

SIECI KOMPUTEROWE

**WPROWADZENIE
DO PROTOKOŁU
BGP**

(BORDER GATEWAY PROTOCOL)

1. Wstęp

Podstawowe zagadnienia

- Zadania warstwy sieciowej modelu OSI/ISO
- Podział routingu
- Metryka
- Routing bezklasowy, CIDR
- System autonomiczny (AS)

Podział protokołów routingu

- Kryterium I (zasada działania)
 - Wektor-odległość: RIP, IGRP
 - Stanu łącza: OSPF, IS-IS
 - Hybrydowe: EIGRP
- Kryterium II (obszar stosowania)
 - W ramach jednego systemu autonomicznego, klasa IGP: RIP, IGRP, EIGRP, OSPF, IS-IS
 - Pomędzy systemami autonomicznymi, klasa EGP: EGP, BGP

Komunikacja w protokołach IGP

- RIP, IGRP: regularne przesyłanie informacji o znanych sobie sieciach
- OSPF, EIGRP:
 - Regularnie wysyłane *hello* informujące o aktywności sąsiada
 - Przesyłanie informacji o zmianach tylko po ich wystąpieniu (i w czasie inicjacji)

Komunikacja w protokołach IGP

- RIP v1
 - Adres: 255.255.255.255
 - Enkapsulacja: UDP/IP (port 520)
- RIP v2
 - Adres: 224.0.0.9
 - Enkapsulacja: UDP/IP (port 520)
- OSPF
 - Adres: 224.0.0.5 i 224.0.0.6
 - Enkapsulacja: IP
- IGRP
 - Adres: 255.255.255.255
 - Enkapsulacja: IP

System autonomiczny

- Obszar sieci będący pod wspólną władzą administracyjną i zarządzany w jednolity sposób
- Identyfikowany przez numer: liczbę 16-to bitową
 - 1-64511 – numery publiczne
 - przyznawane przez RIR (Regional Internet Registries)
 - 8323 – CYFRONET-AS2 Metropolitan Area Network
 - 8267 – CYFRONET-AS Metropolitan Area Network
 - 8364 – poznański MAN
 - muszą jednoznacznie identyfikować właściciela
 - 64512-65534 – numery prywatne
 - mogą być wykorzystywane wielokrotnie
- Rozważa się wprowadzenie 32 bitowych numerów AS

Ważne spostrzeżenie

- *W obszarze zastosowania protokołów EGP nie ma jednej władzy administracyjnej*

ACK Cyfronet

- CYFRONET-AS Metropolitan Area Network
 - 149.156.0.0/16
 - 192.86.14.0/24
 - 192.245.169.0/24
 - 195.150.224.0/19
- CYFRONET-AS2 Metropolitan Area Network
 - 193.193.64.0/19
 - 194.8.45.0/24
 - 194.8.46.0/24
 - 195.150.0.0/16

<http://bgp.potaroo.net/cgi-bin/as-report?as=AS8323>

Połączenia AS 8267

```
import: from AS8501 action pref=100; accept ANY
import: from AS8323 action pref=100; accept ANY
import: from AS6778 action pref=200; accept AS6778
import: from AS5550 action pref=200; accept AS5550
import: from AS8508 action pref=250; accept AS8508
import: from AS12631 action pref=250; accept AS12631
export: to AS8323 announce ANY
export: to AS8501 announce AS8267 AS8323 AS12990 AS13255 AS15967
      AS16290 AS16138 AS15541
export: to AS6778 announce AS8267 AS8323 AS12990 AS13255 AS15967
      AS16290 AS16138
export: to AS5550 announce AS8267 AS8323 AS12990 AS13255 AS15967
      AS16290 AS16138 AS15541
export: to AS8508 announce AS8267 AS8323 AS12990 AS13255 AS15967
      AS16290 AS16138 AS15541
export: to AS12631 announce AS8267 AS8323 AS12990 AS13255 AS15967
      AS16290 AS16138 AS15541
```

8501 – PIONIER

6778 – Exatel SA

5550 – TASK, gdański MAN

8508 – SILWEB, Śląsk

12631 – Formus Polska

12990 – Onet.pl

15541 – Ceti

Niektóre polskie AS

- **AS8938** ENERGIS-IP Energis Polska IP Network
- **AS12827** WIRTUALNAPOLSKA Wirtualna Polska S.A.
- **AS20553** HPPOLAND-AS Hewlett Packard Polska (HPO)
- **AS20778** PKOBP Powszechna Kasa Oszczędności Bank Polski S.A.
- **AS25439** PWPW-AS Polska Wytwórnia Papierów Wartościowych S.A
- **AS25506** TVP-AS Telewizja Polska S.A.
- **AS30759** ALCATEL-POLSKA-AS Alcatel Polska S.A
- **AS33900** TPSA-MPLSNET-AS Telekomunikacja Polska MPLSNET

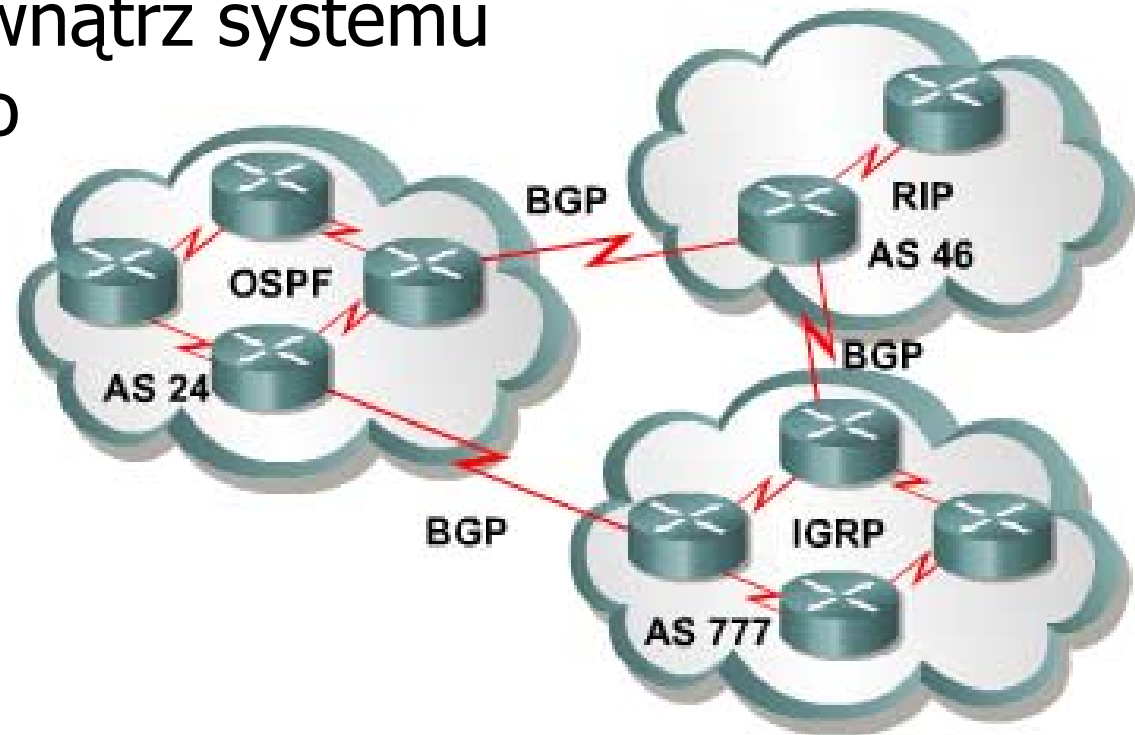
2. Informacje ogólne o BGP

Informacje ogólne

- Protokół routingu zewnętrznego: routing pomiędzy systemami autonomicznymi
- Budowany w oparciu o doświadczenie zdobyte na protokole EGP
- *de facto* standard w komunikacji pomiędzy routerami w Internecie
- tzw. **policy routing**; wymaga dużej ingerencji (i wiedzy) administratora

Obszar zastosowania BGP

- Przede wszystkim pomiędzy systemami autonomicznymi (eBGP)
- Czasem też wewnątrz systemu autonomicznego (iBGP)
 - najczęściej w sieci ISP



Grafika: Cisco Systems

Rodzina protokołów BGP

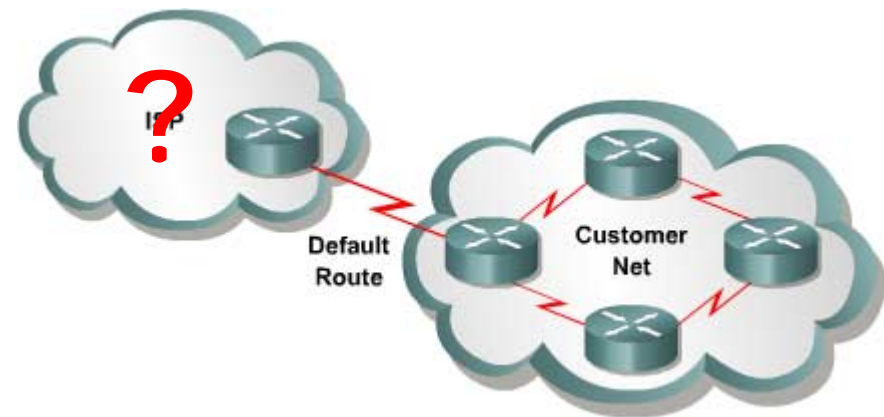
- BGP-1, RFC 1105, 1989
 - zastąpienie EGP
- BGP-2, RFC 1163, 1990
- BGP-3, RFC 1267, 1991
- BGP-4, RFC 1771, 1995
 - Routing bezklasowy

Działanie routera IP

- Informacje o dostępności sieci zgromadzone w tablicy routingu
- Sekwencja wpisów postaci:
 - Adres IP
 - Maska adresu
 - Adres routera następnego skoku/nazwa interfejsu
- Algorytm wyboru: dopasowanie najdłuższej maski; im bardziej specyficzna trasa, tym lepiej
- Wpis domyślny
 - Zmniejszenie rozmiaru tablicy
 - Delegowanie odpowiedzialności za przekazanie pakietu **dalej**

Routing pomiędzy ISP

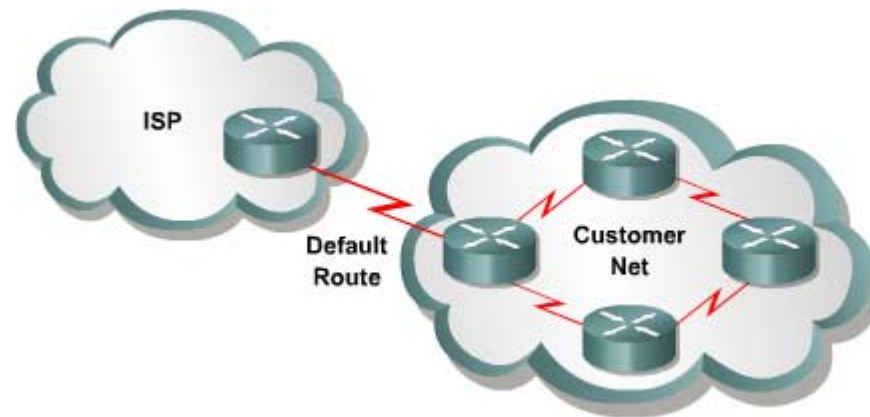
- Prawie zawsze jest wymagana komunikacja obustronna: sieć musi wiedzieć jak przesyłać do nas dane
- Ważna jest symetria
- Ważna jest swoboda wyboru operatora



Grafika: Cisco Systems

Rodzaje systemów autonomicznych (1)

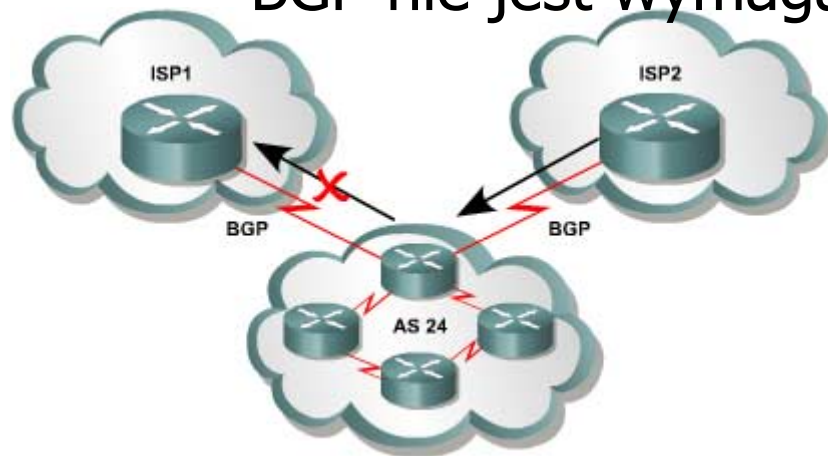
- *Single-homed, stub network*
 - Jedna trasa domyślna na zewnątrz
 - Możliwa konfiguracja
 - Trasa domyślna
 - Routing dynamiczny IGP
 - Routing dynamiczny BGP – po co?



Grafika: Cisco Systems

Rodzaje systemów autonomicznych (2)

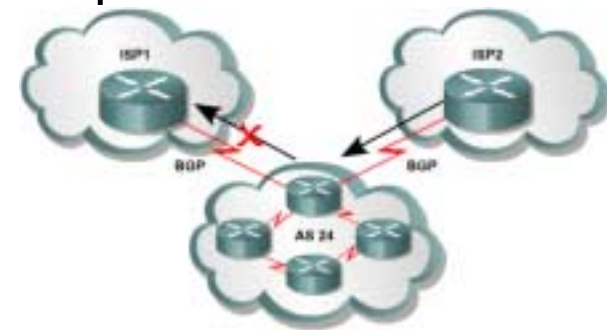
- *Multihomed nontransit*
 - Więcej niż jedno wyjście na świat
 - Dołączone do jednego lub więcej dostawców
 - Nie zezwala się na ruch tranzytowy
 - Nie ogłasza się sieci otrzymanych na drugim wyjściu
 - BGP nie jest wymagany, choć jest polecany



Grafika: Cisco Systems

Rodzaje systemów autonomicznych (2)

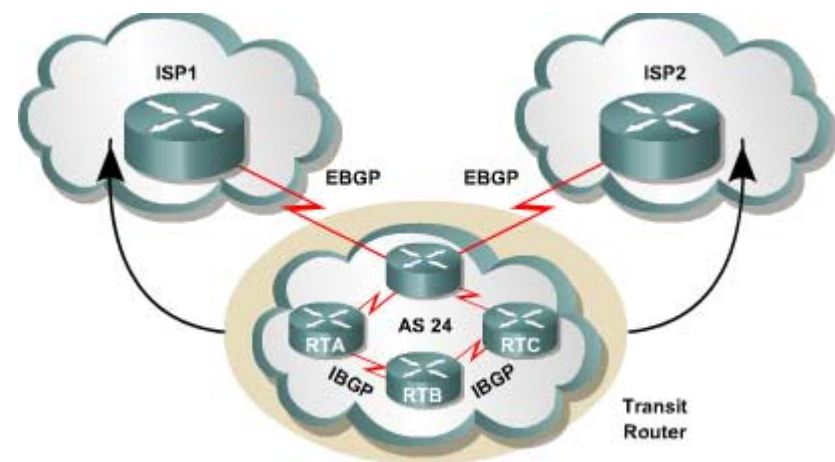
- *Multihomed nontransit*
 - Dołączone do jednego dostawcy
 - Ważna symetria
 - Równoważenie obciążenia lub łącze zapasowe
 - Duże wsparcie BGP
 - Kontrola ruchu wychodzącego
 - Kontrola ruchu przychodzącego
 - Dołączone do wielu dostawców
 - Dywersyfikacja ruchu ☺
 - Wymagania jak wyżej
 - Problem przestrzeni adresowej



Grafika: Cisco Systems

Rodzaje systemów autonomicznych (3)

- *Multihomed transit*
 - Więcej niż jedno wyjście na świat
 - Dołączone do jednego lub więcej dostawców
 - Zezwala się na ruch tranzytowy
 - BGP jest uruchomiony
 - na styku sieci (EBGP)
 - wewnątrz sieci (IBGP)



Grafika: Cisco Systems

Podstawowa terminologia BGP

- Sąsiad, *peer*
 - W tym samym AS:
 - *internal peer*
 - BGP: *internal BGP*, iBGP
 - W innym AS:
 - *external peer*
 - BGP: *external BGP*, eBGP
 - Niekoniecznie bezpośrednio przyległy router (!)
- **Prefiks**
 - Adres z maską, np. 149.156.97.0/24
 - „BGP nie ogłasza mi żadnych prefiksów!”

Komunikaty BGP

- **open**: używany przy nawiązywaniu połączenia BGP
- **keepalive**: podtrzymuje komunikację między sąsiadami
- **notification**: informuje o błędach
- **update**: wysyłany w razie wystąpienia zmian

Komunikat OPEN

- Zestawiane jest połączenie warstwy transportowej
 - Połączenie punkt-punkt (!)
 - TCP/179
 - Użycie niezawodnego protokołu eliminuje
 - konieczność implementacji retransmisji
 - konieczność numerowania pakietów
 - konieczność generowania potwierdzeń, itp.
- Przesyłane są m.in. numer AS, numer wersji BGP
- Ustalany jest czas pomiędzy komunikatami KEEPALIVE

Komunikat KEEPALIVE

- Podtrzymuje aktywność na łączy – monitorowanie istnienia komunikacji
- W stabilnie działającej sieci to jedyne komunikaty przesyłane pomiędzy sąsiadami
- Sugerowane ustawienia
 - *Keepalive interval*: 30 sek.
 - *Hold time interval*: 90 sek.
- Wielkości domyślne na routerach Cisco
 - *Keepalive interval*: 60 sek.
 - *Hold time interval*: 180 sek.

Komunikat NOTIFICATION

- Informuje o błędach
- Wysyłany w razie wystąpienia dowolnego błędu (też po przekroczeniu czasu *hold time*)
- Powoduje zamknięcie połączenia BGP

Komunikat UPDATE

- Pozwala na przesłanie informacji o trasach pomiędzy sąsiadami BGP
- Wysyłany po zestawieniu połączenia i w razie zaistnienia zmian
- Przesyła informacje
 - Identyfikującą opisywaną trasę: prefiks
 - Identyfikujące usuwane (*withdrawn*) trasy
 - Opisujące ścieżkę: jej atrybuty

3. Atrybuty ścieżki BGP

4. Proces decyzyjny BGP

Atrybuty ścieżki

- W protokołach IGP: metryka i jej składowe
- BGP: **atrybuty**
 - Charakteryzują ogłaszaną ścieżkę
 - Sekwencja wiadomości postaci $\langle \text{typ}, \text{długość}, \text{wartość} \rangle$ (TLV)
 - Cztery rozłączne klasy

Klasy atrybutów ścieżki

- *well-known*: obsługa jest wymagana przez każdą implementację BGP
 - *mandatory*: musi być zawarty w każdej wiadomości *Update*
 - *discretionary*: nie musi być zawarty w każdej wiadomości *Update*
- *optional*: obsługa nie jest wymagana przez implementację BGP
 - *transitive*: komunikat *Update* powinien być przekazany dalej, nawet gdy parametr nie jest zrozumiany
 - *nontransitive*: komunikat *Update* nie jest przekazywany dalej gdy parametr nie jest zrozumiany

Atrybut ORIGIN

- Klasa *well-known mandatory*
- Opisuje pochodzenie wiadomości
 - IGP: z protokołu wewnętrznego
 - EGP: z protokołu zewnętrznego
 - *Incomplete*, nieokreślone lub nieznanne, np. z redystrybucji
- Kolejność uwzględniania w metryce: IGP, EGP, *incomplete*
- Atrybut ustawiany przez AS, który wygenerował informację o sieci

Atrybut AS_PATH

- Klasa *well-known mandatory*
- Zawiera sekwencję numerów AS, przez które wiedzie trasa
- Proces dodawania kolejnych numerów (na początek listy): *AS prepending*
- Routery iBGP nie zwiększają tej listy
- Możliwe kilkukrotne dodanie tego samego numeru AS
- Pozwala na wykrycie pętli

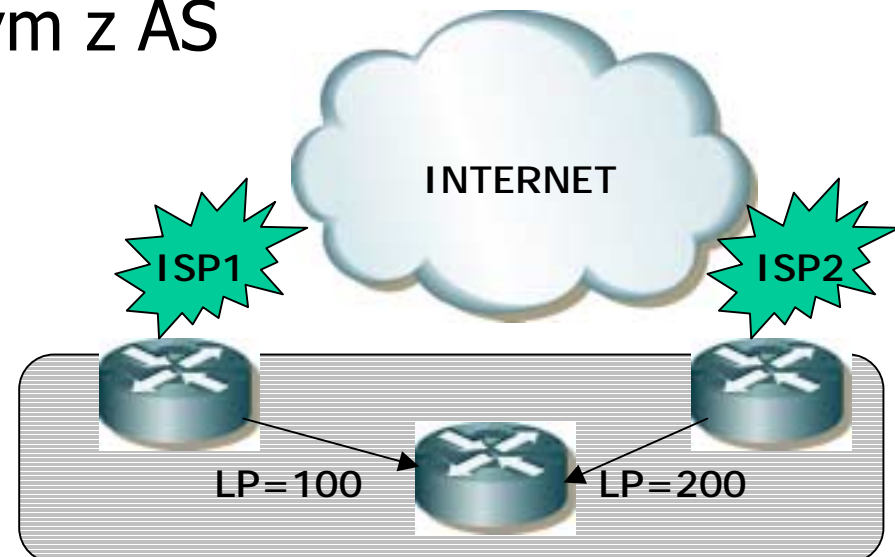
Atrybut NEXT_HOP

- Klasa *well-known mandatory*
- Adres routera następnego skoku
- Nie zawsze adres sąsiedniego routera
 - Gdy routery otrzymujący i wysyłający są w różnych AS, adres ogłaszającego routera
 - Gdy routery otrzymujący i wysyłający są w tym samym AS:
 - prefiks jest z tego samego AS: adres routera ogłaszającego
 - prefiks jest spoza AS: adres routera z poprzedniego AS
 - Niekiedy wymagany jest tzw. *recursive lookup*



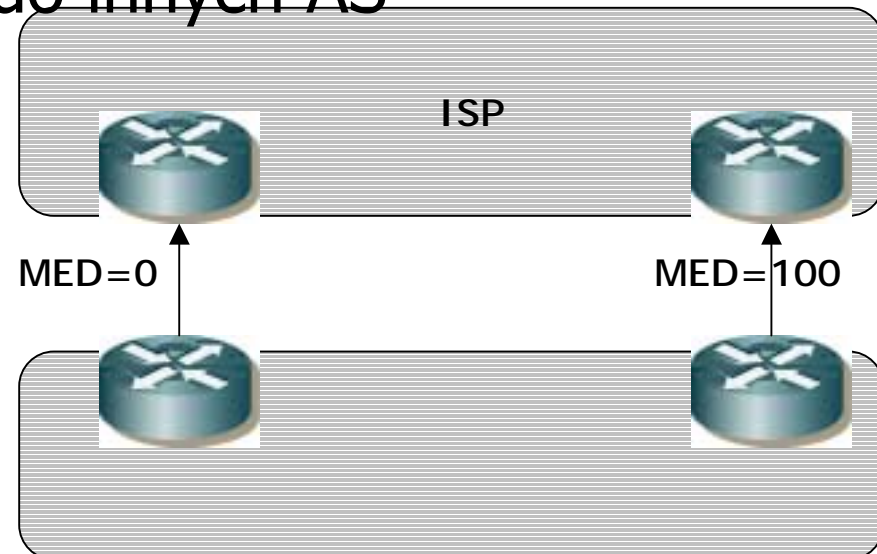
Atrybut LOCAL_PREF

- Klasa *well-known discretionary*
- Ma sens jedynie w komunikacji pomiędzy routerami iBGP, nie jest przekazywany do innych AS
- Używany do preferowania któregoś z routerów przy ruchu wychodzącym z AS
- Wyższe wartości są preferowane
- 100 jest wartością domyślną



Atrybut MULTI_DISC_EXIT

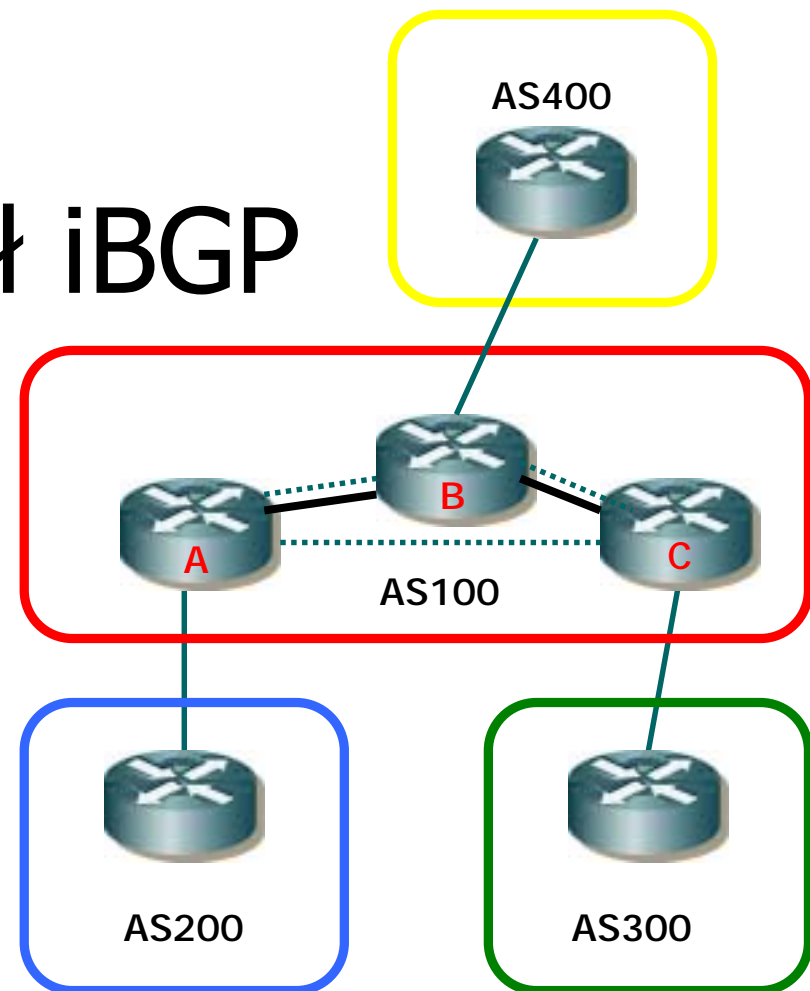
- Klasa *optional nontransitive*
- Używany do poinformowania sąsiedniego AS o preferowanym punkcie **wejścia** do naszej sieci
- Ma sens jedynie w komunikacji do jednego AS, nie jest przekazywany do innych AS
- Niższe wartości są preferowane



Proces decyzyjny BGP (skrót)

1. Najwyższa wartość LOCAL_PREF
2. Preferuj trasy uzyskane lokalnie (na tym routerze)
3. Trasy o najmniejszej długości listy AS_PATH
4. Trasy o najmniejszej wartości *origin code*
IGP < EGP < *incomplete*
5. Najmniejsza wartość MULTI_EXIT_DISC
6. Najmniejsza wartość metryki do routera
NEXT_HOP

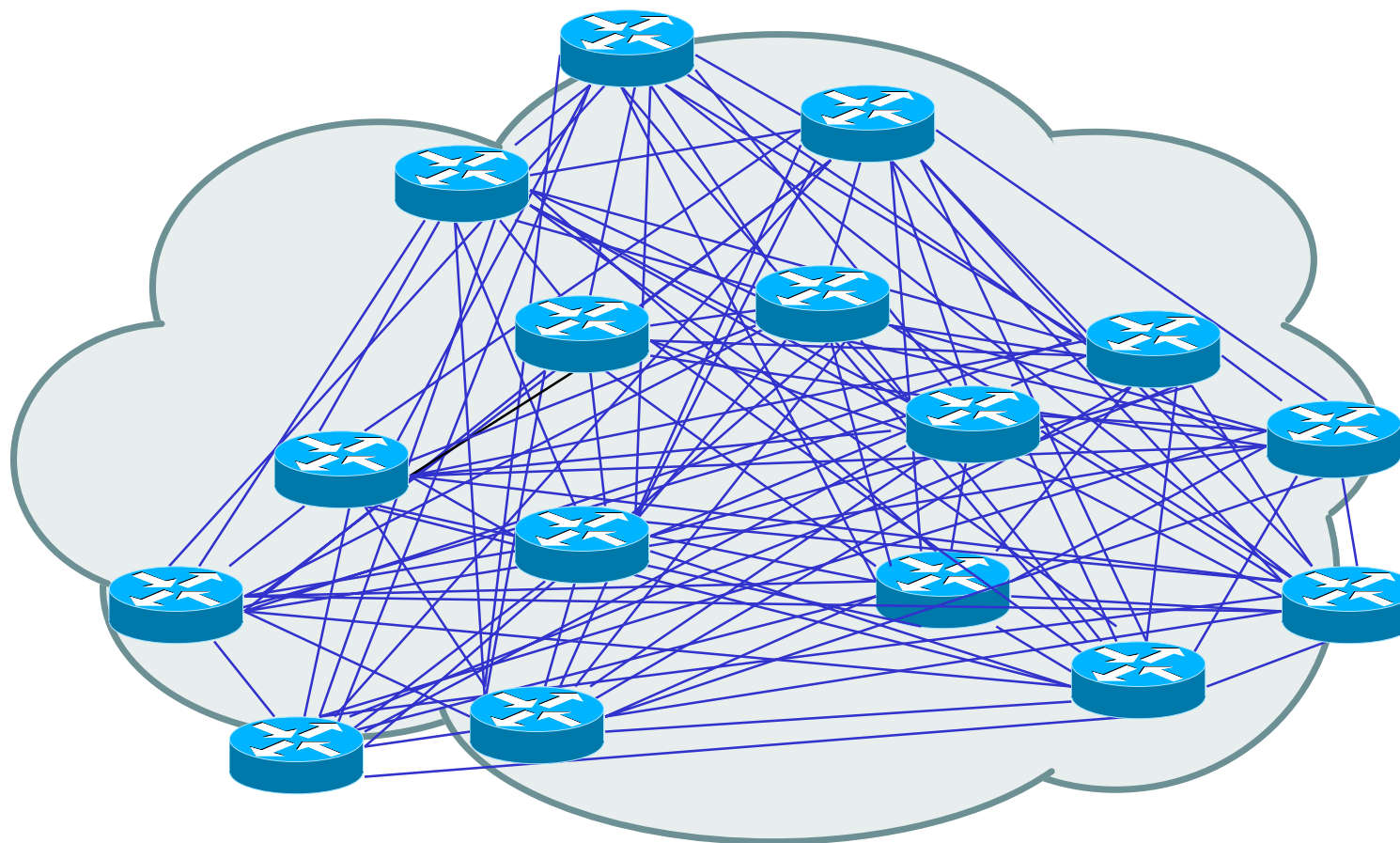
5. Protokół iBGP



Protokół iBGP

- iBGP używany praktycznie tylko w sieciach *multihomed*: właściwe rozgłaszanie informacji w przypadku wielu dostawców
- Nie jest zmieniana wartość AS_PATH (nie ma ochrony przez pętlami)
- iBGP nie ogłasza tras otrzymanych od innych routerów iBGP (pętle!)
- Połączenie routerów musi być typu każdy z każdym (*fully meshed*)
 - Wada, nie używać jako jedyne go protokołu IGP

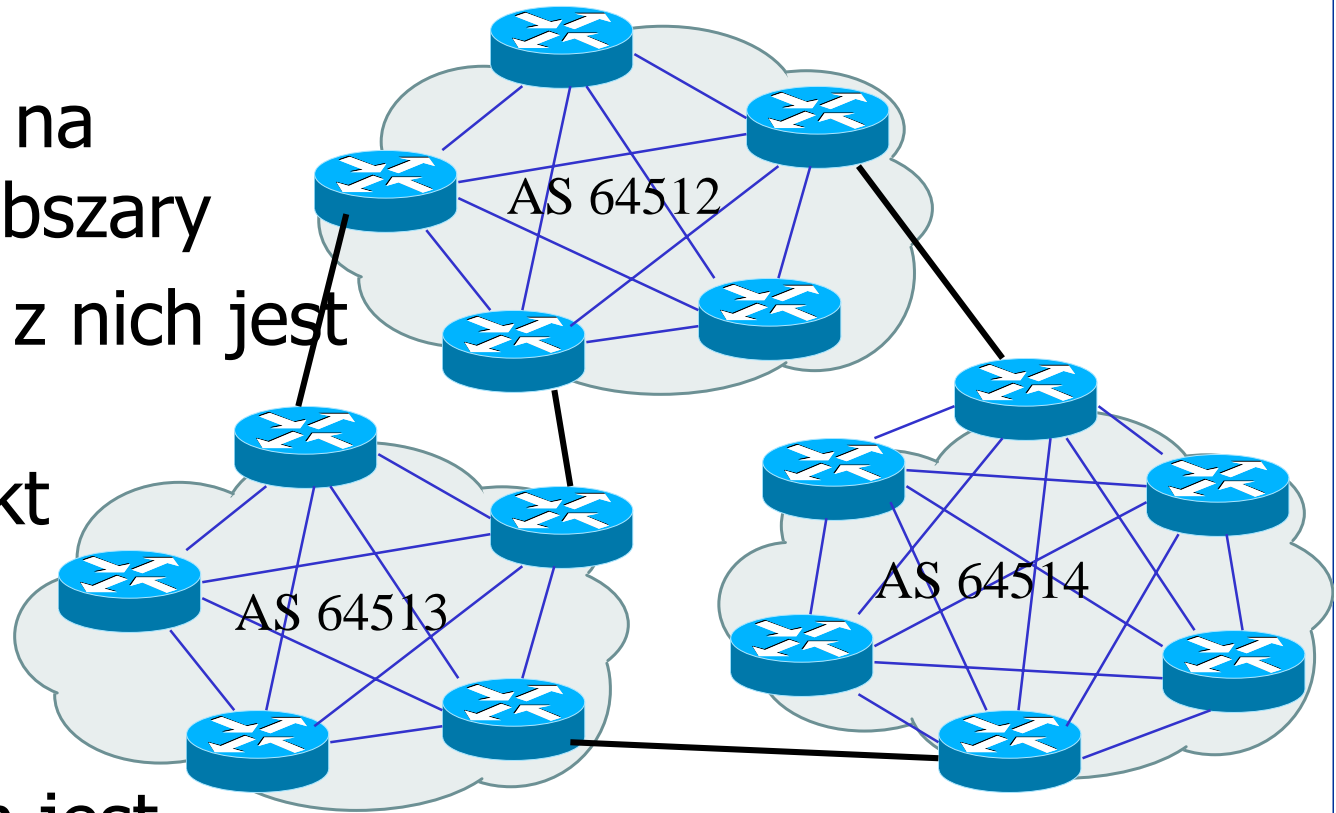
iBGP w dużym AS



Grafika: Intro to BGP: All Day Tutorial, Avi Freedman

Konfederacje

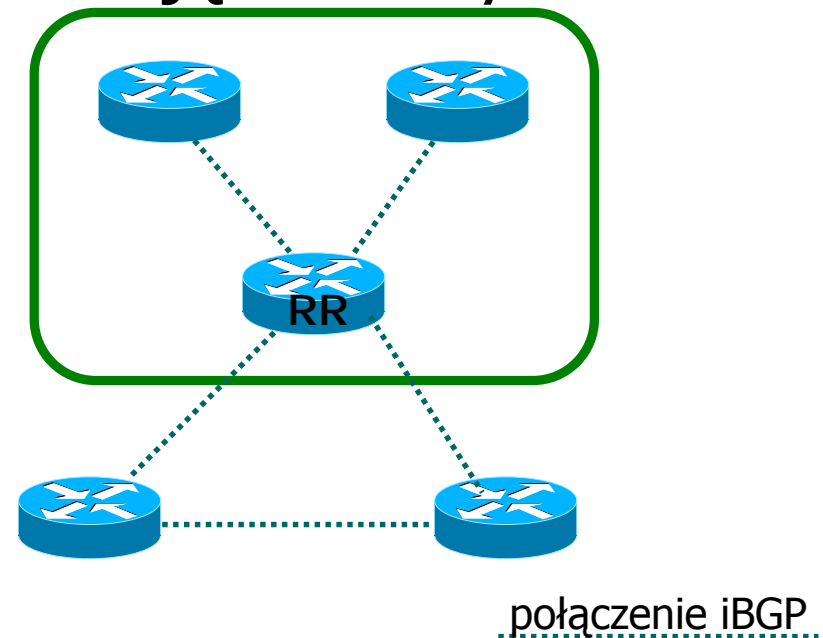
- Podział AS na mniejsze obszary
- W każdym z nich jest połączenie punkt-punkt
- Mechanizm jest przejrzysty dla sieci zewnętrznych



Grafika: Intro to BGP: All Day Tutorial, Avi Freedman

Route reflector

- Podział routerów iBGP w ramach AS na klientów i nie-klientów
- Nie-klient rozgłasza informację do wszystkich swoich klientów



6. Stabilność tras

Stabilność tras

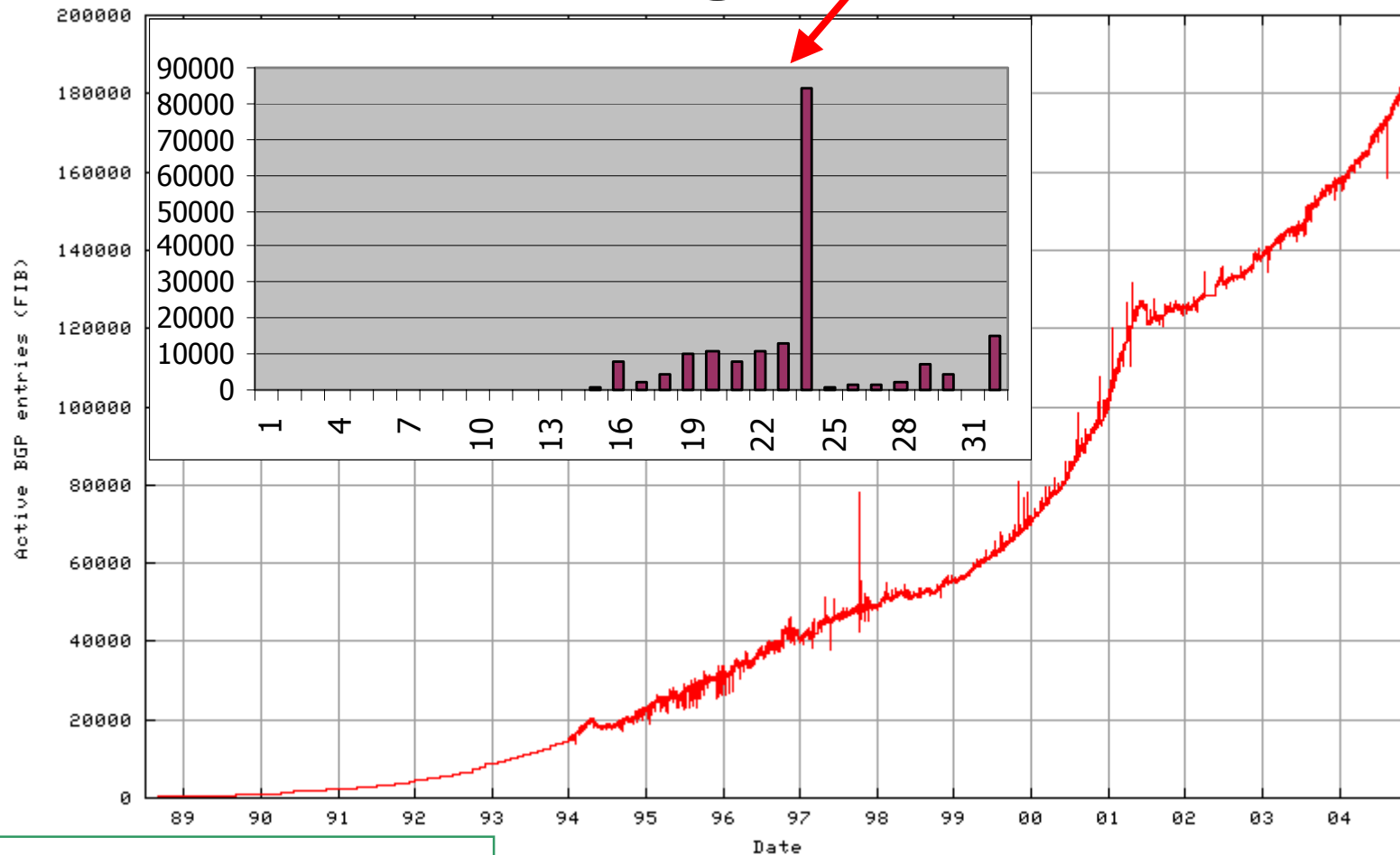
- Gdy sieć się pojawia: informacja jest propagowana
- Gdy sieć znika: informacja jest propagowana
- Często migotające sieci (*flapping*) powodują nadmierny ruch
- Mechanizm *route dampening* pozwala identyfikować takie sieci

Route dampening

- (Niestabilna) sieć otrzymuje karę (*penalty*) za każdym razem gdy pojawia się i znika
- Jeśli kary przekroczą założony limit – informacja o sieci nie jest rozgłaszana
- Wielkość kary jest automatycznie obniżana o połowę co ustalony czas
- Po obniżeniu się poniżej ustalonego progu informacja o sieci znów może być ogłaszana

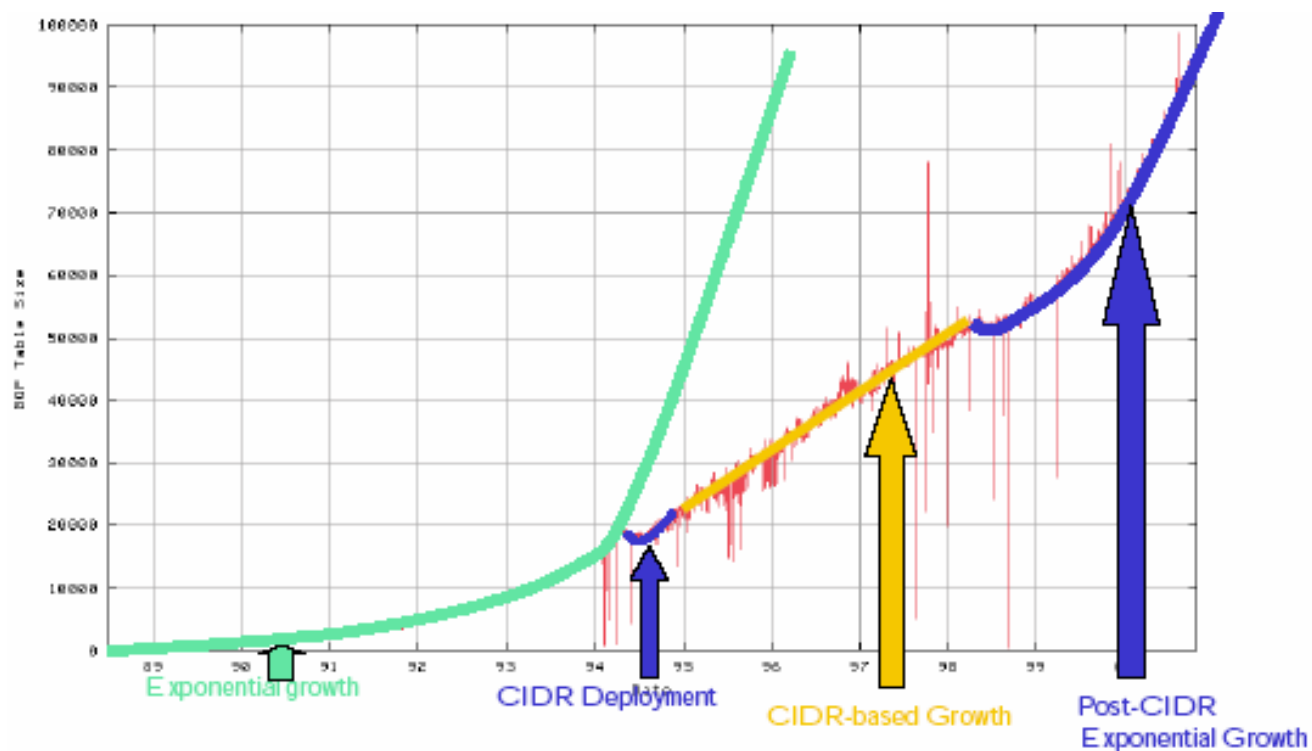
7. Podsumowanie

Telstra – AS1221: ilość wpisów BGP



Grafika: <http://bgp.potaroo.net>

Telstra – AS1221: CIDR a wielkość tablic BGP



Grafika: <http://bgp.potaroo.net>

Przykładowe komendy BGP

```
router bgp 60000
  network 1.1.1.0 mask 255.255.255.0
  neighbor 111.111.111.111 remote-as 60001
  neighbor 111.111.111.111 update-source loopback 0
  neighbor 111.111.111.111 next-hop-self
  neighbor 111.111.111.111 ebgp-multihop 2
  neighbor 1.1.1.1 filter-list 1 in
  neighbor 1.1.1.1 distribute-list 1 out
  neighbor 1.1.1.1 weight 777
  neighbor 1.1.1.1 route-map USTAWWAGE in
  neighbor 1.1.1.1 route-reflector-client
  bgp default local-preference 777
```


Literatura

- RFC 1771
 - A Border Gateway Protocol 4 (BGP-4)
- RFC 3065
 - Autonomous System Confederations for BGP
- RFC 2796
 - BGP Route Reflection - An Alternative to Full Mesh IBGP
- RFC 2439
 - BGP Route Flap Damping

KONIEC