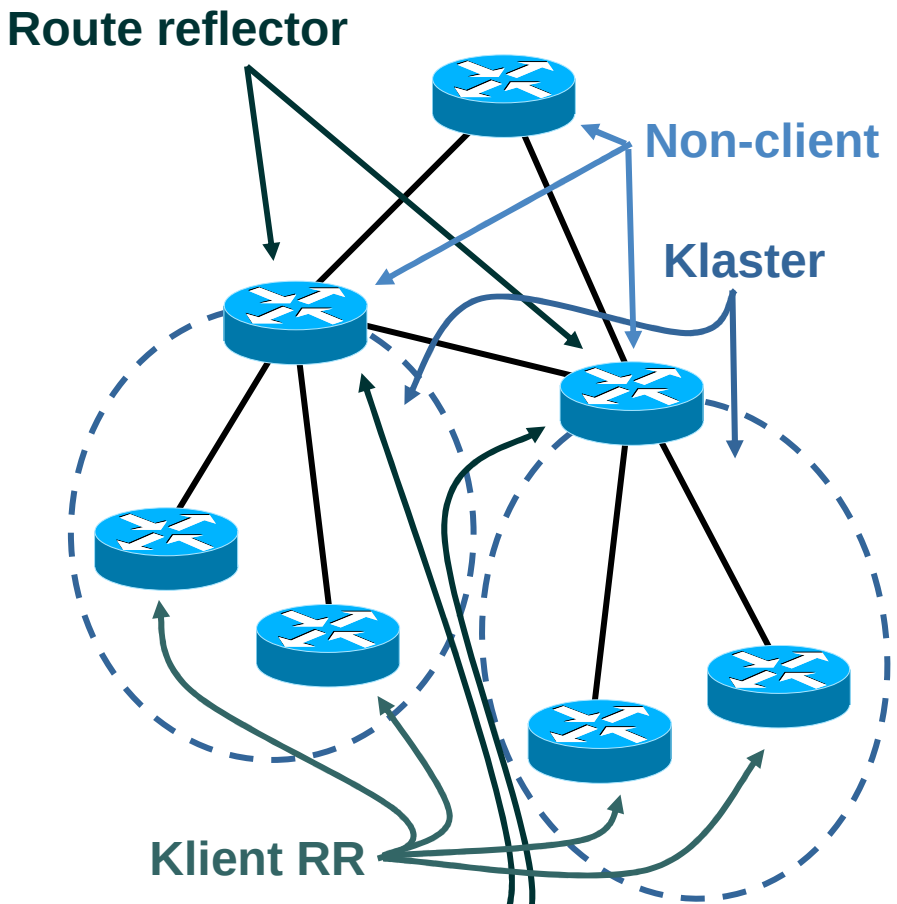
- **Klaster** to RR i jego klienci
- Klastry mogą się nakładać w sieci




```
neighbor <X> route-reflector-client
neighbor <X> route-reflector-client
```

Route Reflector

- non-client to taki sąsiad RR iBGP który nie jest klientem RR
- Każdy z RR jest również non-clientem dla innych RR
- RR muszą być połączone w pełną siatkę z non-clientami



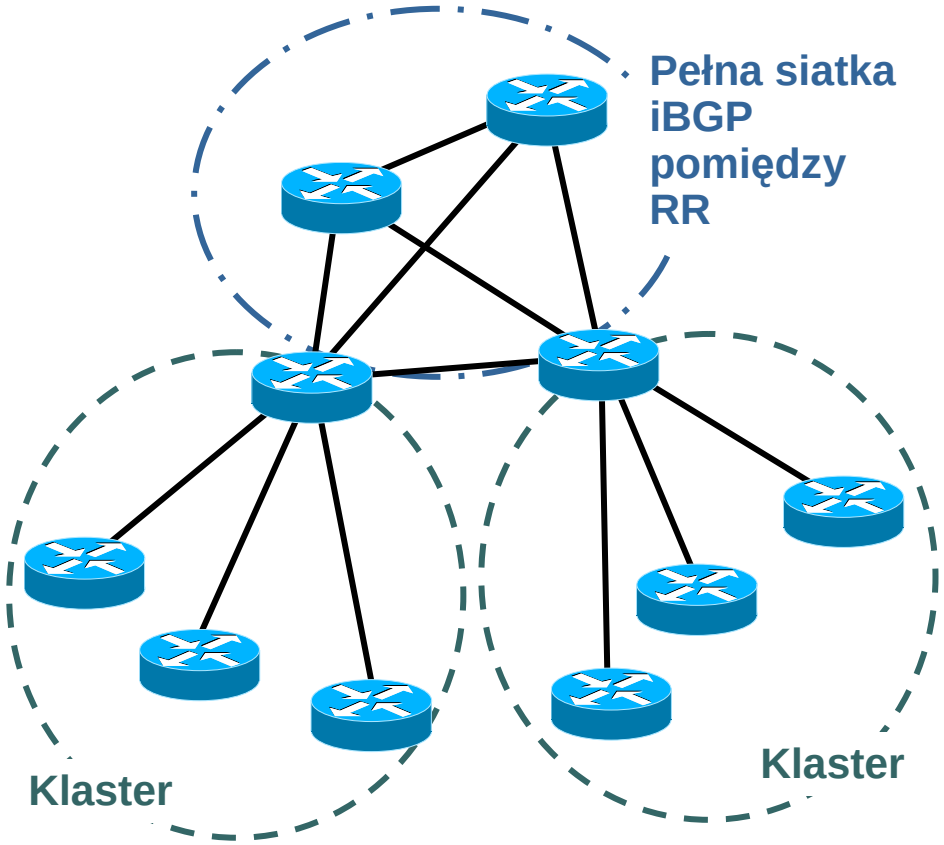

```
neighbor <X> route-reflector-client
neighbor <X> route-reflector-client
```

Hierarchia RR

- Wszystkie RR muszą być połączone w pełną siatką

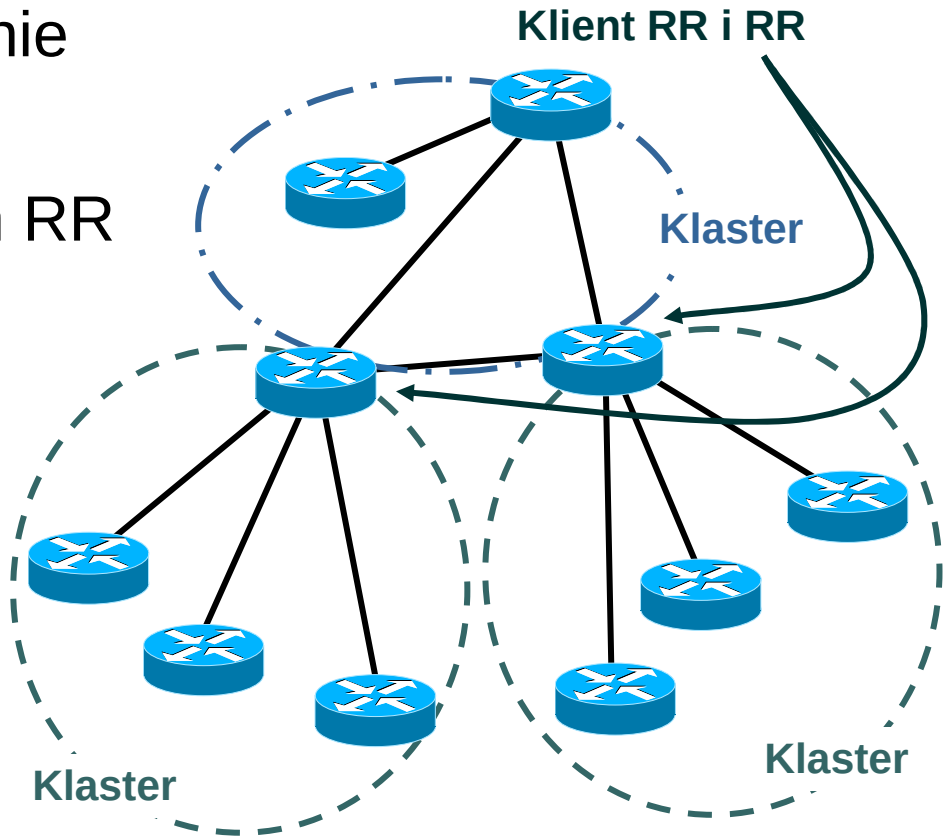
  RR nadal obowiązują podstawowe zasady rozgłaszania przez iBGP

- Ta siatka może rozrosnąć się do dużych rozmiarów oznaczając dokładnie takie same problemy ze skalowalnością jak na początku



Hierarchia RR

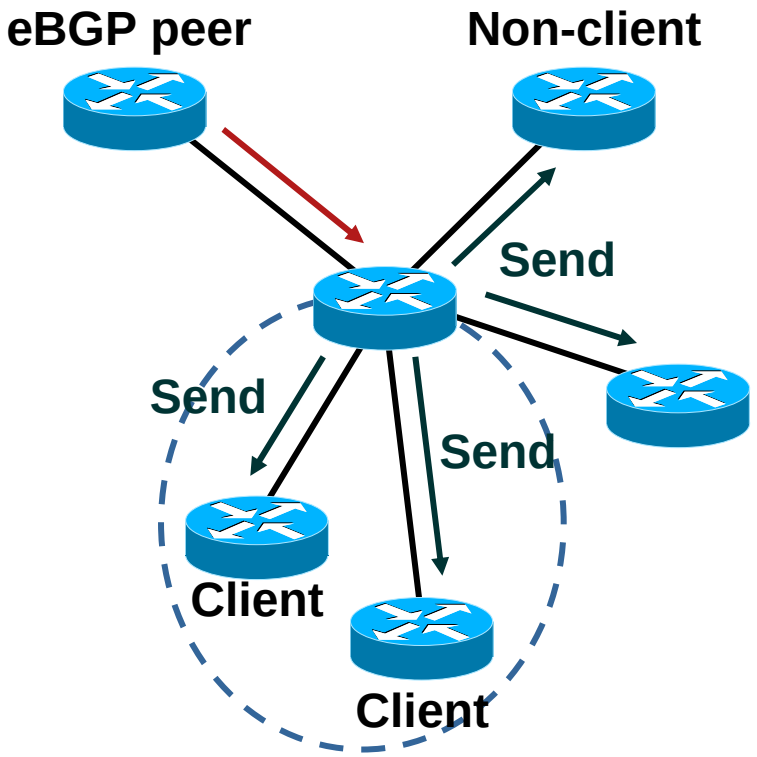
- RR mogą tworzyć hierarchie
- Jeden router może być zarówno RR jak i klientem RR



Działanie RR

RR otrzymując trasę od sąsiada eBGP:

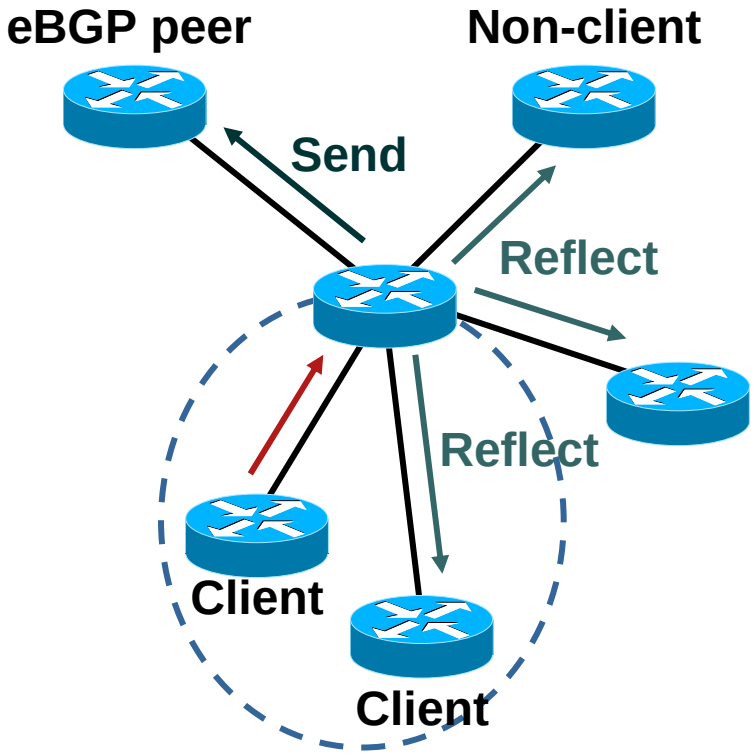
- Wysyła trasę do wszystkich klientów i nie-klientów



Działanie RR

RR otrzymując trasę od klienta:

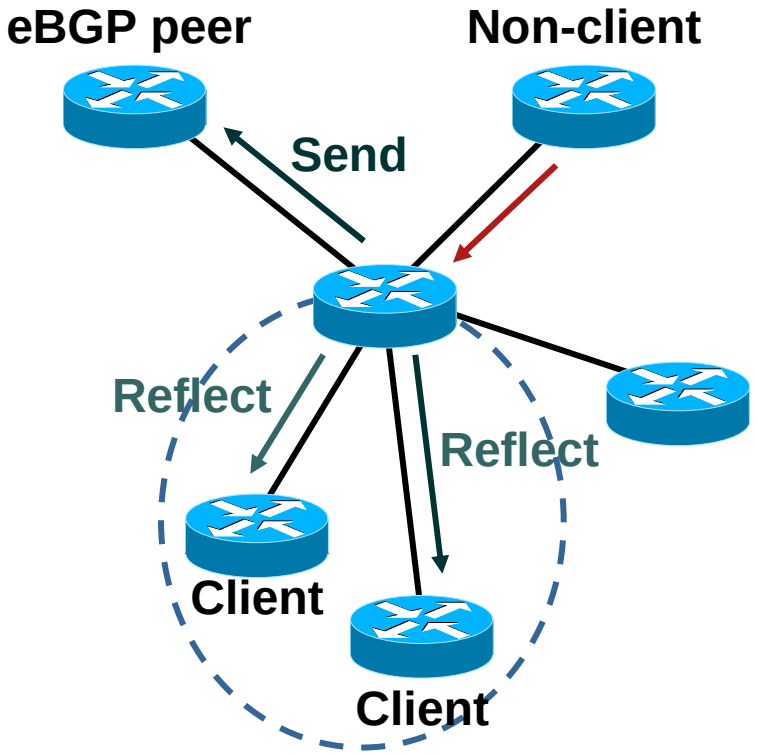
- Odbija ją do wszystkich klientów
- Odbija ją do wszystkich nie-klientów
- Wysyła trasę do wszystkich sąsiadów eBGP



Działanie RR

RR otrzymuje trasę od nie-klienta:

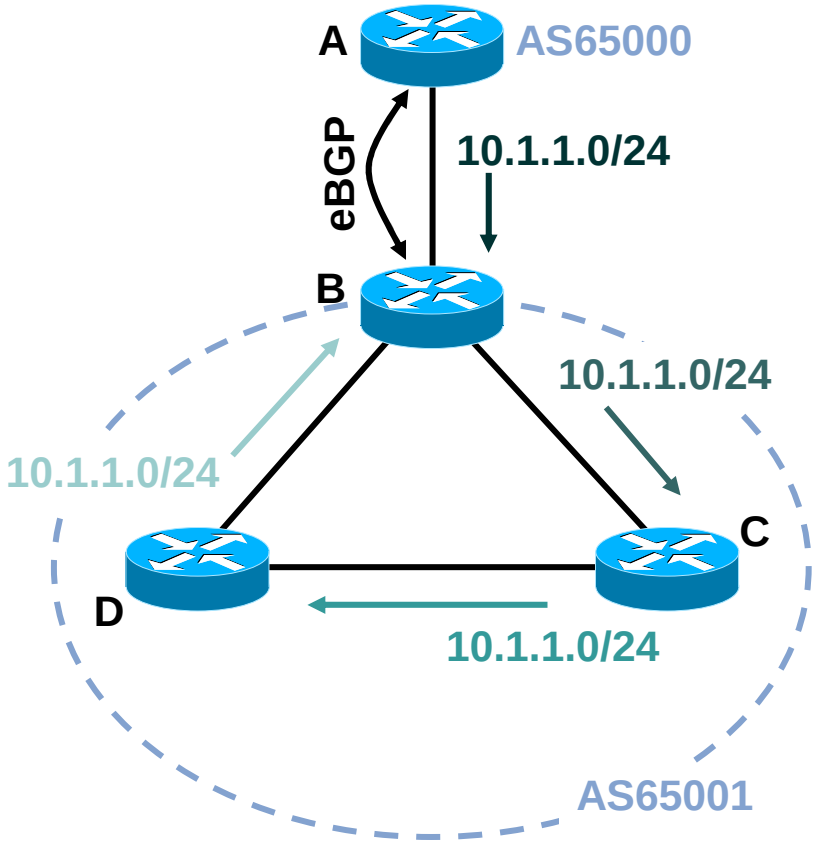
- Odbija ją do wszystkich klientów
- Wysyła trasę do wszystkich sąsiadów eBGP



Działanie RR

Wiemy że iBGP nie gwarantuje trasy bez pętli przez AS

- B otrzymuje 10.1.1.0/24 z AS Path równą {65000}
- C otrzymuje 10.1.1.0/24 z AS Path równą {65000}
- D otrzymuje 10.1.1.0/24 z AS Path równą {65000}
- B otrzymuje tą samą trasę z tymi samymi atrybutami, tworząc pętle!



Działanie RR

- Potrzebujemy jakiegoś mechanizmu żeby zapobiec możliwości powstania takiej sytuacji w obrębie AS!
- RFC2796 definiuje dwa atrybuty BGP mające zapobiec możliwości powstania pętli
- Originator ID

  Ustawiony na ID routera który rozgłasza trasę do AS
- Cluster List

  Każdy RR przez który rozgłaszana jest trasa dodaje swój ID klastra. **ID klastra = router ID (domyślnie)**

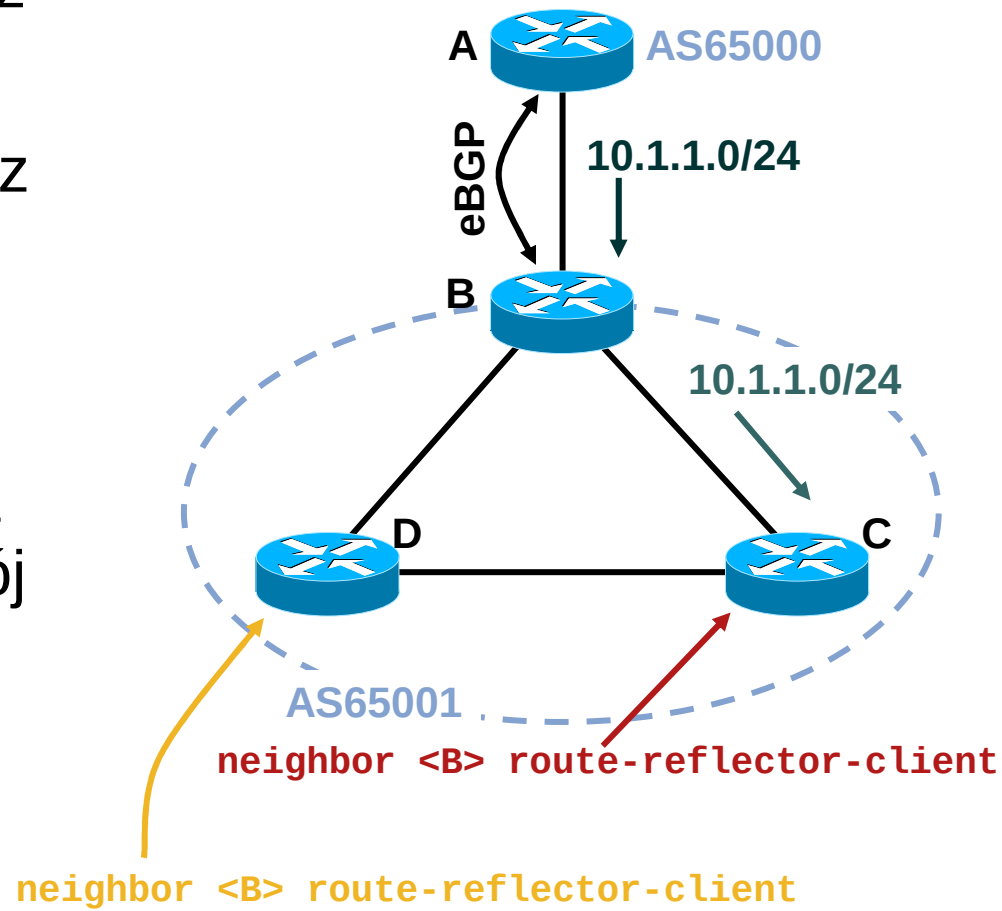
“bgp cluster-id A.B.C.D” nie jest zatem wymagane

Działanie RR

- Podczas odbijania trasy, RR zawsze:
  - Tworzy atrybut Cluster List jeśli jeszcze taki nie istnieje
  - Dodaje swój router ID (lub skonfigurowany ID klastra) do atrybutu
  - Dodaje router ID sąsiada który rozgłosił tą trasę jako Originator ID
- Podczas rozgłaszania trasy, RR zawsze przechodzi normalną procedurę BGP

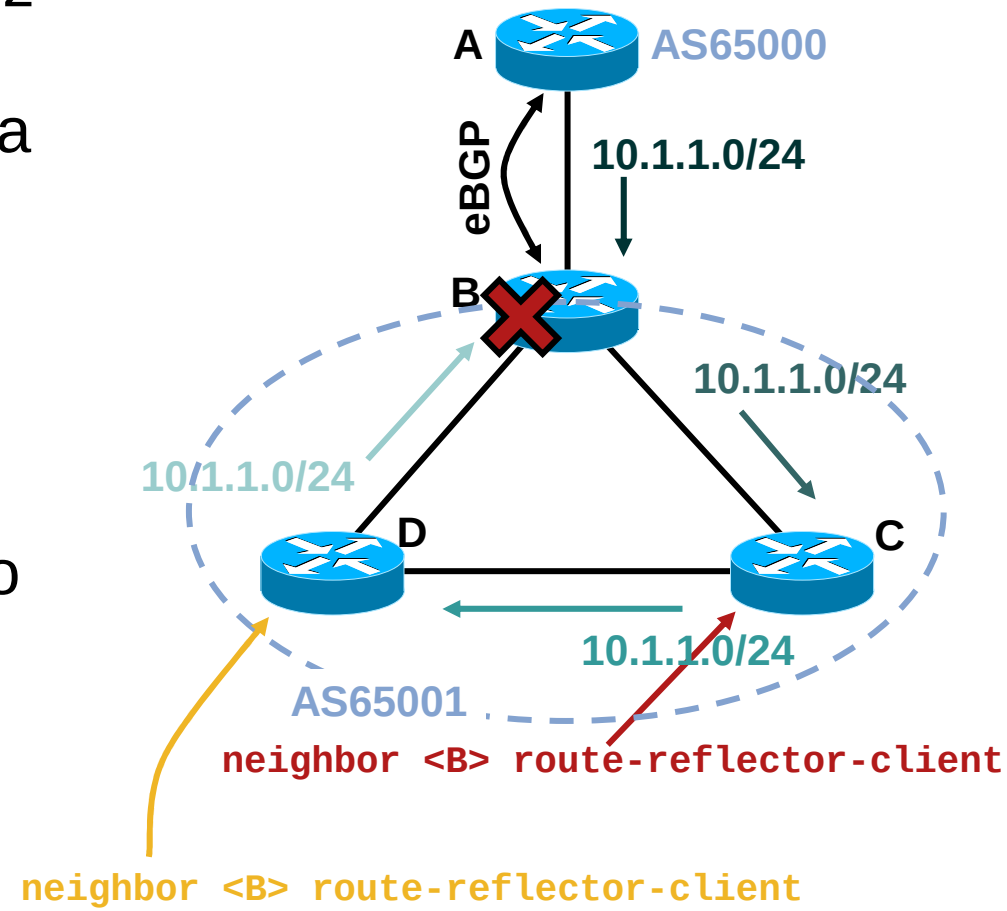
Działanie RR

- B otrzymuje 10.1.1.0/24 z AS Path równą {65000}
- C otrzymuje 10.1.1.0/24 z AS Path równą {65000}, ale dodaje RouterID B jako Originator ID
- C tworzy również atrybut Cluster List, i dodaje swój RouterID do listy

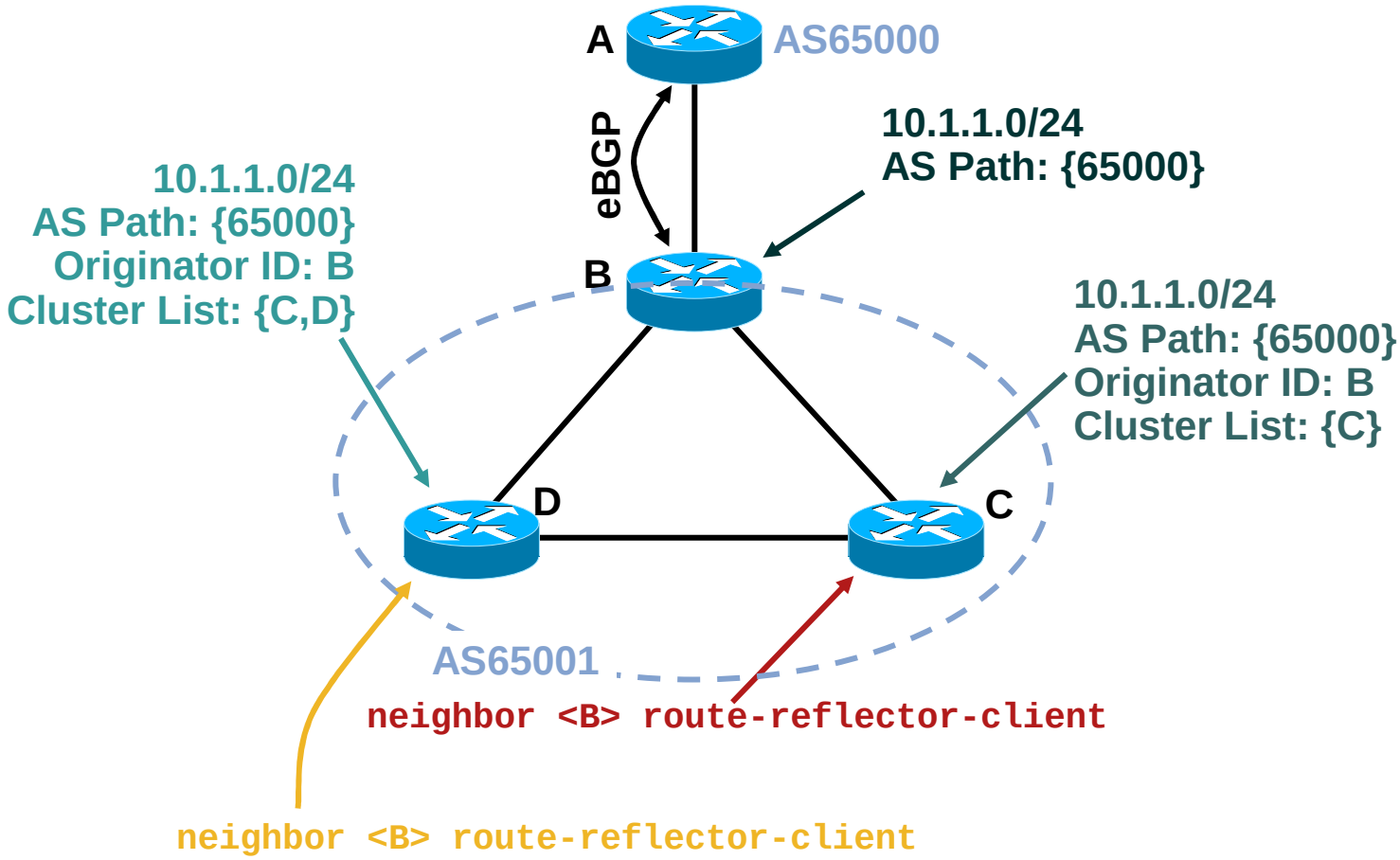


Działanie RR

- D otrzymuje 10.1.1.0/24 z AS Path równą {65000} oraz Originator ID routera B
- Router D dodaje swój RouterID do atrybutu Cluster list
- Router D wysyła trasę do routera B. Router B wykrywa pętle ponieważ Originator ID pasuje do jego własnego identyfikatora



Działanie RR



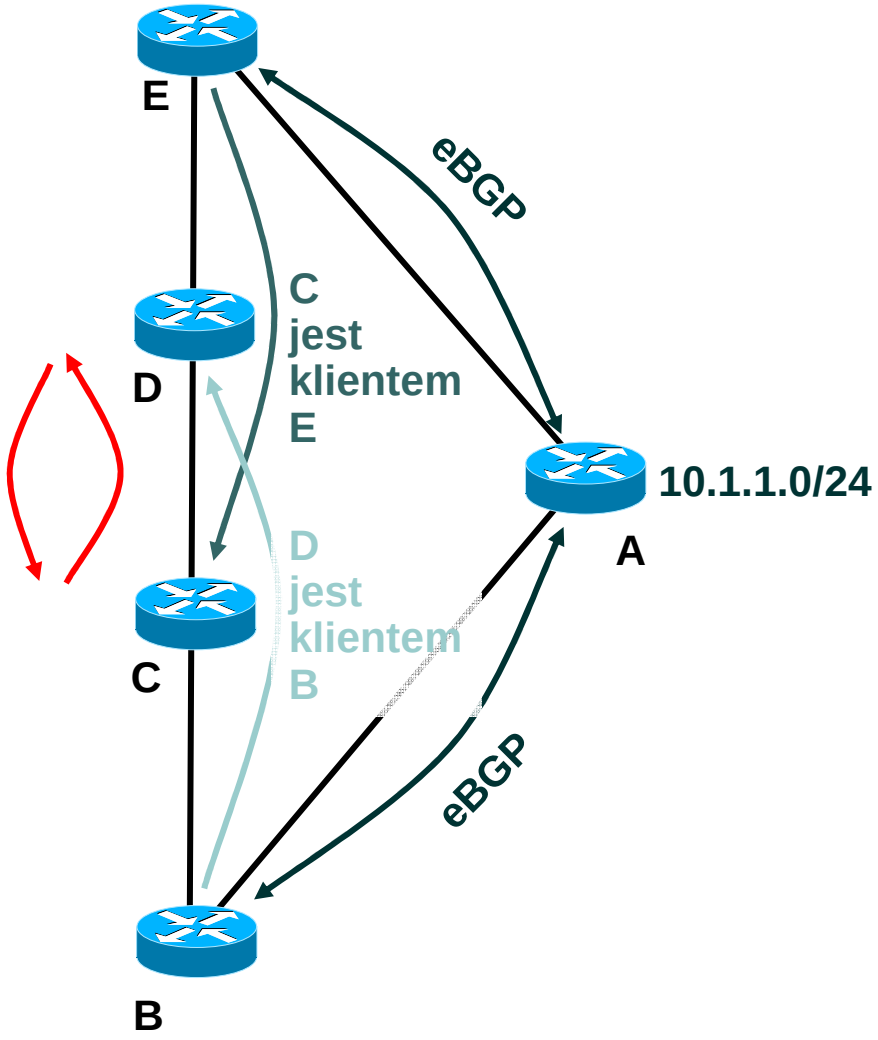
Działanie RR

Router E:
BGP Next-Hop: Router A
Local Next-Hop: Router A
Set: Next-Hop-Self

Router D:
BGP Next-Hop: Router B
Local Next-Hop: Router C

Router C:
BGP Next-Hop: Router E
Local Next-Hop: Router D

Router B:
BGP Next-Hop: Router A
Local Next-Hop: Router A
Set: Next-Hop-Self



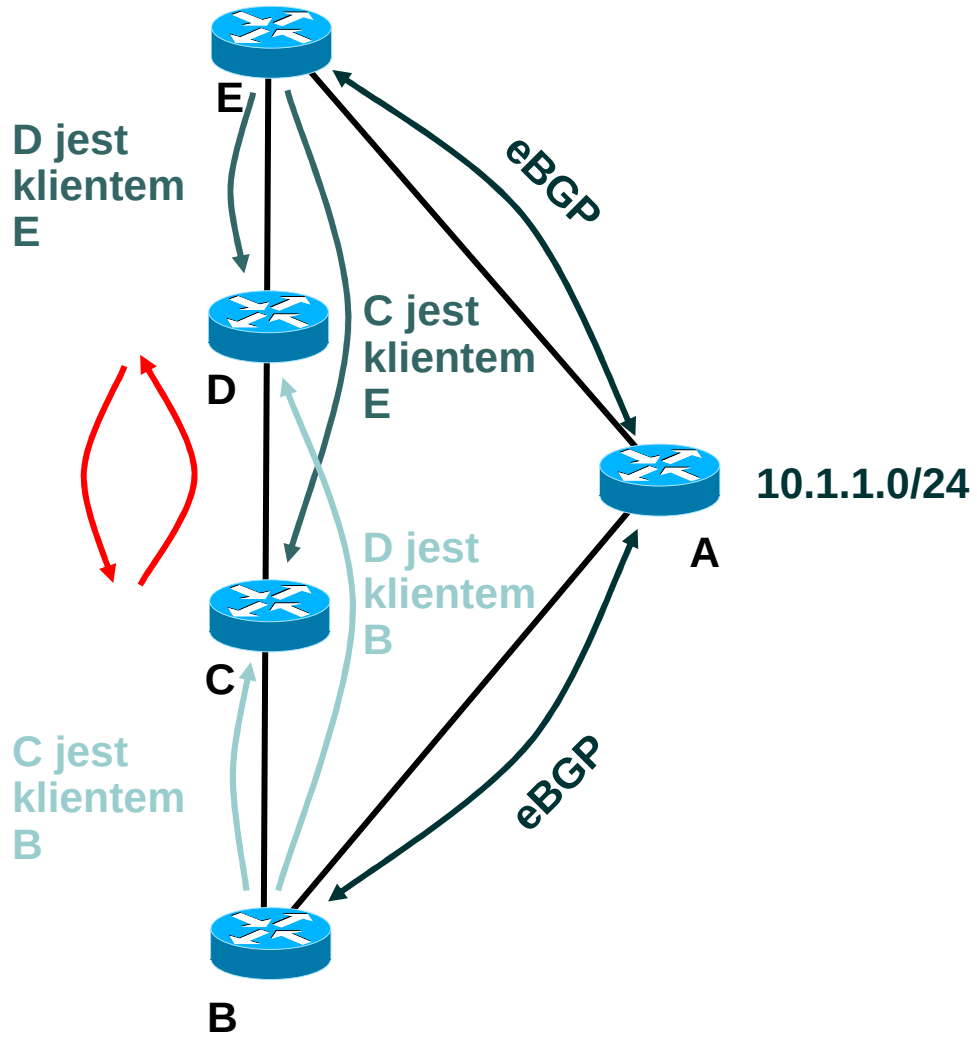
Działanie RR

Router E:
BGP Next-Hop: Router A
Local Next-Hop: Router A
Set: Next-Hop-Self

Router D:
BGP Next-Hop: Router B
Local Next-Hop: Router E

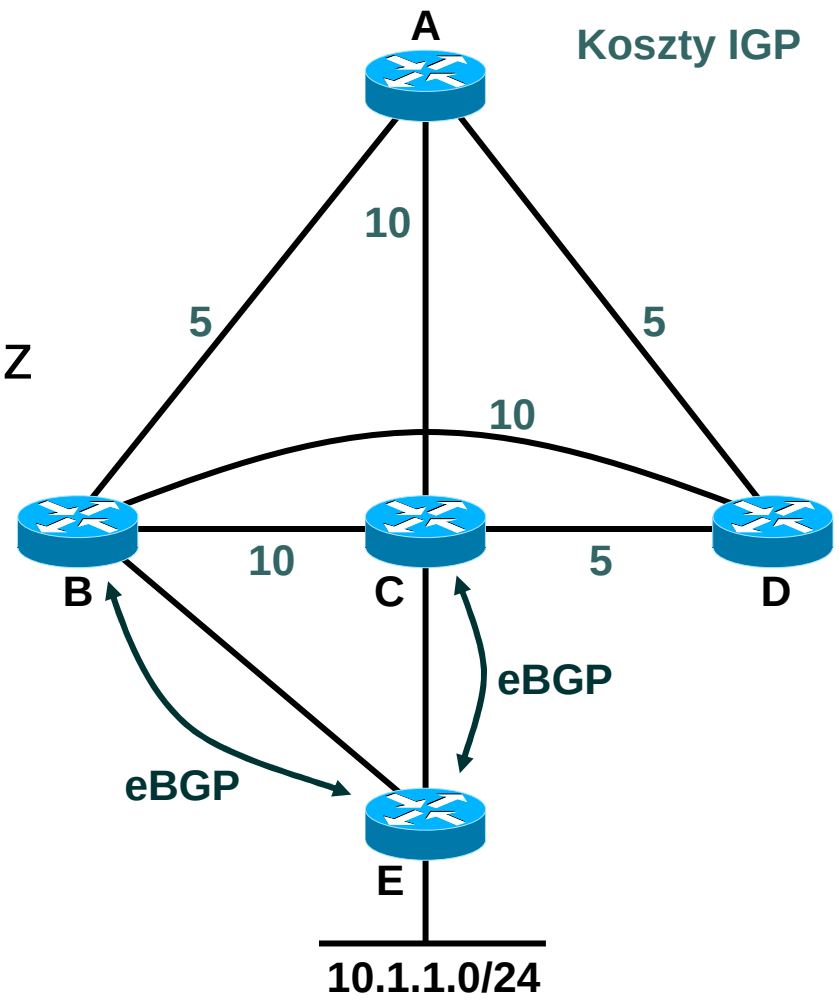
Router C:
BGP Next-Hop: Router B
Local Next-Hop: Router B

Router B:
BGP Next-Hop: Router A
Local Next-Hop: Router A
Set: Next-Hop-Self



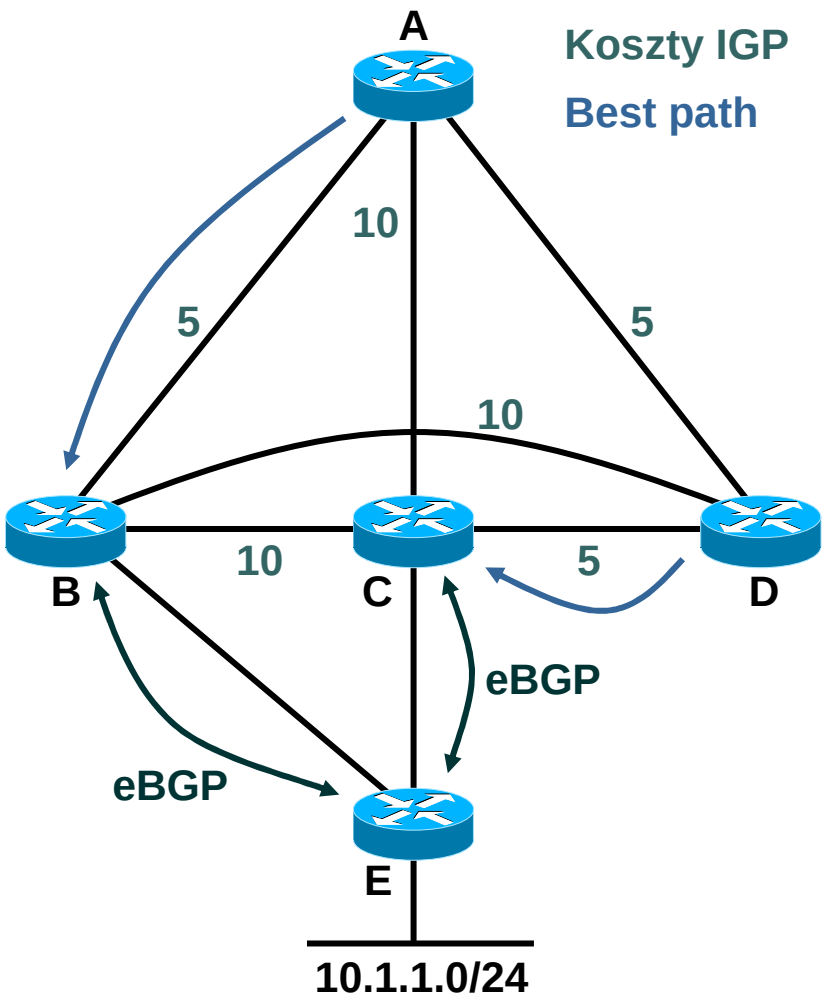
Działanie RR

- RR mogą spowodować suboptymalny routing w porównaniu do pełnej siatki iBGP
- E rozgłasza 10.1.1.0/24 przez eBGP do B i C
- Atrybuty local preference, MED, i długość AS Path są takie same dla 10.1.1.0/24 zarówno na B jak i na C



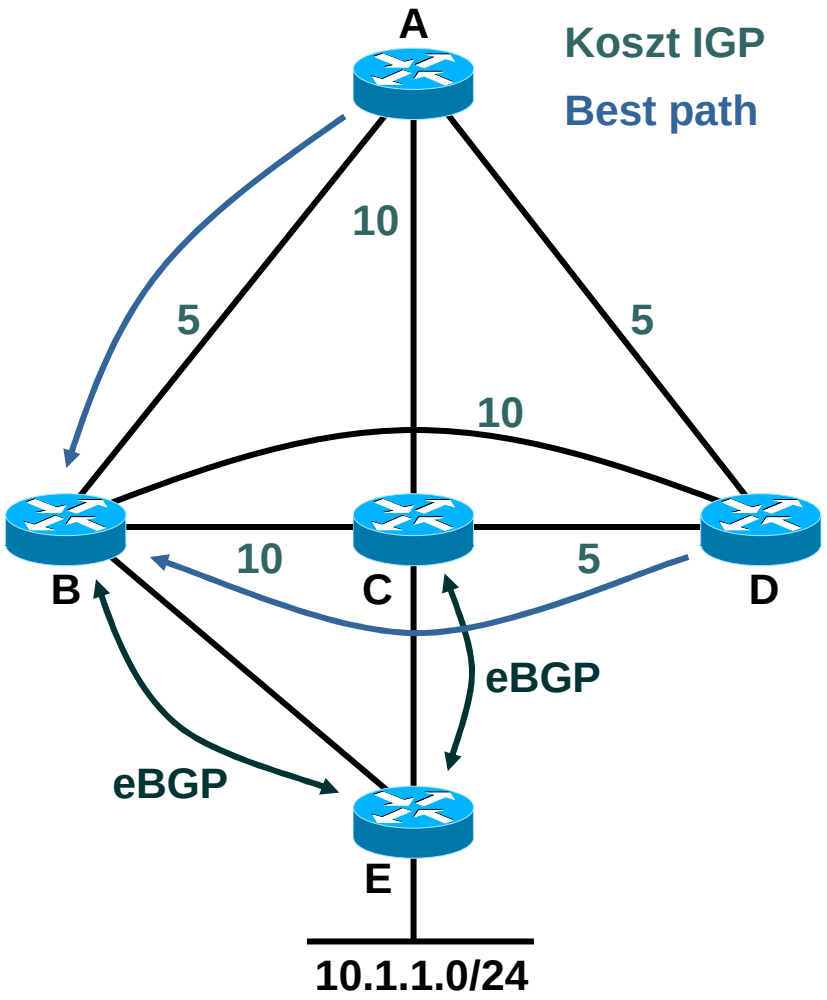
Działanie RR

- Załóżmy, że A, B, C i D są skonfigurowane do pełnej siatki połączeń iBGP
- A wybiera B jako punkt wyjścia z uwagi na koszt IGP
- D wybiera C z tego samego powodu



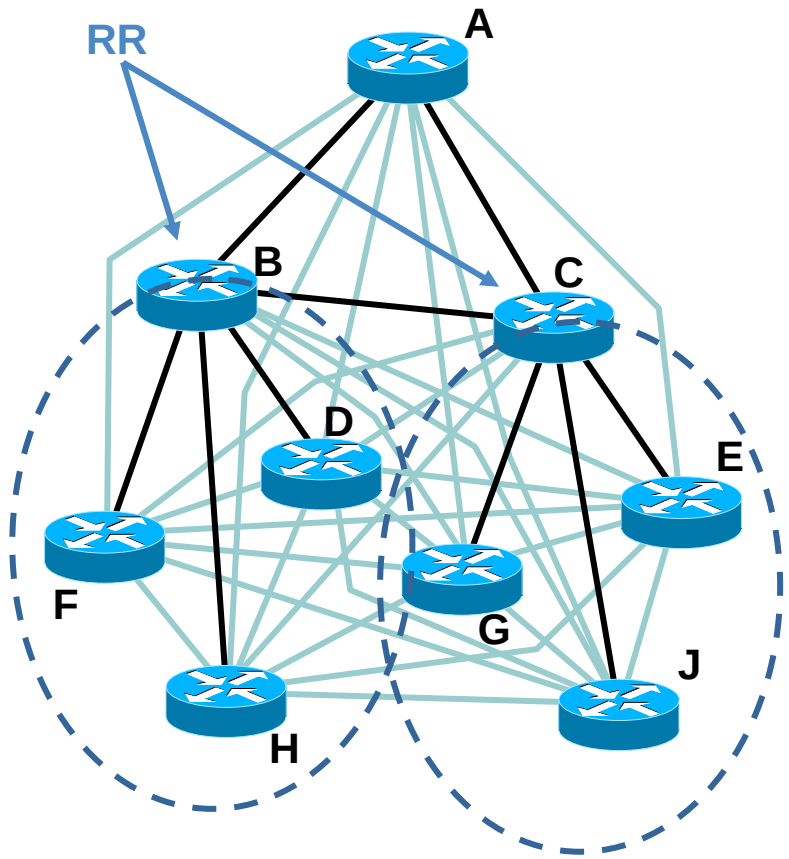
Działanie RR

- Teraz załóżmy, że B, C i D są klientami RR routera A
- A wybiera B jako najlepsze wyjście z uwagi na koszt IGP
- A odbija trasę do C, ale C wybiera lokalną trasę przez eBGP przez B
- A odbija ten wybór do D, a D wybiera B mimo że trasa przez C jest lepsza



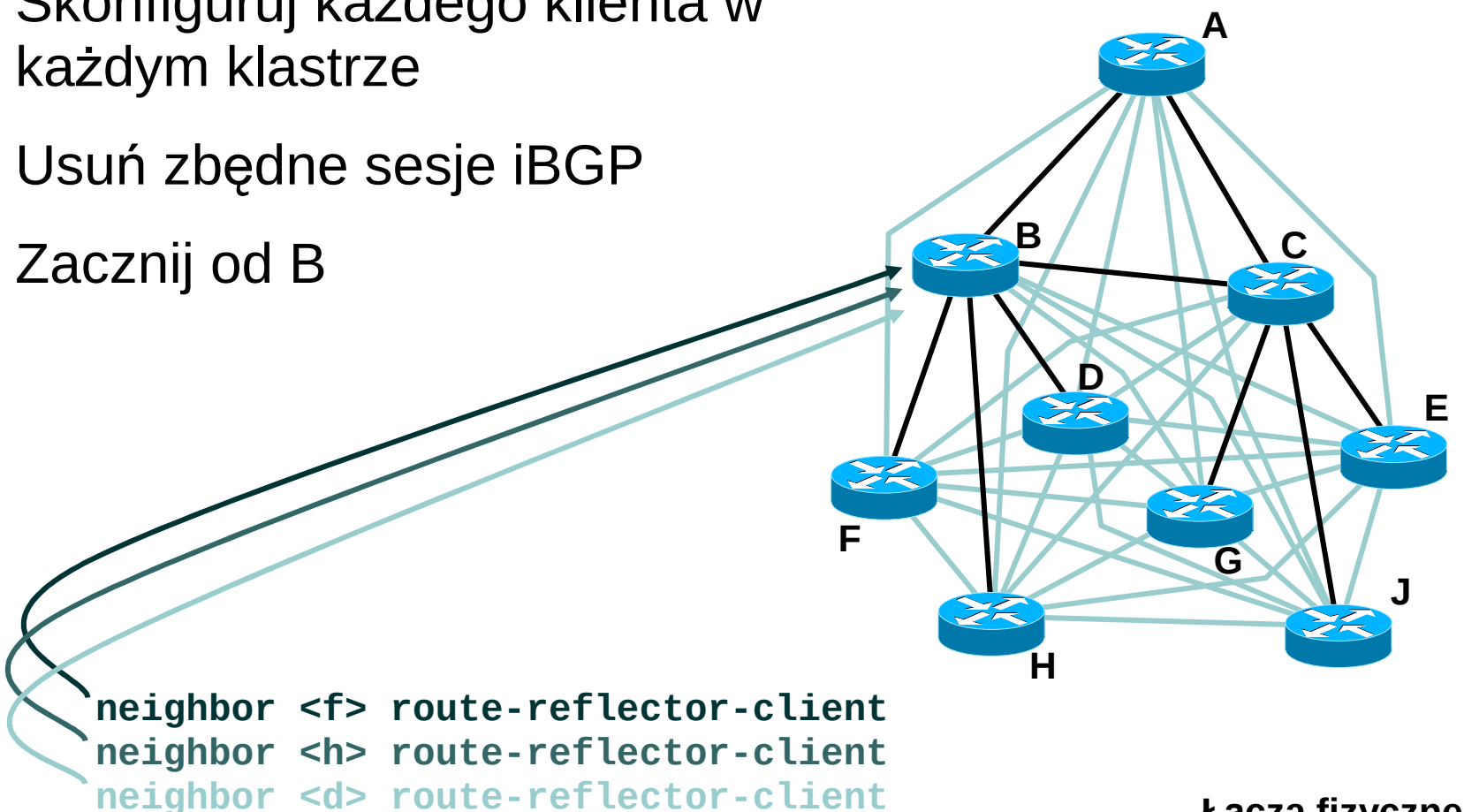
Przykład wdrożenia RR do sieci

- 9 routerów i 36 sesji iBGP na początku...
- Wybierz klastry posługując się topologią fizyczną
- Następnie wybierz RR posługując się również topologią fizyczną



Przykład wdrożenia RR do sieci

- Skonfiguruj każdego klienta w każdym klastrze
- Usuń zbędne sesje iBGP
- Zacznij od B

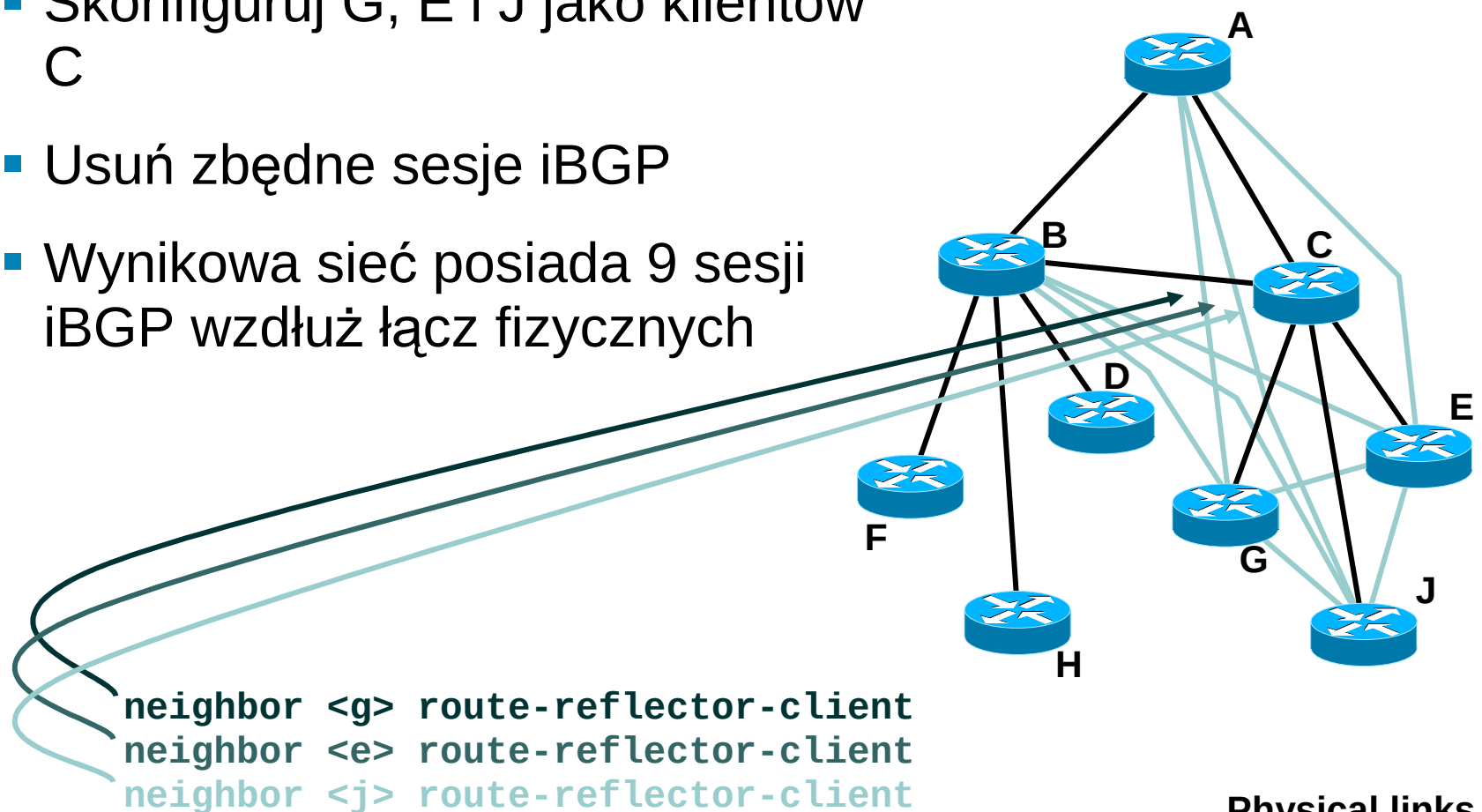


Łącza fizyczne

Sesje iBGP

Przykład wdrożenia RR do sieci

- Skonfiguruj G, E i J jako klientów C
- Usuń zbędne sesje iBGP
- Wynikowa sieć posiada 9 sesji iBGP wzdłuż łącz fizycznych



Redundancja RR

- Klient może być połączony z więcej niż jednym RR w różnych klastrach

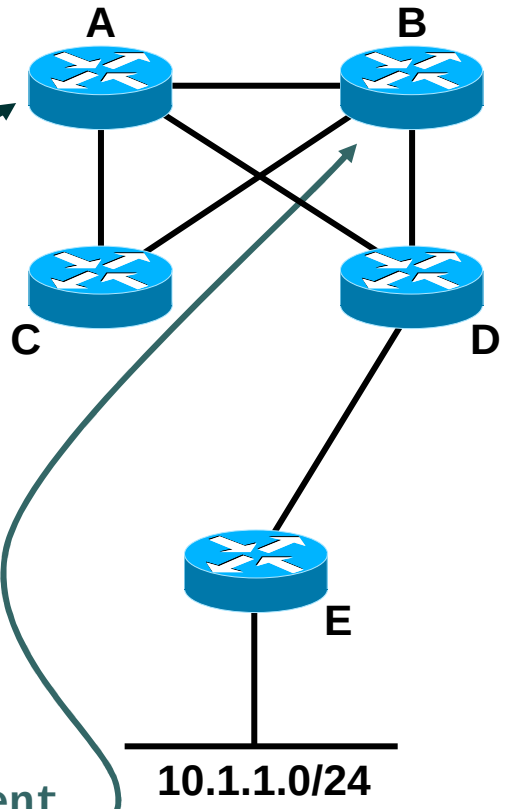
  Klient z jedną sesją do jednego RR = potencjalny pojedynczy punkt awarii

  Klienci powinni być podłączeni do przynajmniej dwóch RR

- Do ilu? Czy jest jakiś magiczny wzór?

Redundancja RR

- Załóżmy że A i B mają skonfigurowanych tych samych klientów
- **Są redundantne**
- Powinny mieć to samo ClusterID czy różne ClusterID?

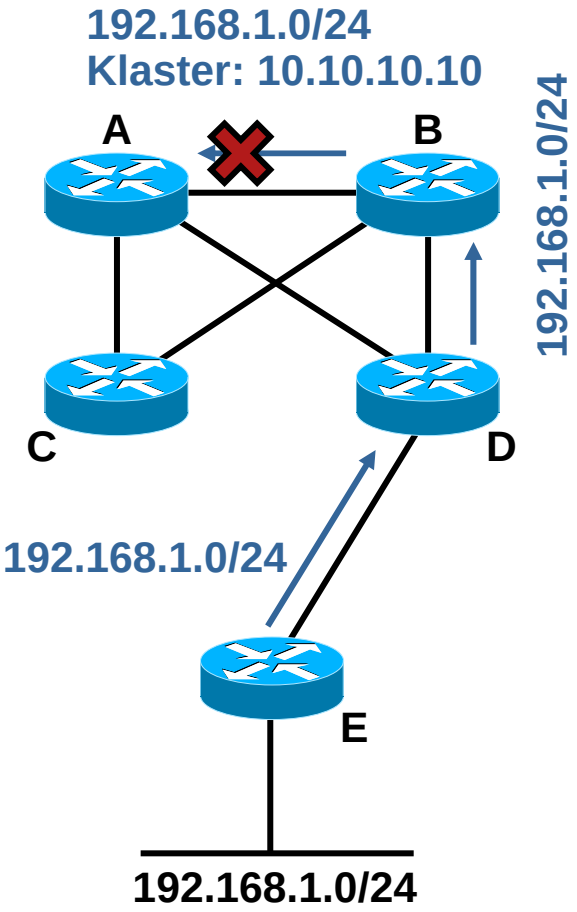



```
neighbor <c> route-reflector-client
neighbor <d> route-reflector-client
```

```
neighbor <c> route-reflector-client
neighbor <d> route-reflector-client
```

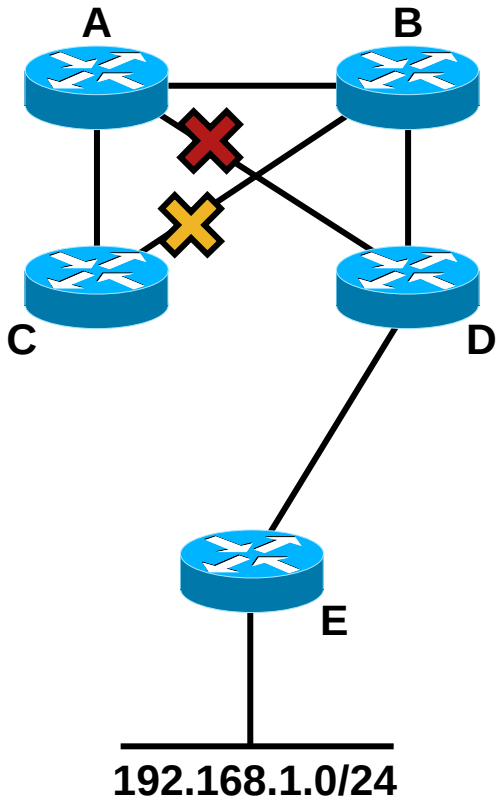
Redundancja RR

- **Załóżmy dalej, że mają to samo ClusterID równe 10.10.10.10**
- E rozgłasza 192.168.1.0/24 do D
- D wysyła ten prefiks do B
- B dodaje Cluster List i Originator ID, a następnie odbija tą trasę do A i C
- A odbierając trasę zauważa że lokalny ClusterID jest już na liście (skoro A i B mają ten sam) i odrzuca trasę



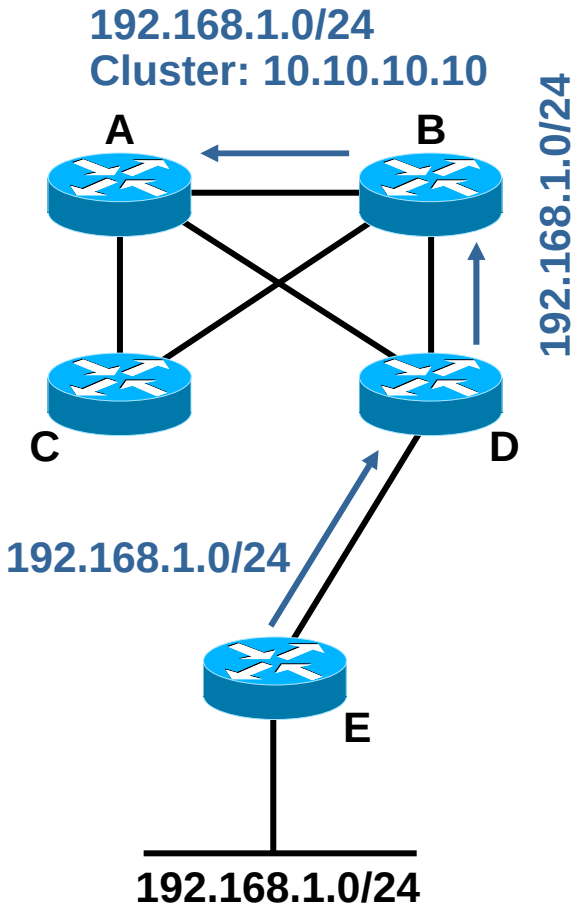
Redundancja RR

- Jeśli połączenie pomiędzy A i D zostanie zerwane, A nie będzie miał żadnej trasy do 192.168.1.0/24 ponieważ odrzucił trasę
- Jeśli połączenie między B i C zostanie zerwane, C nie bedzie miał również trasy do 192.168.1.0/24 z tego samego powodu

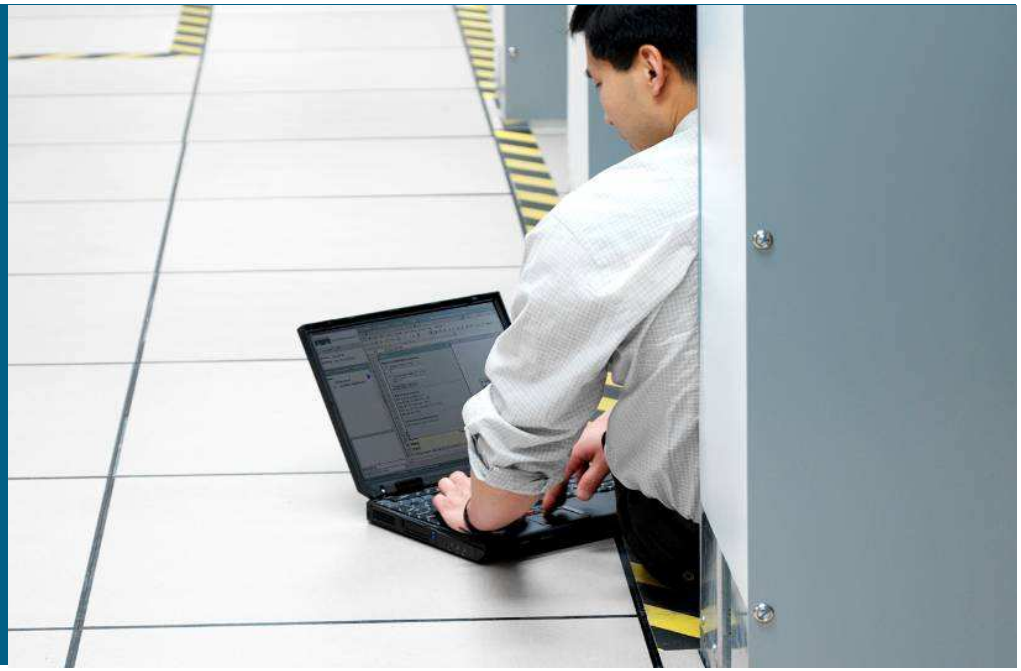


Redundancja RR

- Rekomenduje się konfigurację różnych ClusterID
- A odbierając trasę od B zatrzyma ją
- A wykona algorytm bestpath BGP i rozgłosi swoją ścieżkę do C przez D i B



Konfederacje

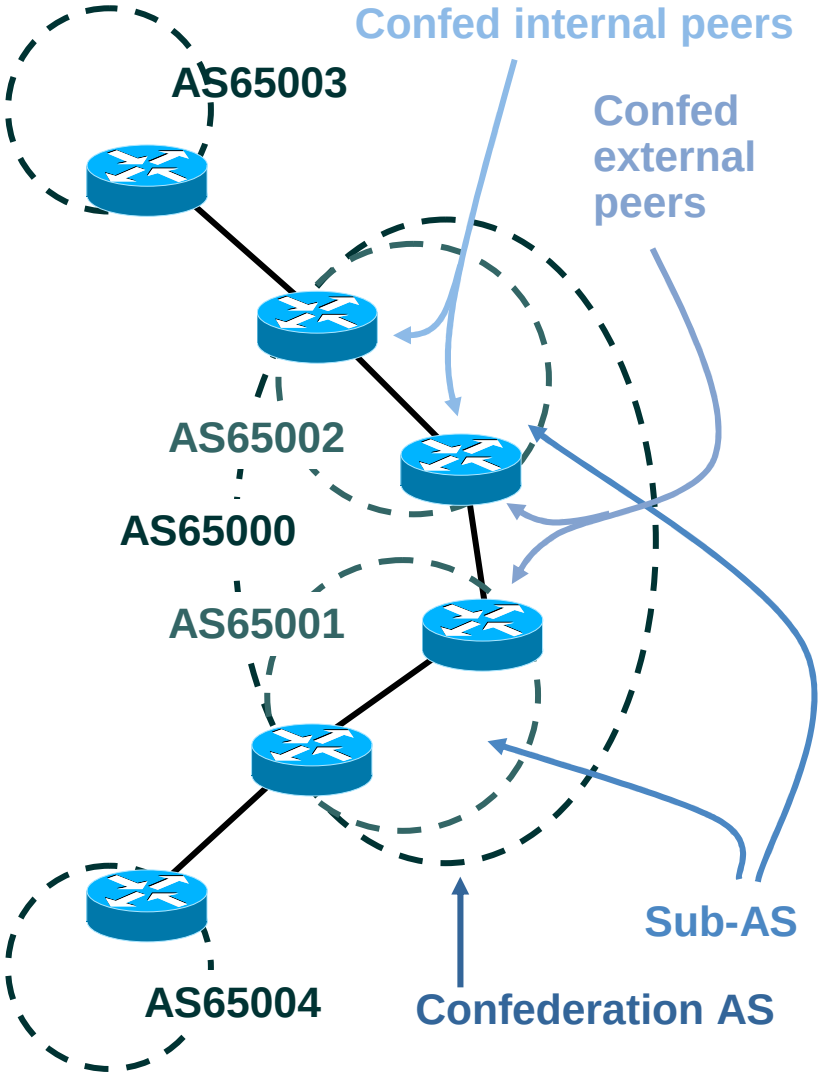


Konfederacje

- Konfederacje dostarczają innej opcji skalowania BGP wewnątrz AS
- Zamiast dodawać nowe atrybuty, konfederacje dodają informacje do AS Path
- Konfederacje definiuje RFC3065

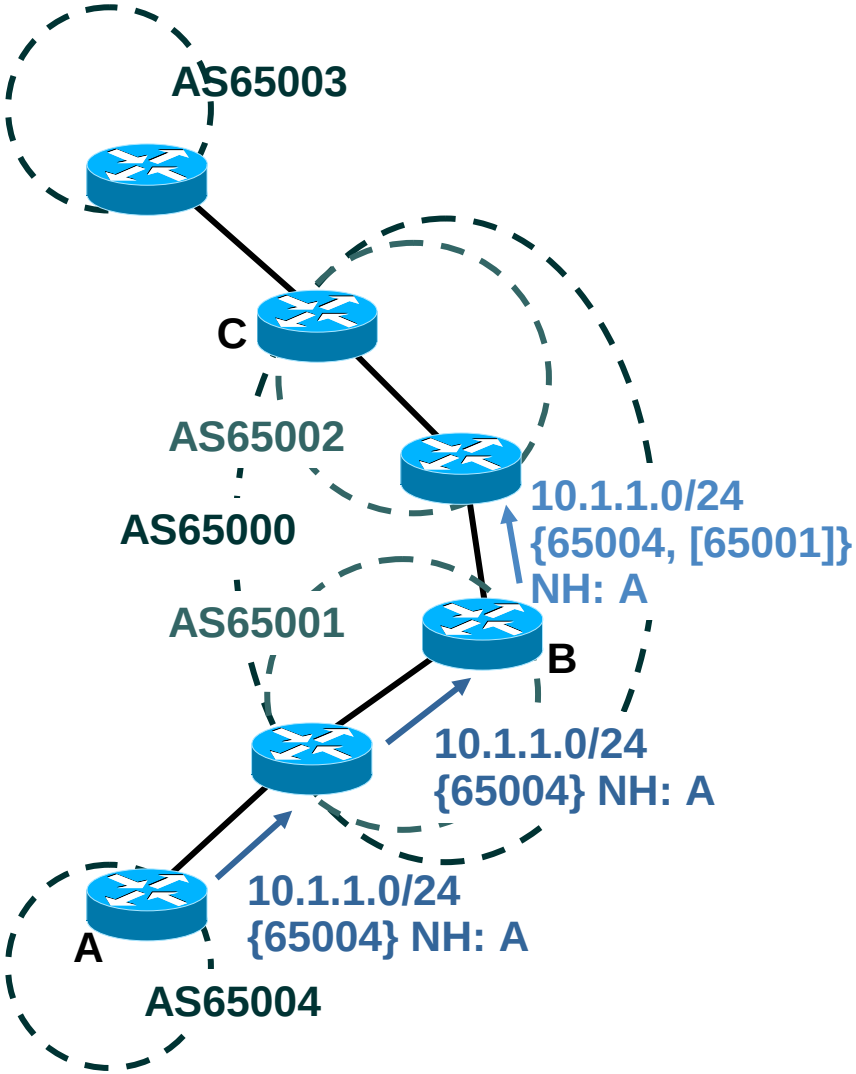
Konfederacje

- W konfederacji, duży AS zostaje podzielony
  - „Zewnętrzny” AS nazywany dalej **AS Konfederacji**
  - Wewnętrzne ASy – **sub-AS** (każdy ma swój numer)
- Sąsiedzi w tym samym sub-AS są **wewnętrznymi sąsiadami konfederacji**
- Sąsiedzi w różnych sub-AS są **zewnętrznymi sąsiadami konfederacji**



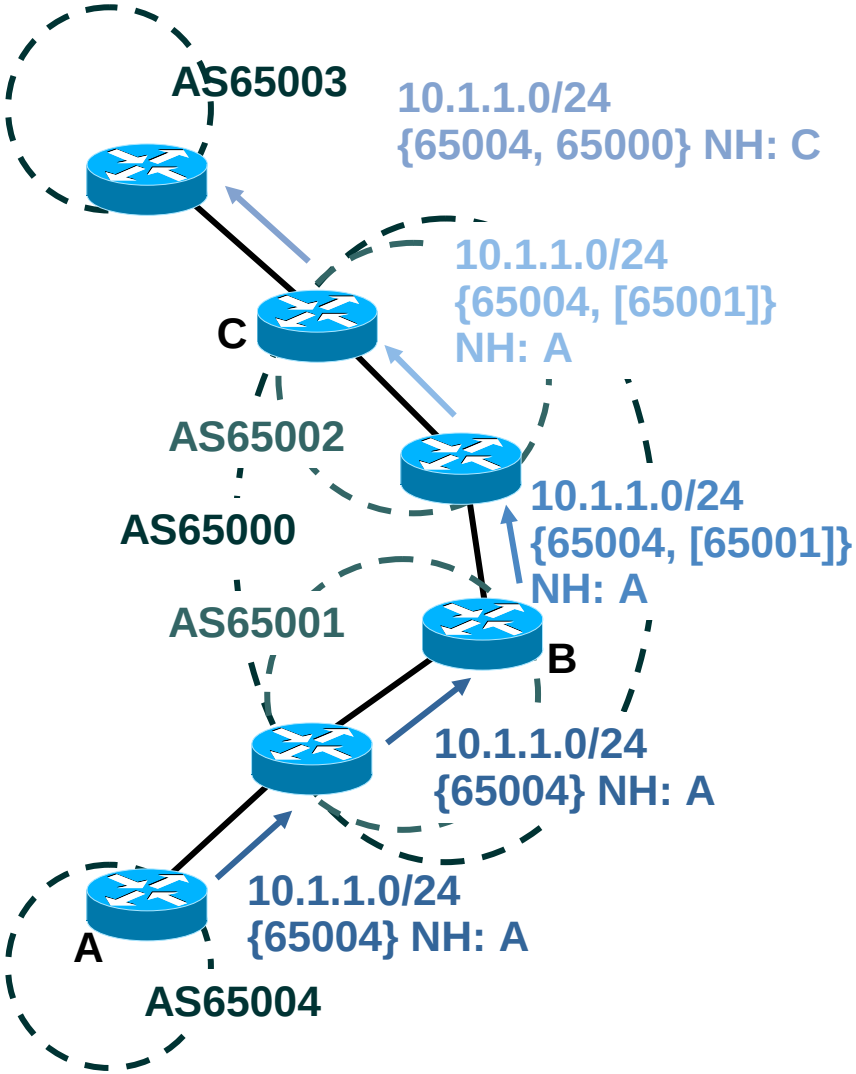
Konfederacje

- A wysyłając trasę dokleja AS65004 do AS Path z next-hop na siebie
- Wewnątrz sub-AS 65001 atrybuty pozostają niezmienne
- B tworzy **AS Confederation Sequence,** i dodaje sub-AS – 65001 do AS Path



Konfederacje

- W obrębie sub-AS 65002, atrybuty się nie zmieniają
- Next-hop pozostaje na A w obrębie całej konfederacji, nawet mimo tego, że trasa rozgłaszana jest przez wiele sub-ASów
- Na C AS Confederation Path jest usuwana z AS Path, dodawany jest AS konfederacji

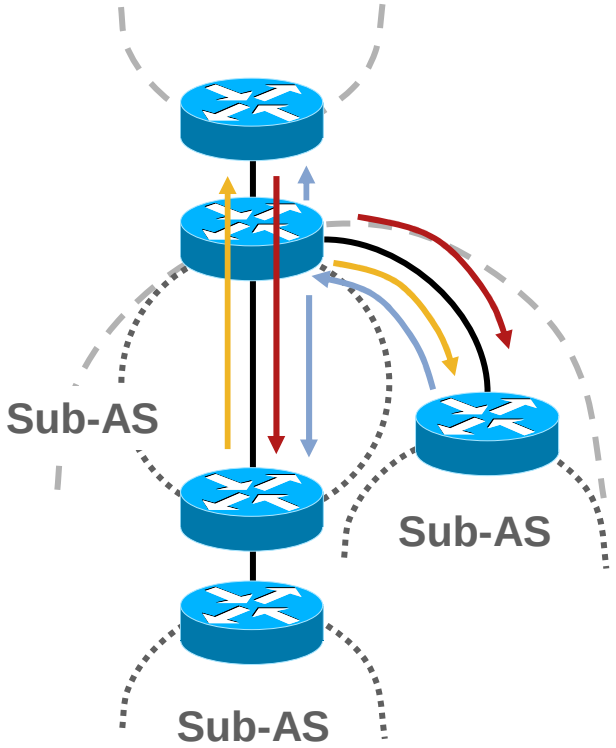


Konfederacje

- Zewnętrzni sąsiedzi konfederacji otrzymują trzy niezmienne atrybuty:
  - Next Hop
  - Local Preference
  - MED
- Dzięki temu możliwy jest przewidywalny i taki sam routing w konfederacji mimo jej wewnętrznego podziału
- Ponieważ nie zmieniamy next-hop, pojedyncza konfederacja powinna pracować pod kontrolą jednego IGP

Konfederacje

- Trasa poznana od sąsiada eBGP jest rozgłaszana do wszystkich wewnętrznych i zewnętrznych sąsiadów
- Trasa poznana od sąsiada w konfederacji (wewnętrznego) rozgłaszana jest do sąsiadów zewnętrznych w konfederacji i eBGP
- Trasa poznana od sąsiada w konfederacji (zewnętrznego) rozgłaszana jest do sąsiadów wewnętrznych w konfederacji i eBGP



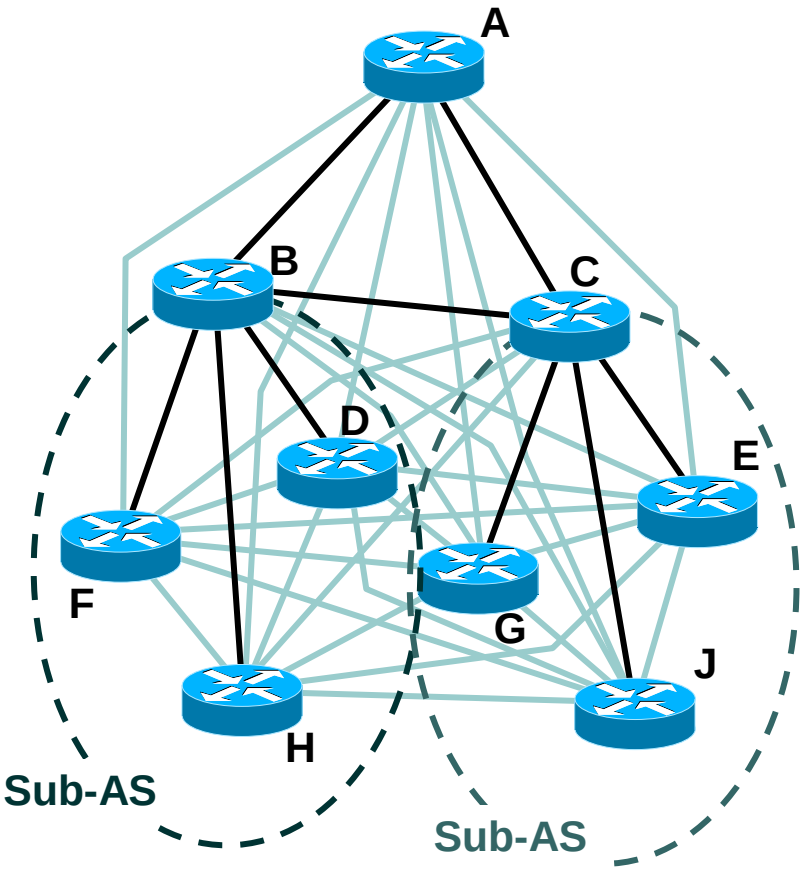
Trasa z eBGP

Trasa od sąsiada wewnętrznego

Trasa od sąsiada zewnętrznego

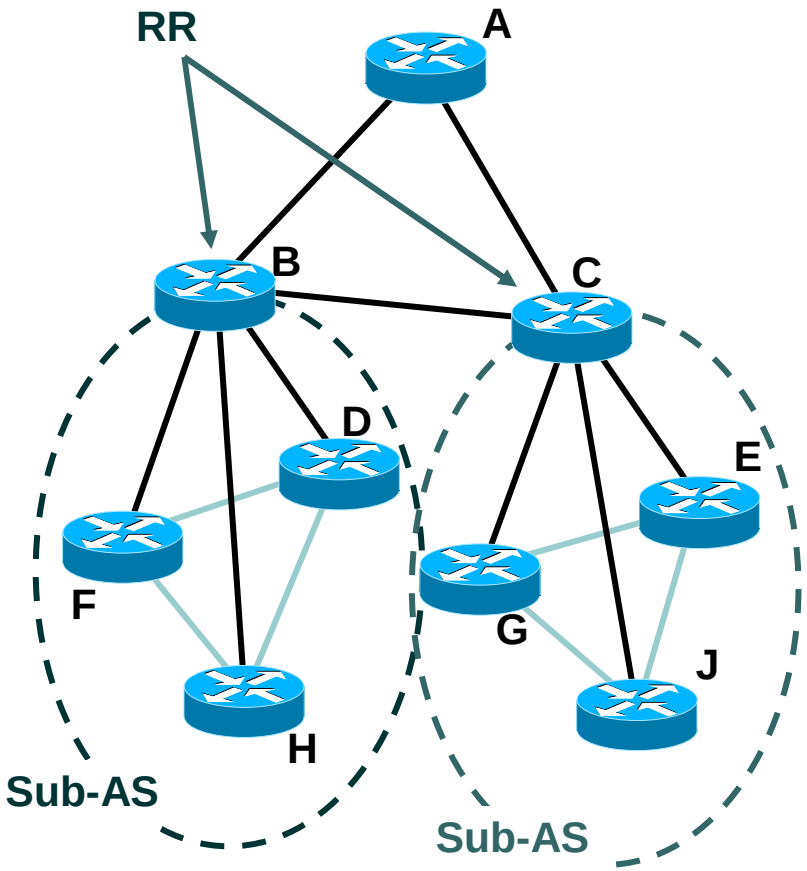
Konfederacje

- Wszystkie routery BGP w konfederacji muszą nawiązać pełną siatkę sąsiedztw iBGP
- Konfederacje zmniejszają jednak łączną ilość sesji, ponieważ routery w różnych subAS nie zestawiają ze sobą sesji (oprócz routerów brzegowych)

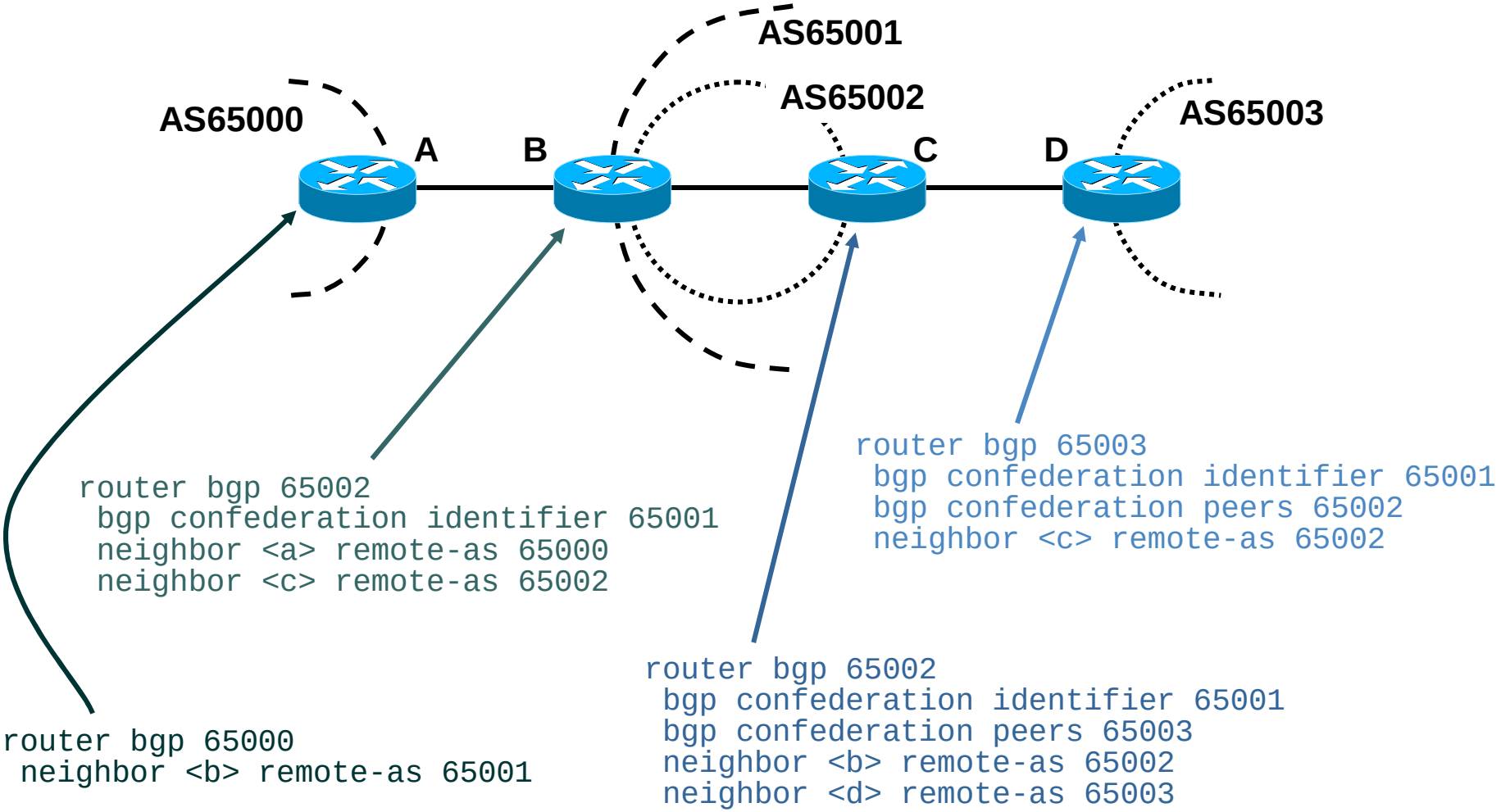


Konfederacje

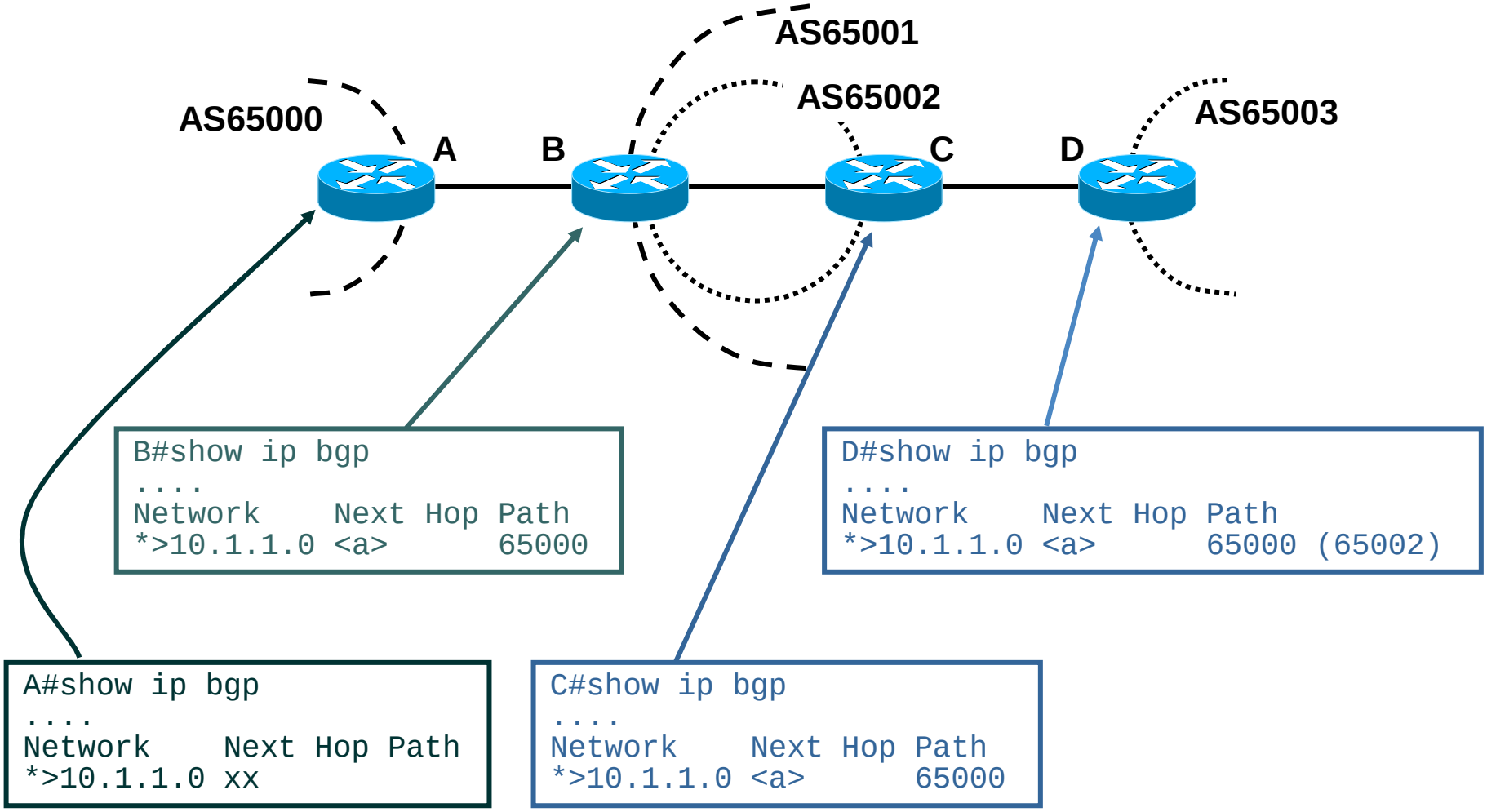
- W konfederacjach można zastosować RR żeby jeszcze bardziej zmniejszyć ilość sesji
- Dzieje się tak bardzo często w świecie rzeczywistym



Konfederacje

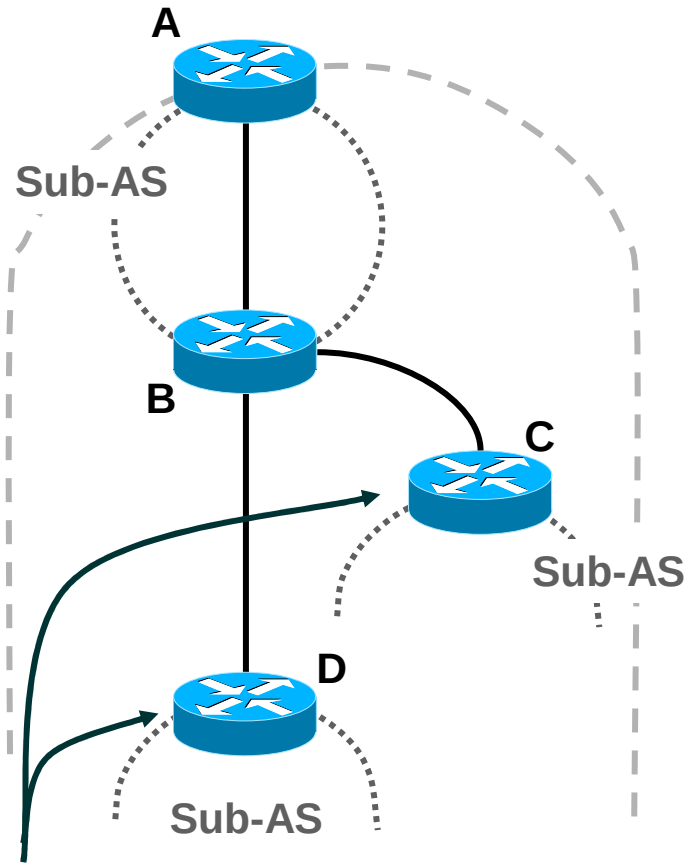


Konfederacje



Wdrażanie konfederacji

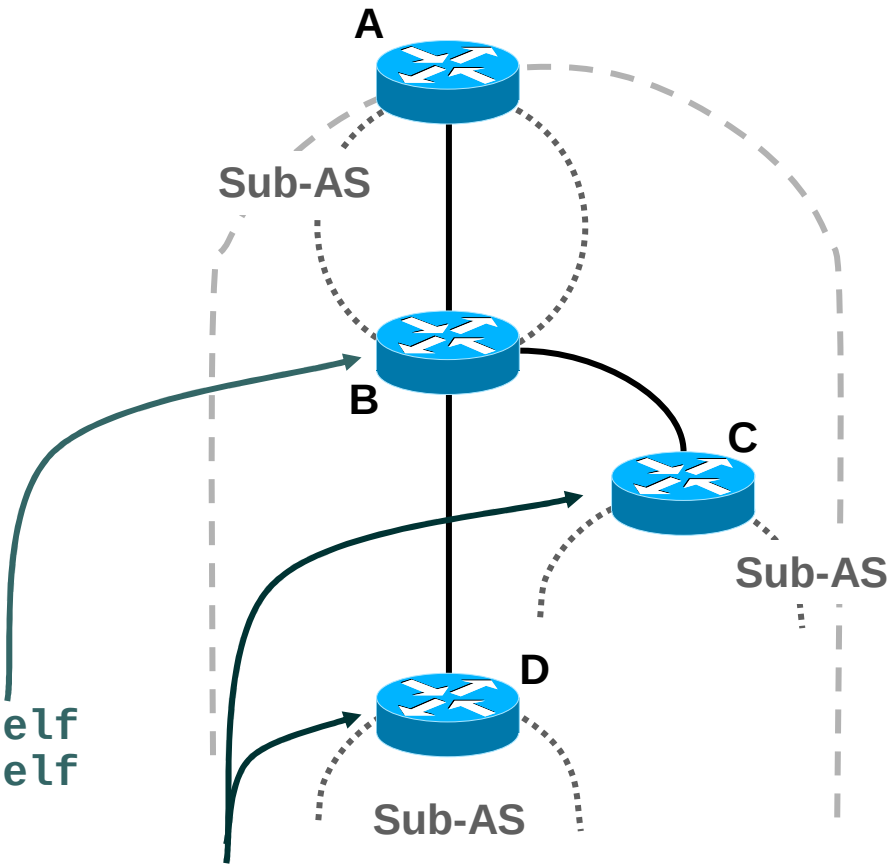
- W normalnych warunkach pole next-hop nie jest zmieniane w trakcie rozgłaszania trasy pomiędzy subASami
- Zwykle wymaga to uruchomienia jednego protokołu routingu IGP w konfederacji



Muszą być w stanie osiągnąć A

Wdrażanie konfederacji

- Oczywiście B może zmienić pole next-hop na siebie podczas rozgłaszania trasy do C i D
- **Po co chcielibyśmy zrobić coś takiego?**

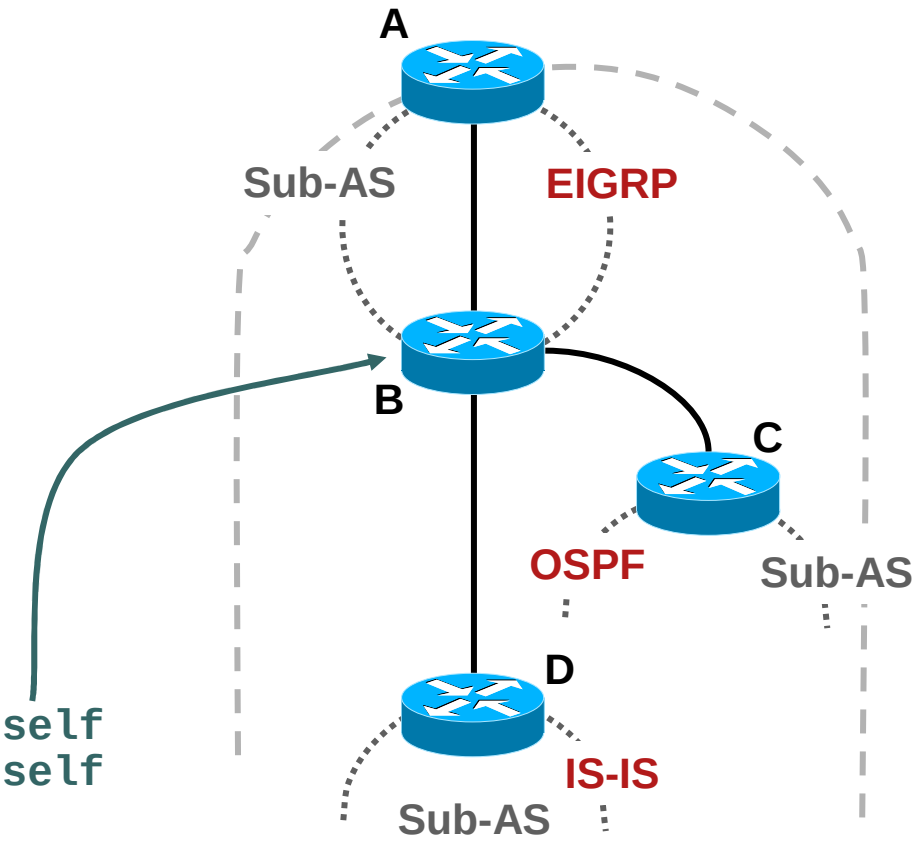



```
neighbor <c> next-hop self
neighbor <d> next-hop self
```

Muszą być w stanie osiągnąć A

Wdrażanie konfederacji

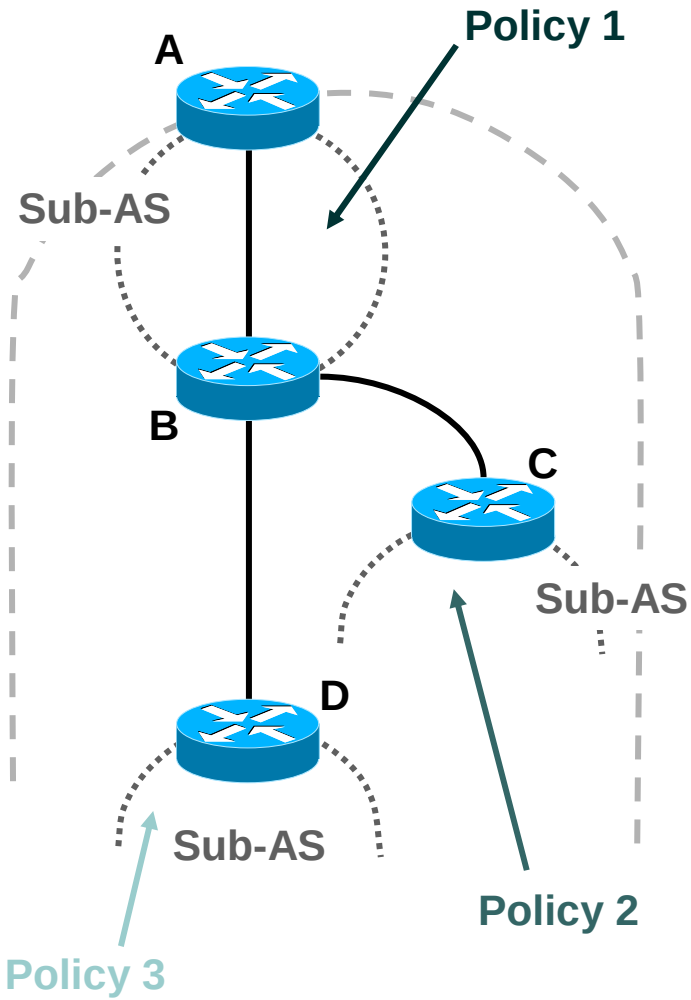
- Każdy subAS może używać innego protokołu IGP
- Z punktu widzenia IGP i BGP rozbija to konfederację, ale zwiększa skalowalność sieci




```
neighbor <c> next-hop self
neighbor <d> next-hop self
```

Wdrażanie konfederacji

- Każdy subAS może mieć również swoją politykę
  - akceptowanie lub usuwanie MED
  - local-preference
  - route dampening
  - itp.
- Polityka może dotyczyć również ruchu pomiędzy subASami



Konfederacje a RR

	Konfederacje	Route Reflector
Zapobieganie pętlom	AS Confederation Set	Originator/Cluster ID
Podział AS	Sub-AS	Klastry
Redundancja	Wiele połączeń pomiędzy sub-ASami	Połączenie klienta do wielu RR
Połączenia zewnętrzne	W dowolnym miejscu	W dowolnym miejscu
Hierarchia	RR w sub-AS	Klastry w klastrach
Kontrola polityki	Na brzegu AS i pomiędzy sub-ASami	Na brzegu AS
Skalowalność	Średnia – wymaga siatki iBGP w każdym sub-AS	Bardzo duża
Migracja	Bardzo trudna	Średnia/łatwa

Skalowanie BGP: konfederacje i RR

Warsztaty

Praktyka konfiguracji styku z Internetem

Prezentacja

Multihoming?

- Więcej niż jedno połączenie zewnętrzne do lokalnej sieci
  - Dwa lub większa ilość łącz od tego samego ISP
  - Dwa lub większa ilość łącz od różnych ISP
- Zwykle dwa routery brzegowe
  - Jeden router zabezpiecza tylko przed awarią ISP do niego podłączonych

Narzędzia kontroli ruchu

- Local preference

  Wybór miejsca wyjścia z własnej sieci

- Metric (MED)

  Wybór miejsca wejścia ruchu do sieci (lokalnie)

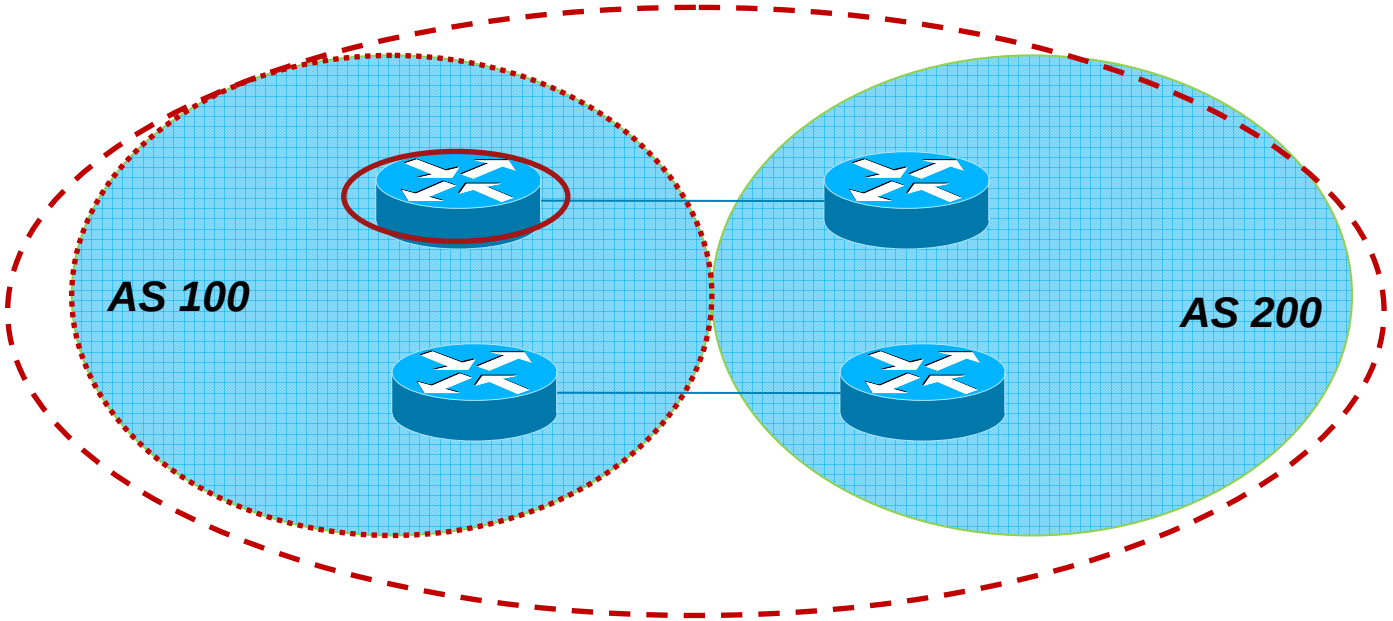
- AS-PATH prepend

  Wybór miejsca wejścia ruchu do sieci (globalnie)

- Communities

  Specyficzne ustalenia z ISP

Narzędzia kontroli ruchu



Weight

Local Preference

MED

Zastosowanie prywatnych numerów AS

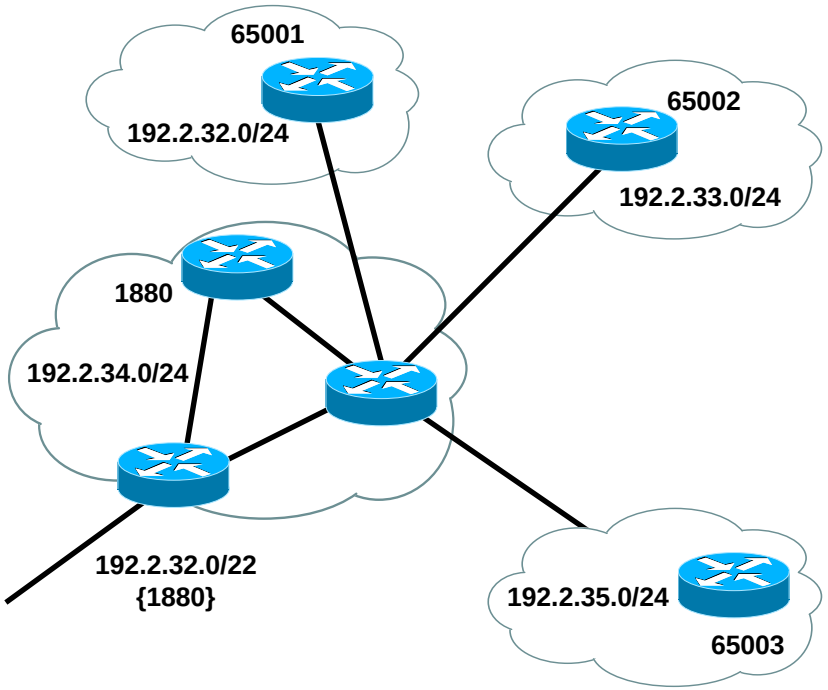
- ISP z klientami podłączonymi w wielu miejscach swojej sieci (RFC2270)

  -lub-

- Sieć firmowa z wieloma regionami ale połączeniem do Internetu tylko w jednym miejscu

  -lub-

- W konfederacji BGP



Zastosowanie prywatnych numerów AS

- Prywatne ASN **muszą** zostać usunięte ze wszystkich rozgłaszanych do Internetu prefiksów
- Dokładnie tak jak z przestrzenią adresową opisaną w RFC1918, prywatne ASN powinny być używane tylko wewnętrznie
- Cisco IOS

```
neighbor x.x.x.x remove-private-AS
```

Scenariusz: jedno łącze jako backup

- Na obu łączach akceptujemy tylko trasę domyślną (default)

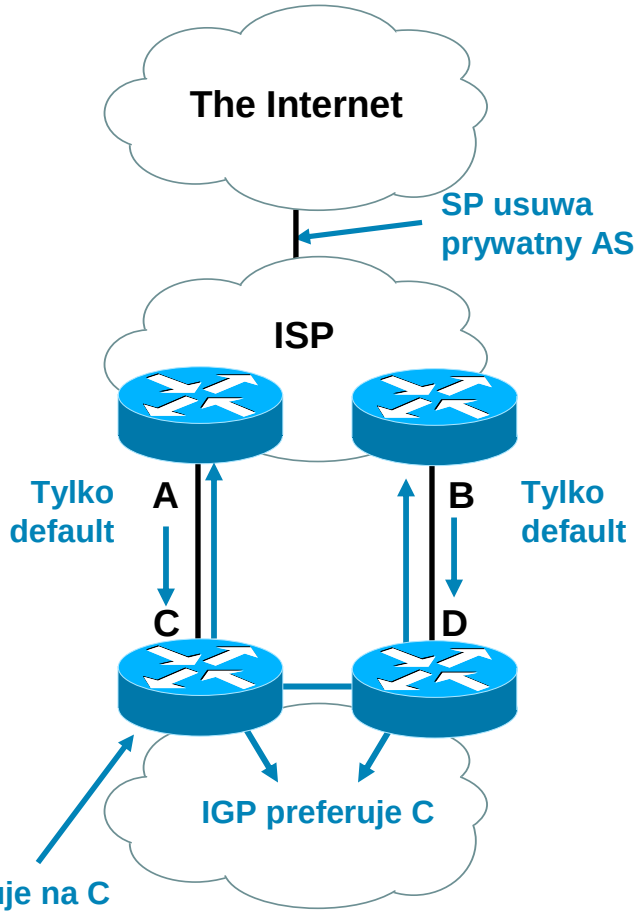
  Za pomocą local preference kierujemy ruch wyjściowy przez łącze „podstawowe”

  …a protokołem IGP kierujemy ruch w stronę routera

- Rozgłaszamy tą samą klasę adresową przez oba łącza

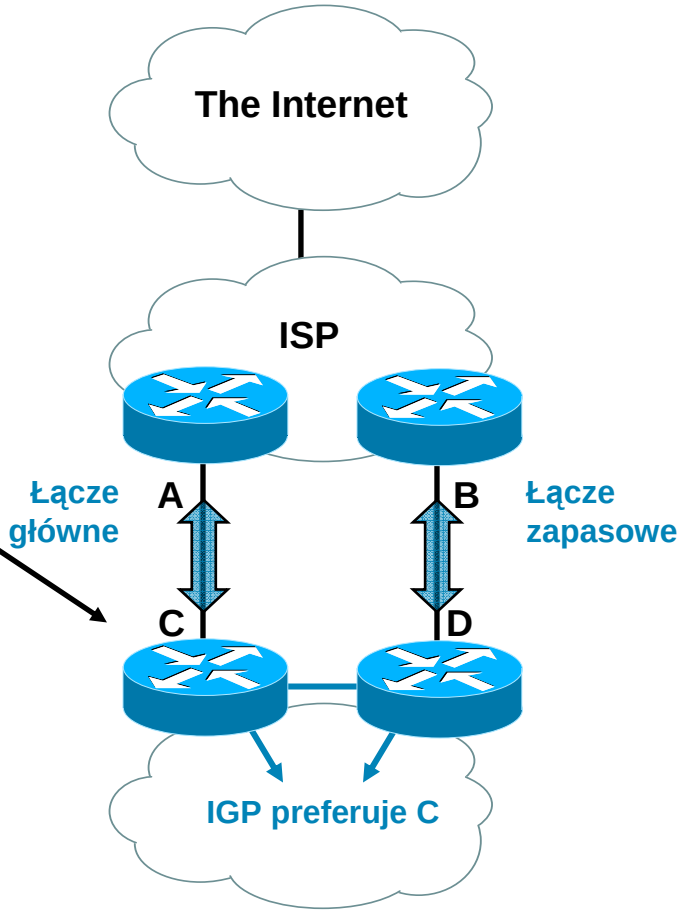
  To SP będzie preferował jedno łącze

  Inna możliwość to rozgłoszenie warunkowe



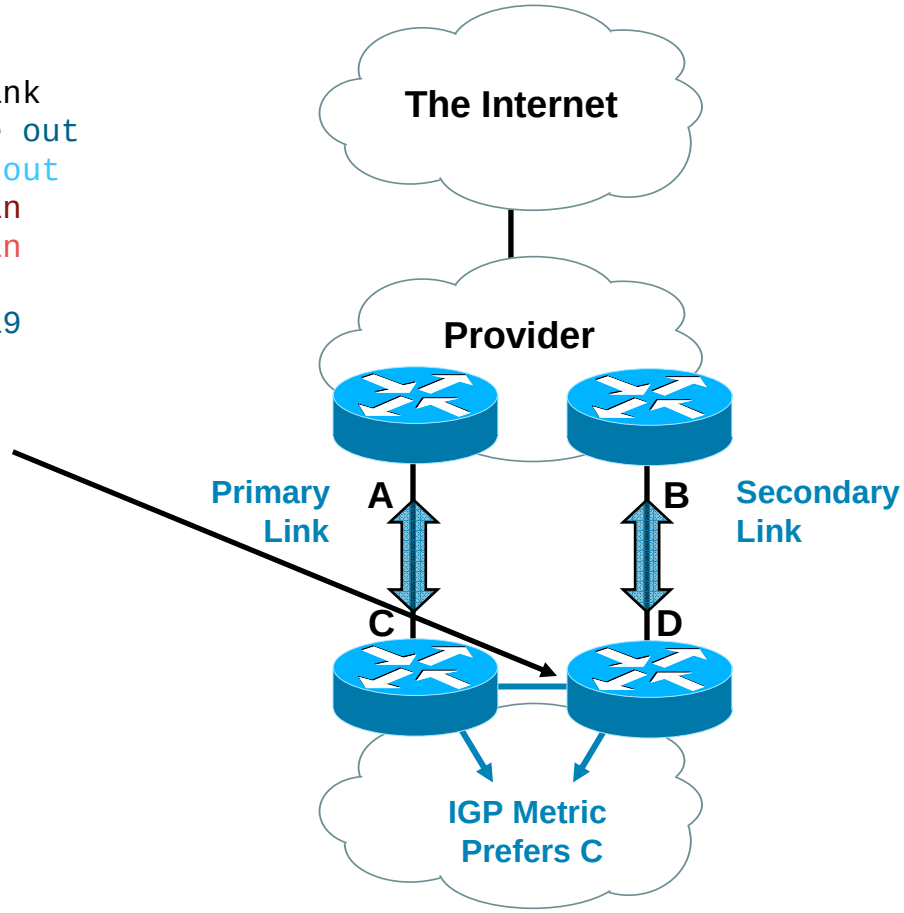
Scenariusz: jedno łącze jako backup

```
router bgp 65534
 network 121.10.0.0 mask 255.255.224.0
 neighbor 122.102.10.2 remote-as XXX
 neighbor 122.102.10.2 description primary-link
 neighbor 122.102.10.2 prefix-list aggregate out
 neighbor 122.102.10.2 prefix-list default in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip route 121.10.0.0 255.255.224.0 null0
```



Scenariusz: jedno łącze jako backup

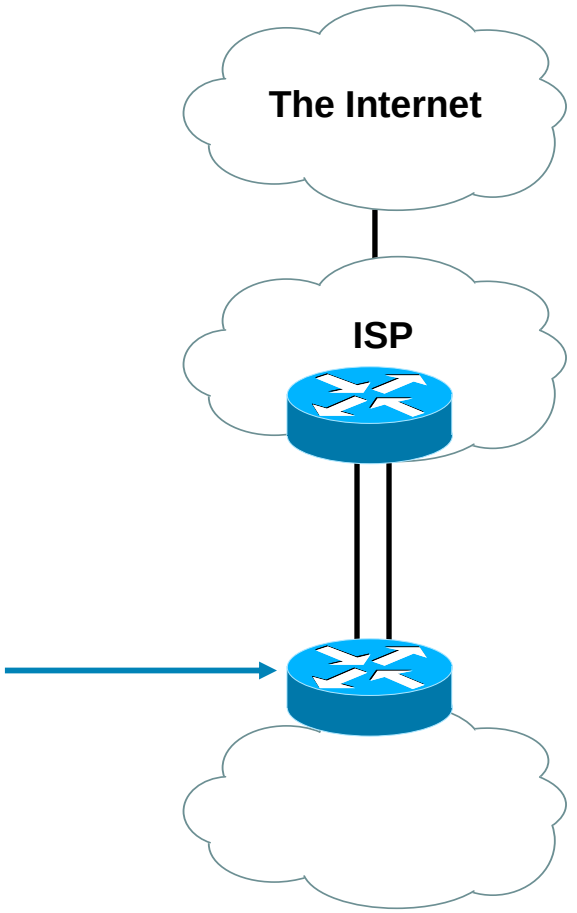
```
router bgp 65534
 network 121.10.0.0 mask 255.255.224.0
 neighbor 122.102.10.6 remote-as XXX
 neighbor 122.102.10.6 description backup-link
 neighbor 122.102.10.6 prefix-list aggregate out
 neighbor 122.102.10.6 route-map backup-out out
 neighbor 122.102.10.6 prefix-list default in
 neighbor 122.102.10.6 route-map backup-in in
!
ip prefix-list aggregate permit 121.10.0.0/19
ip prefix-list default permit 0.0.0.0/0
!
ip route 121.10.0.0 255.255.224.0 null0
!
route-map backup-out permit 10
 match ip address prefix-list aggregate
 set metric 10
route-map backup-out permit 20
!
route-map backup-in permit 10
 set local-preference 90
!
```



Scenariusz: rozkładanie obciążenia

- eBGP multihop – jeśli masz więcej niż jedno łącze pomiędzy parą routerów
  - eBGP do adresów loopback
  - prefiksy eBGP uczone są z adresem interfejsu loopback routera dostawcy

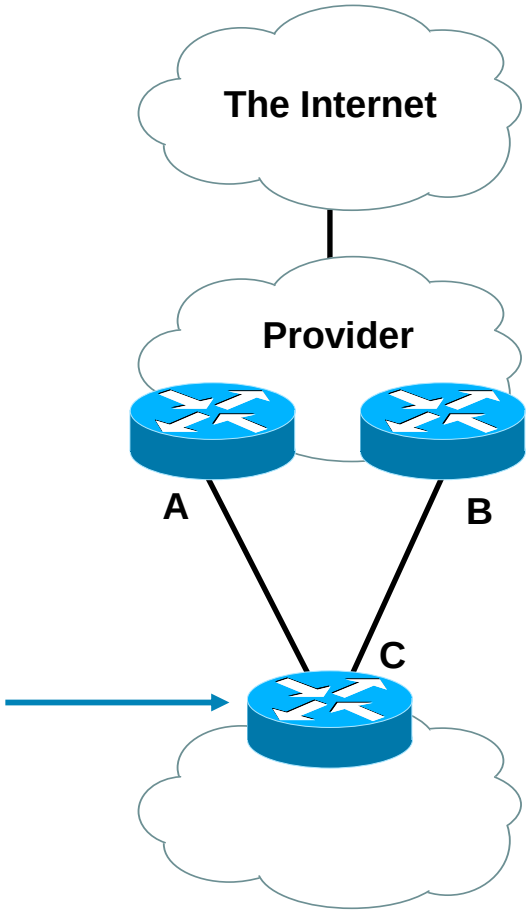
```
router bgp 65534
 neighbor 1.1.1.1 remote-as XXX
 neighbor 1.1.1.1 ebgp-multihop 2
 neighbor 1.1.1.1 update-soure Loopback0
!
ip route 1.1.1.1 255.255.255.255 serial 1/0
ip route 1.1.1.1 255.255.255.255 serial 1/1
```



Scenariusz: rozkładanie obciążenia

- Jeśli wiele łącz prowadzi do różnych routerów – eBGP multipath
  - Wiele sesji eBGP do tego samego dostawcy (ASN)
  - Sesje terminowane na tym samym routerze

```
router bgp 201
 neighbor 1.1.2.1 remote-as XXX
 neighbor 1.1.2.5 remote-as XXX
 maximum-paths 2
```

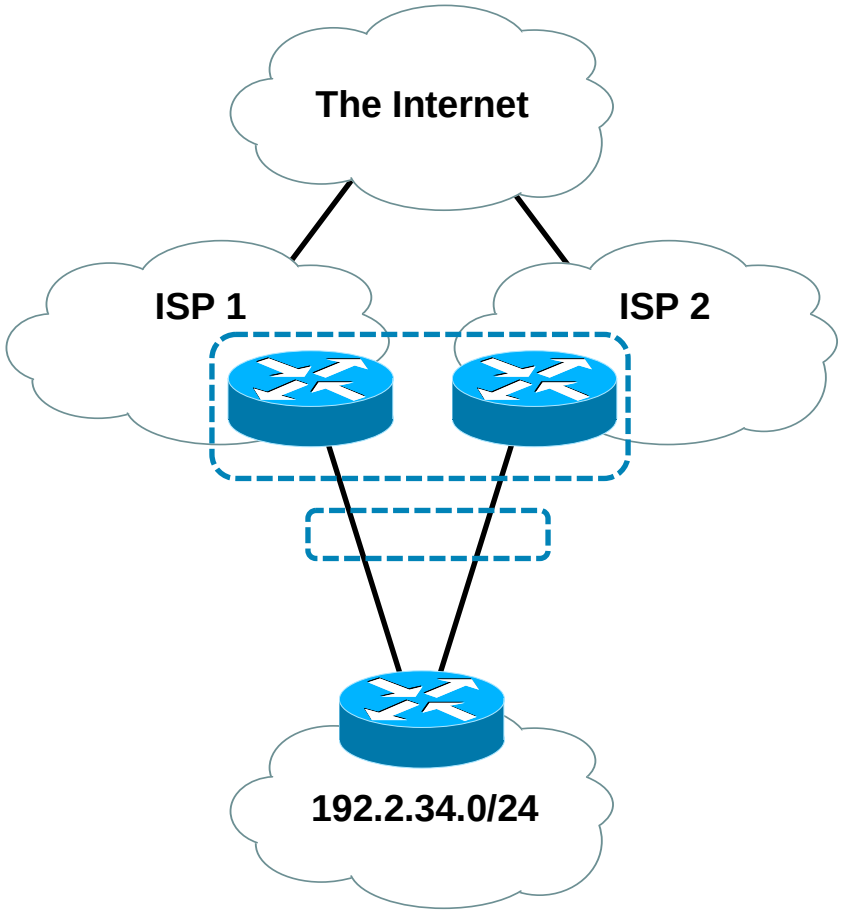


Wielu dostawców – parę kwestii

- Pojedynczy punkt awarii
- Nie zostań ASem tranzytowym…
- Jedno łącze jako backup
- Kontrola ruchu przychodzącego
- Kontrola ruchu wychodzącego

Pojedynczy punkt awarii

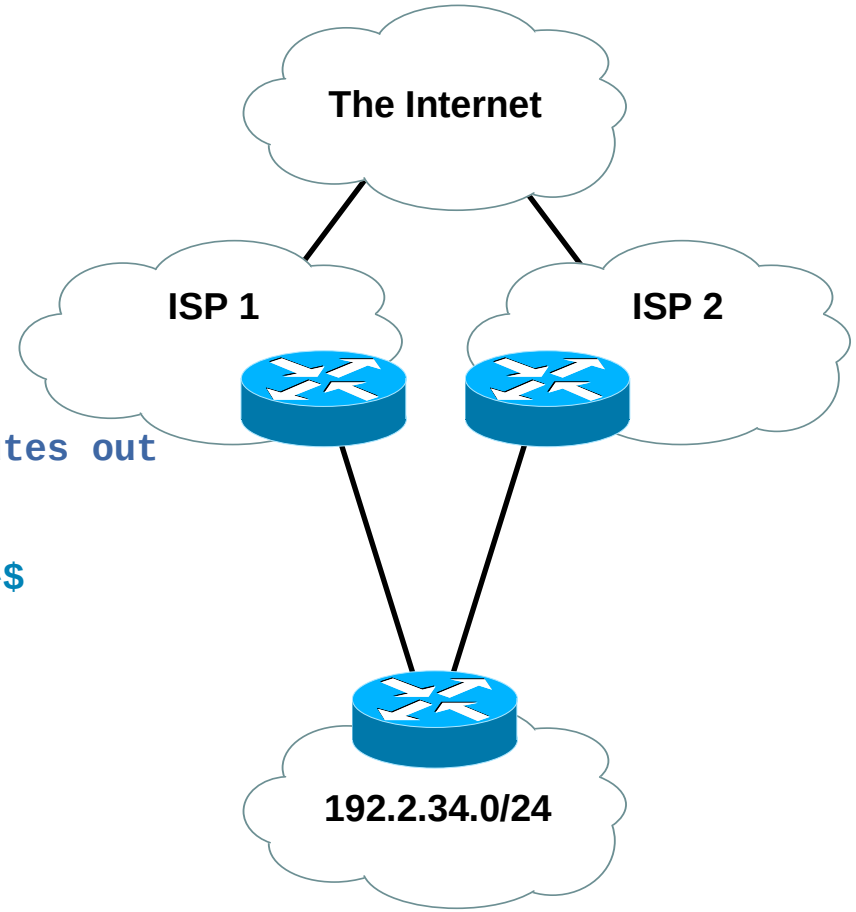
- Wielu dostawców wcale nie gwarantuje redundancji
- Tzw. „fate sharing” to między innymi:
  - jeden router
  - jeden zasilacz, lub dwa ale z jednym źródłem zasilania
  - Łącza fizyczne przez tą samą studzienkę



Nie zostań ASem tranzytowym

- Przykładowa konfiguracja:

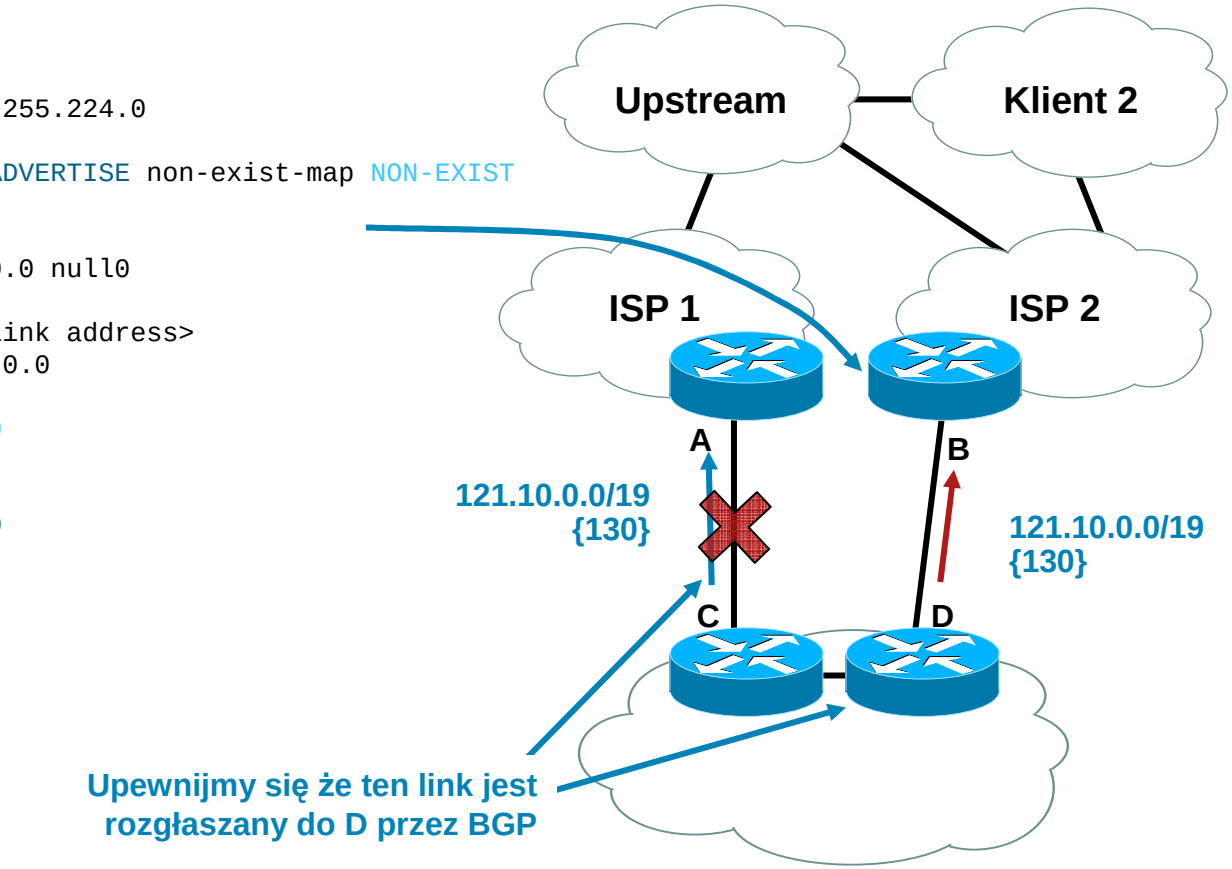
```
router bgp 130
 neighbor x.x.x.x remote-as 1
 neighbor x.x.x.x route-map my_routes out
!
ip as-path access-list 20 permit ^$
ip as-path access-list 20 deny .*
!
route-map my_routes permit 10
 match as-path 20
```



Jedna trasa jako backup

Rozgłoszenie warunkowe

```
router bgp 130
 bgp log-neighbor-changes
 network 121.10.0.0 mask 255.255.224.0
 neighbor <B> remote-as XXX
 neighbor <B> advertise-map ADVERTISE non-exist-map NON-EXIST
 neighbor <C> remote-as 130
!
ip route 10.1.0.0 255.255.240.0 null0
!
access-list 60 permit <a->c link address>
access-list 65 permit 121.10.0.0
!
route-map NON-EXIST permit 10
 match ip address 60
!
route-map ADVERTISE permit 10
 match ip address 65
```

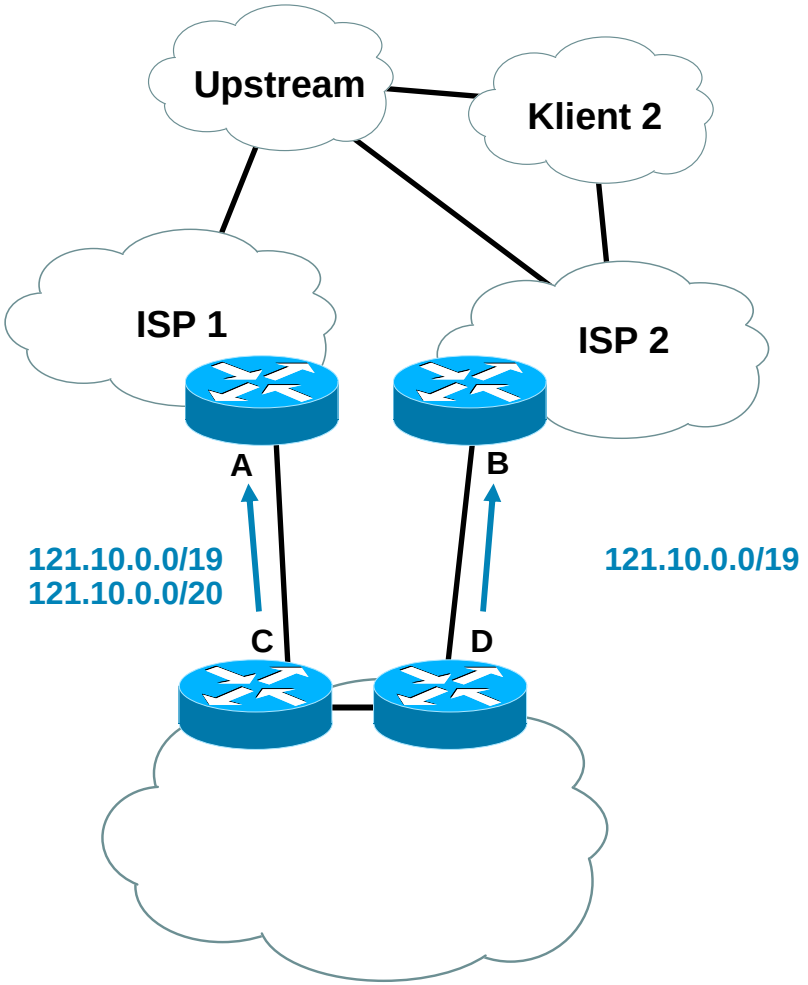


Równoważenie ruchu przychodzącego

- Dokładniejsze prefiksy

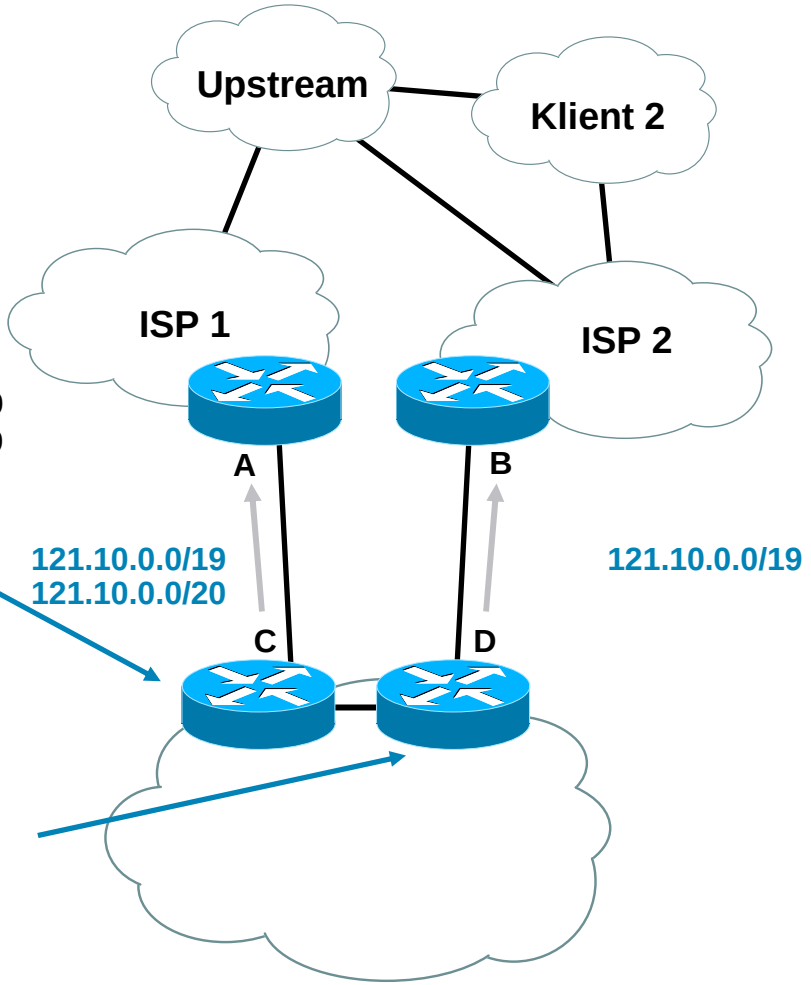
  Rozgłoś 121.10.0.0/19 i 121.10.0.0/20 przez jedno połączenie

  Rozgłoś 121.10.0.0/19 przez drugie połączenie



Równoważenie ruchu przychodzącego

```
router bgp 65555
 network 121.10.0.0 mask 255.255.224.0
 network 121.10.0.0 mask 255.255.240.0
 neighbor x.x.x.x remote-as <provider 1>
 neighbor x.x.x.x prefix-list firstblock out
!
ip prefix-list firstblock permit 121.10.0.0/20
ip prefix-list firstblock permit 121.10.0.0/19
```




```
router bgp 130
 network 121.10.0.0 mask 255.255.224.0
 neighbor x.x.x.x remote-as <provider 2>
 neighbor x.x.x.x prefix-list secondblock out
!
ip prefix-list secondblock permit 121.10.0.0/19
```

Równoważenie ruchu przychodzącego

- „Pogorszenie” atrakcyjności własnego prefiksu 10.10.10.0/24 przez AS200

```
router bgp 100
 neighbor 172.16.10.1 remote-as 200
 neighbor 172.16.10.1 route-map bgp-prepend-2x out

route-map bgp-prepend-2x permit 10
 set as-path prepend 100 100

rtr-AS200# show ip bgp 10.0.10.0
   Network          Next Hop     LocPrf  Path
*  10.0.10.0/24     172.16.10.2  100     100 100 100 i
*>                  172.16.99.1  100     300 100 i
```

Równoważenie ruchu przychodzącego

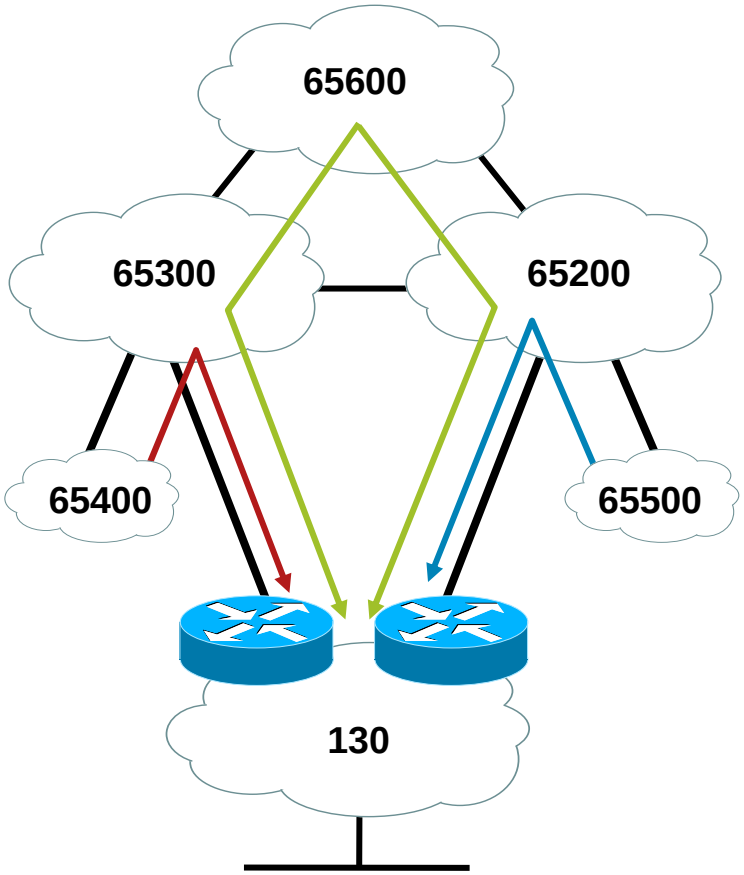
- Problem z AS prepend

  Ruch z AS65500 będzie nadal przechodziło przez 65200

  Ruch z AS65400 będzie nadal przechodził przez AS65300

  Tylko ruch z AS65600 zostanie zmieniony - przez AS path

  Może, nie musi działać



Ruch przychodzący

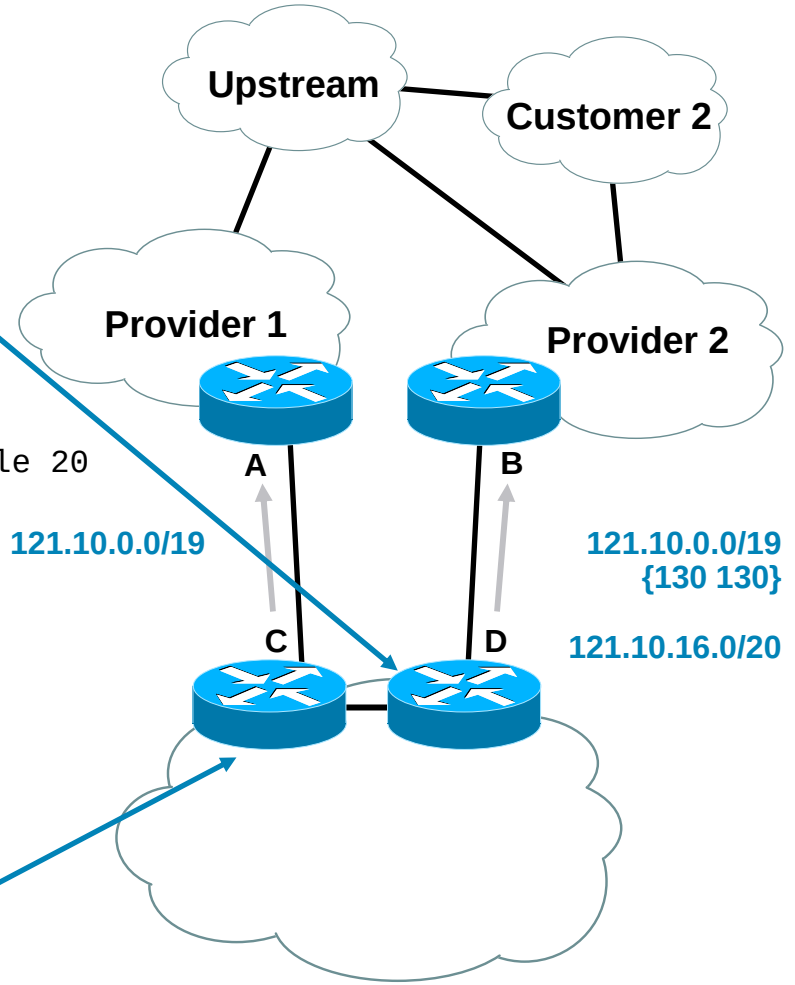
- Tylko prefiks 8.8.8.0/24 lub bardziej dokładny jest zawsze lepszy przez AS300

```
router bgp 100
 neighbor 172.16.20.1 remote-as 300
 neighbor 172.16.20.1 route-map bgp-8880-local-preference in

ip prefix-list JedynaDroga permit 8.8.8.0/24 le 32

route-map bgp-8880-local-preference permit 10
 match ip prefix-list JedynaDroga
 set local-preference 5000
```

Dobry opis jak działają prefix-listy:
http://www.groupstudy.com/archives/ccielab/200404/msg00539.html

Rozwiązania można łączyć

```
router bgp 130
 network 121.10.0.0 mask 255.255.224.0
 network 121.10.16.0 mask 255.255.240.0
 neighbor x.x.x.x remote-as <provider 2>
 neighbor x.x.x.x prefix-list subblocks out
 neighbor x.x.x.x route-map traffic-eng out
!
route-map traffic-eng permit 10
 match ip address prefix-list aggregate
 set as-path prepend 130 130
route-map traffic-eng permit 20
!
ip prefix-list subblocks permit 121.10.0.0/19 le 20
ip prefix-list aggregate permit 121.10.0.0/19
```

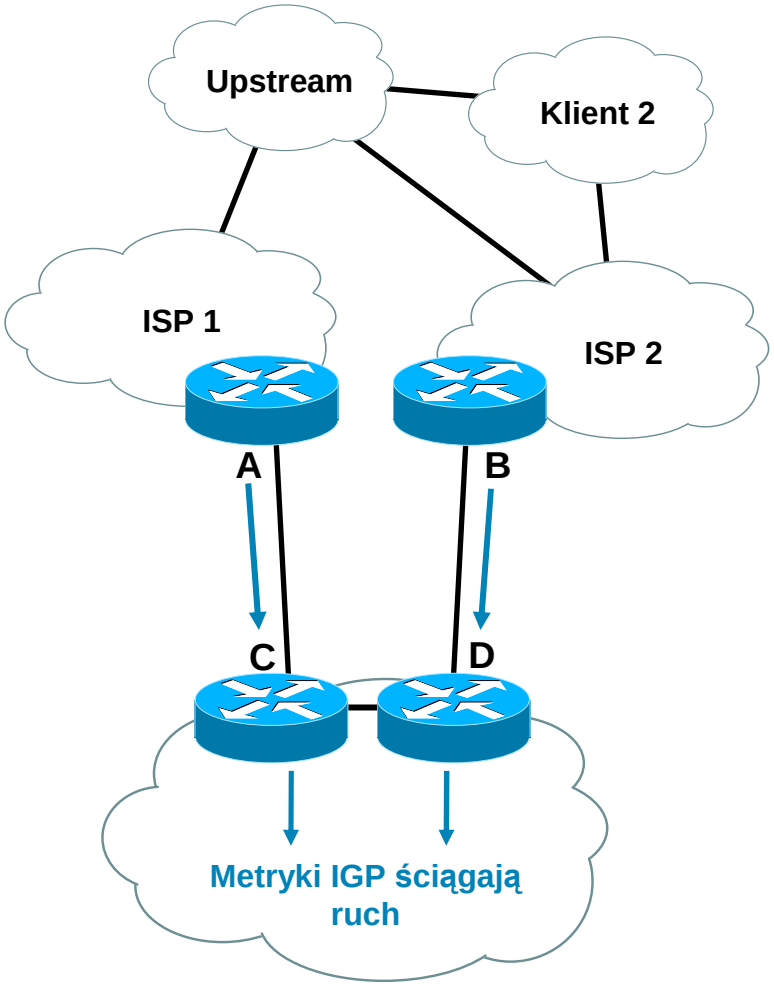



```
router bgp 130
 network 121.10.0.0 mask 255.255.224.0
 neighbor x.x.x.x remote-as <provider 1>
 neighbor x.x.x.x prefix-list aggregate out
!
ip prefix-list aggregate permit 121.10.0.0/19
```

Ruch wychodzący

- Jak rozgłosić swoje prefiksy?
- Pierwsza opcja: akceptujemy tylko trasy domyślne

  Metryka protokołów IGP powoduje ściągnięcie ruchu do konkretnych punktów wyjścia

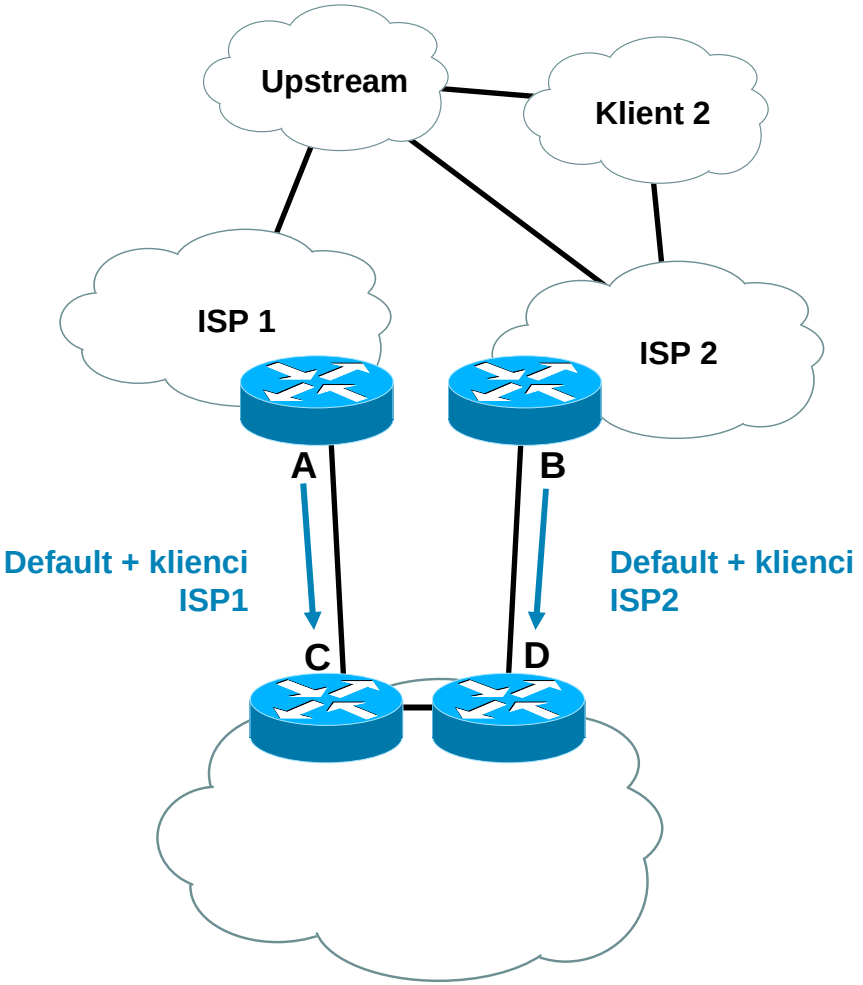


Ruch wychodzący

- Druga opcja: częściowe prefiksy

  Domyślna trasa plus sieci bezpośrednio podłączone do operatora

  W tym przypadku – ISP2 wyśle sieci klienta 2



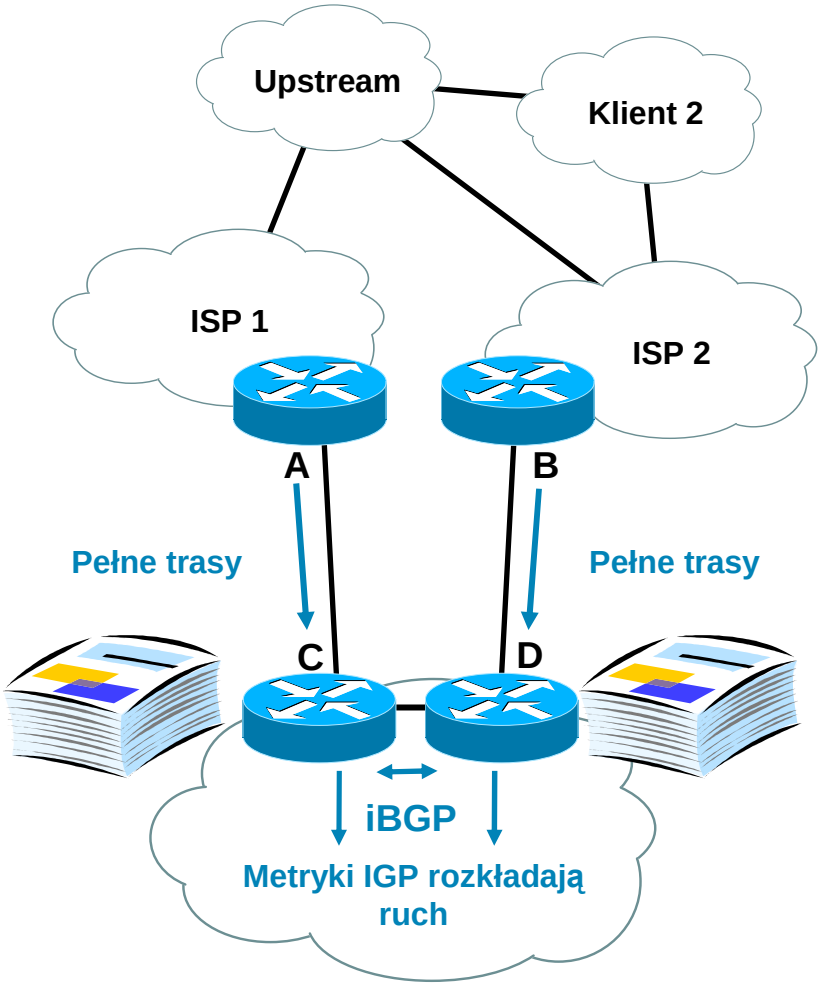
Ruch wychodzący

- Wszystkie prefiksy otrzymane z AS200 są lepsze niż inne, z domyślnym (100) local-preference

```
router bgp 100
 neighbor 172.16.10.1 remote-as 200
 neighbor 172.16.10.1 route-map bgp-local-preference in

route-map bgp-local-preference permit 10
 set local-preference 5000
```

Ruch wychodzący

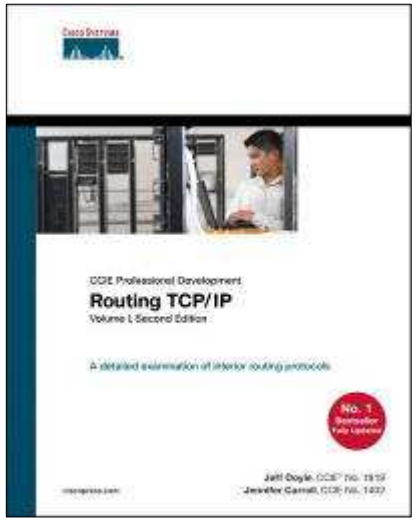
- Pełne trasy
- Dokładna kontrola gdzie wysyłać ruch
- Niezbędne tylko gdy jesteś ASem tranzytowym, nie muszą być konieczne dla klienta chcącego optymalizować ruch

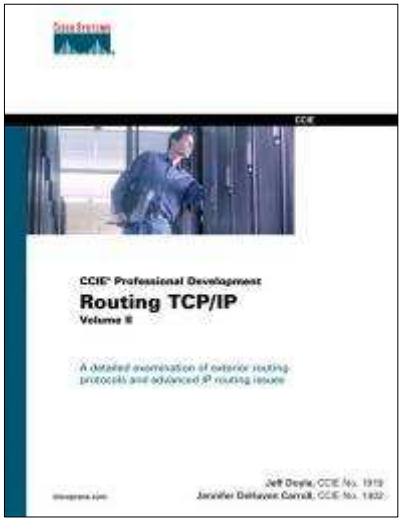


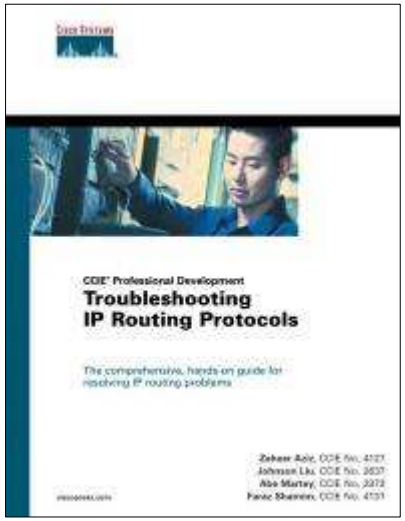
Praktyka konfiguracji styku z Internetem

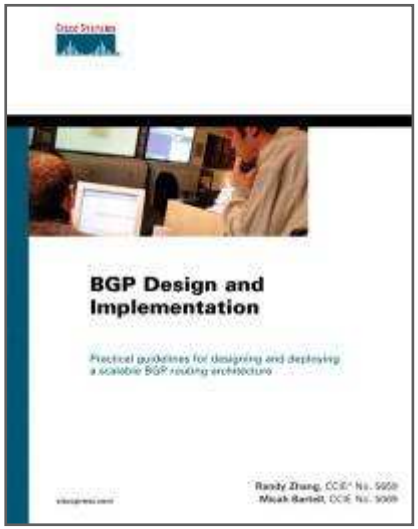
Warsztaty

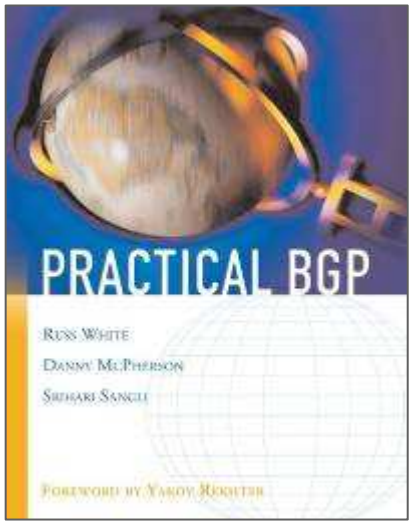
Książki

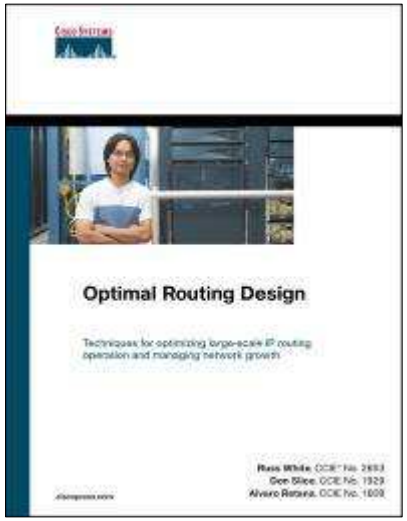


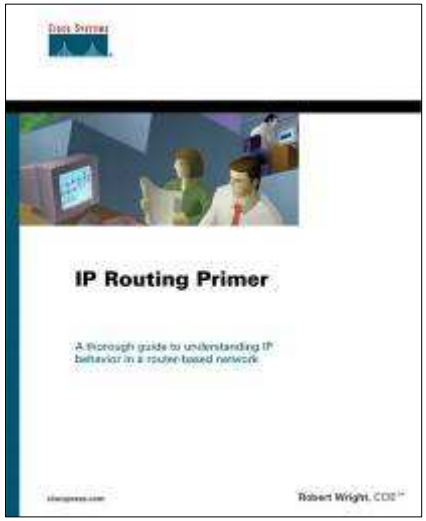


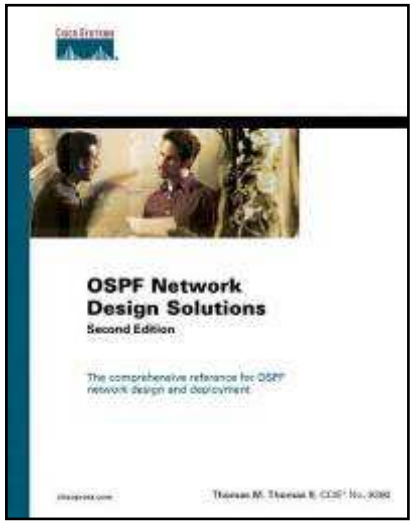


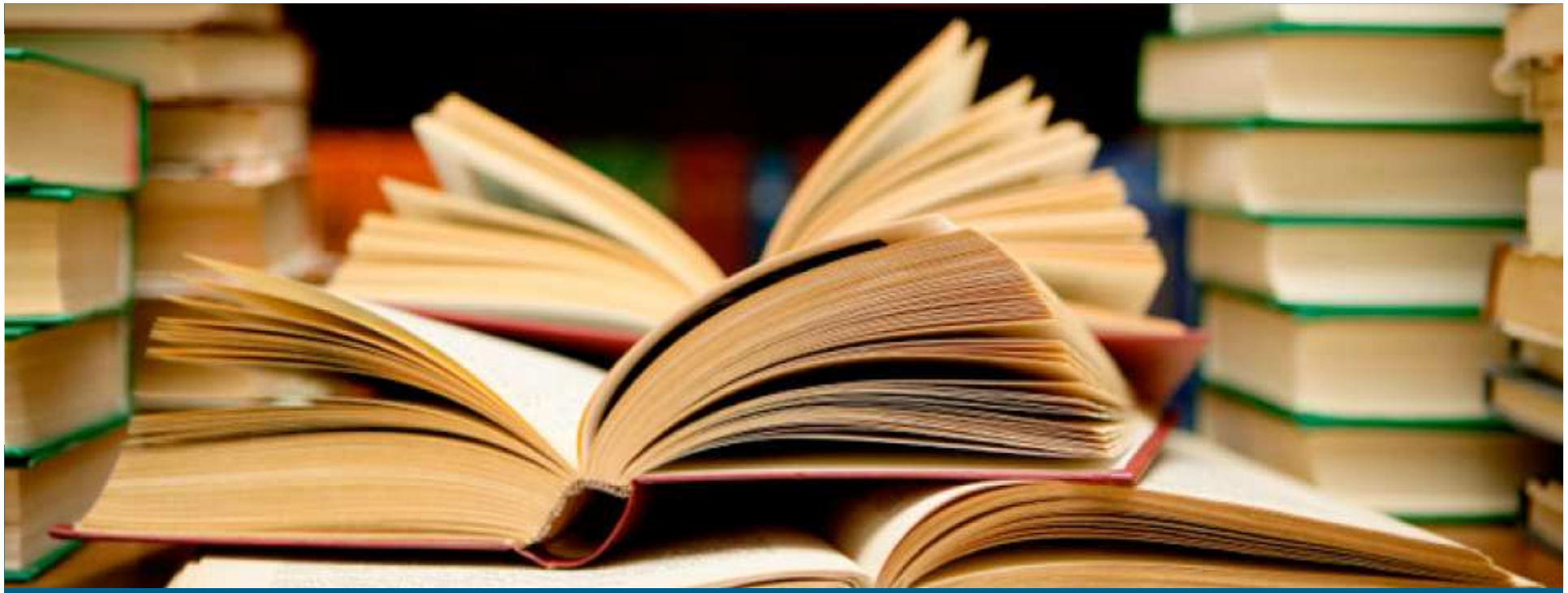












Q&A