Маршрутен протокол BGP

Понятие за автономна система (AS)

Автномната система (**autonomous system - AS**) е сбор от свързани помежду си IP мрежи (префикси), които са под административното и техническо управление на мрежов оператор или др. организация (напр. СУ).

В рамките на AS е възможно да работят различни вътрешни протоколи за маршрутизация (IGP).

AS поддържат строго дефинирана политика за маршрутизация в Internet (вж. RFC 1930).

Понятие за AS

AS трябва да има глобален уникален номер (ASN - Autonomous System Number)

Този номер се използва при обмен на маршрутизираща информация със съседни AS-и и като идентификатор на самата AS.

Кога ни трябва AS (ASN)

AS са задължителни при обмен на външни маршрути с други ASs с помощта на протоколи за външна маршрутизация.

В момента такъв е BGP (Border Gateway Protocol).

Но това не е достатъчно условие, за да искаме да имаме AS.

Кога да, кога не - AS

AS ни е необходима единствено и само тогава, когато имаме политика за маршрутизация (routing policy), различна от тази на други партньори - съседи (peers).

routing policy – как останалата част от Internet взима решения за маршрутизация на базата на информация от нашата AS.

Кога да, кога не - AS

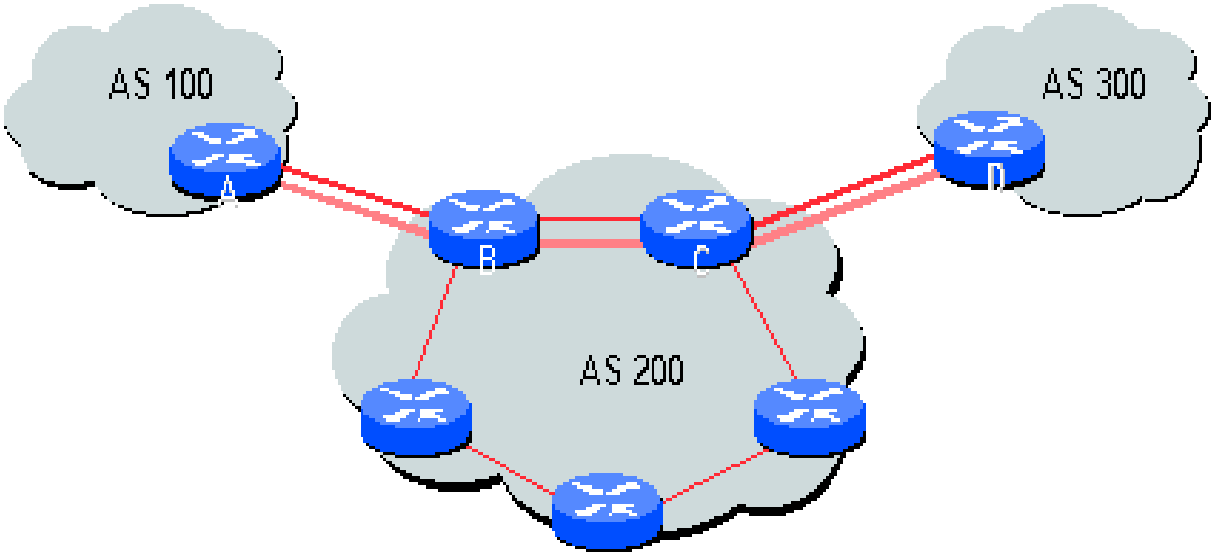
Single-homed site, единствен или множество префикси, свързан към един единствен доставчик (т.е. една AS).

Не ни е необходима AS. Префиксът/те се поставя в AS на провайдера.

Multi-homed site. Необходима е AS.

multi-homed означава префикс или група от префикси, която се свързва към повече от един доставчик (т.е. повече от една AS, всяка със своя политика).

AS 5421 (CY) - Multi-homed



ASNs

До 2007 г. ASNs бяха единствено 16-битови, максимален брой 65536.

IANA е резервирала следния блок от AS номера за частно ползване (да не се анонсират в глобалния Internet):

64512 - 65535

ASN 0 означава немаршрутизирана мрежа.

Раздаване на ASNs

Всички останали номера (1–64495) са раздадени от IANA.

IANA ги разпределя на съответните RIR, които от своя страна присвояват AS номера на организации в тяхната област, които отговарят на дадените по-горе критерии.

(RIR за територията на България е RIPE, http://ripe.net).

Разпределението на ASN ресурса от страна на IANA до момента, можете да видите на :

http://iana.org/assignments/as-numbers/as-num

32-битови AS номера

Поради големия брой PI оператори адресното пространство на 16-битовите автономни системи застрашително се запълва.

Към 01.04.2011 са раздадени номера до 58367 включително.

58368-64495 са резервирани от IANA.

Затова се въведоха 32-битови AS номера, RFC 4893, които вече се раздават от IANA.

16-битови AS се явяват подмножество на 32-битови AS (с 16 нули отляво).

Гарантира се плавен преход за разлика от IPv4 към IPv6. Все пак зависи от софтуера.

Border Gateway Protocol (BGP)

Border Gateway Protocol (BGP) е основният протокол за маршрутизация в Internet.

Поддържа таблица от IP мрежи (префикси), които определят достижимостта на мрежите между автономните системи.

BGP е протокол с вектор на пътищата, path vector protocol.

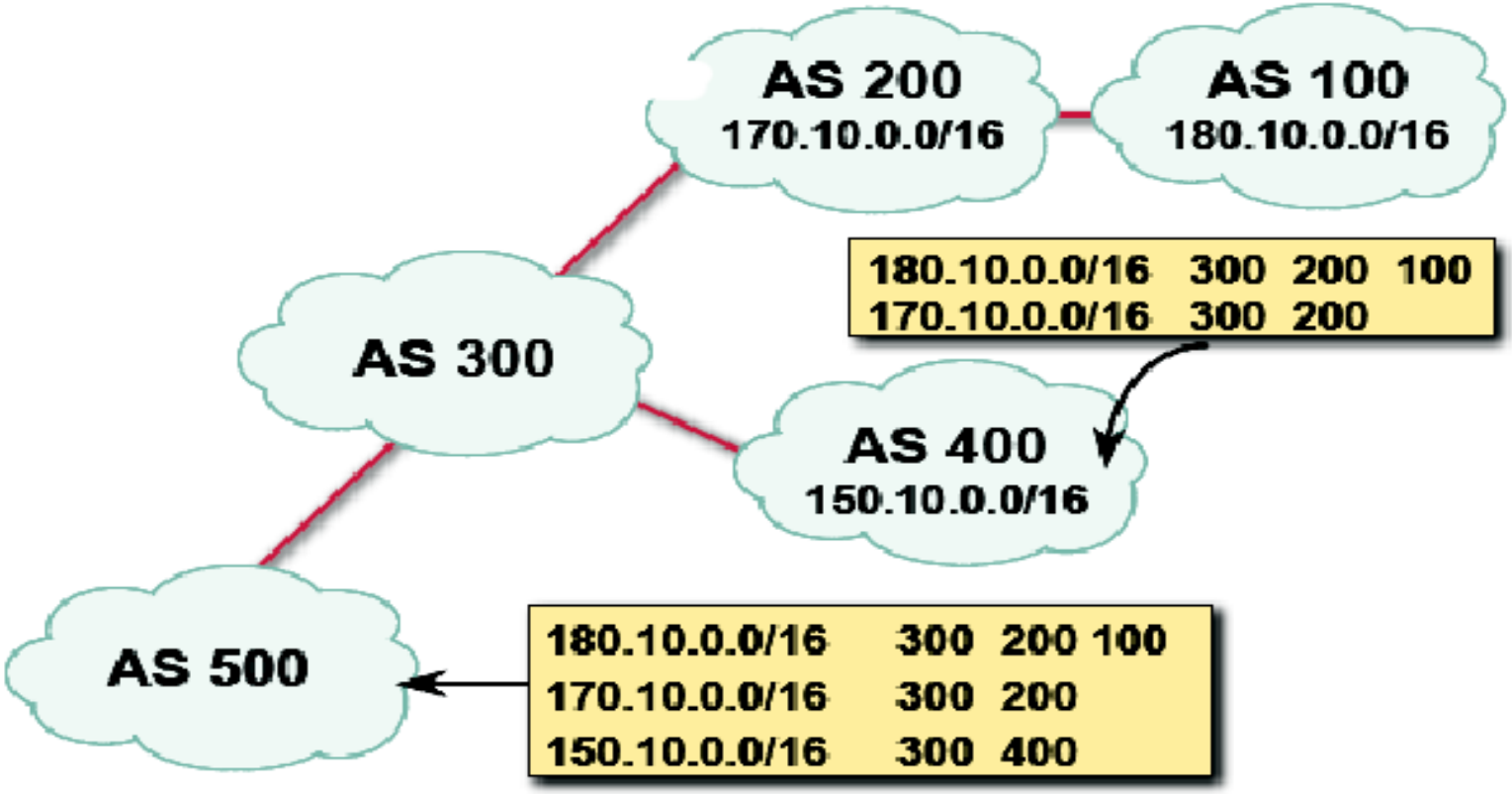
Вектор на пътищата

BGP не използва метриката на вътрешните протоколи, а взема решения за определяне на маршрути на база на пътя между ASs, мрежови политики и/или правила.

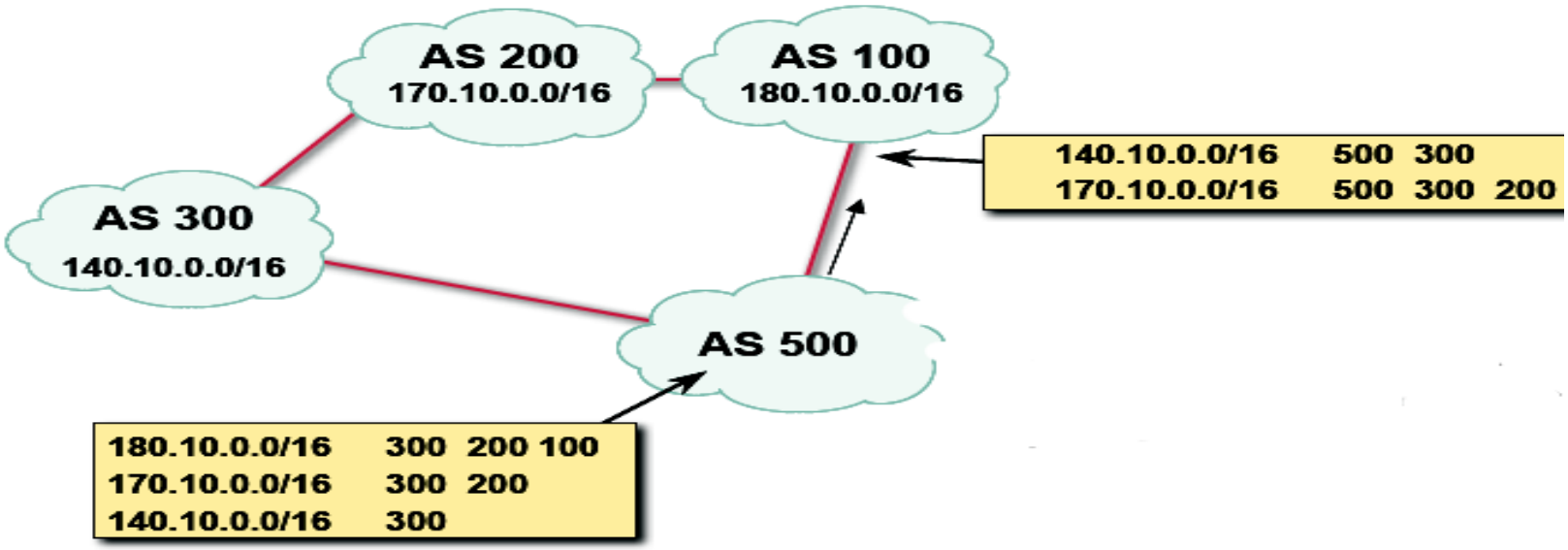
От 1994 г. насам се използва версия 4 на протокола, която поддържа CIDR и обединяване (агрегация) на маршрути, с което се намалява размера на маршрутните таблици.

От януари 2006 г. версия 4 е стандартизирана в RFC 4271. BGP4/4+ в момента.

Вектор на пътищата



Защита от зацикляне



180.10.0.0/16 не се приема от AS100.

Префиксът има AS100 в своя AS-PATH.
Разпознат е цикъл (loop).

Принцип на действие

BGP съседите (neighbors или peers) - маршрутизатори, се формират, след като ръчно са зададени.

Между тях се установава TCP сесия по порт 179.

Всеки BGP възел периодично изпраща 19-байтови “keep-alive” съобщения за поддържане на връзката.

BGP единствен от маршрутизиращите протоколи използва TCP за танспорт, което го прави приложен протокол до известна степен.

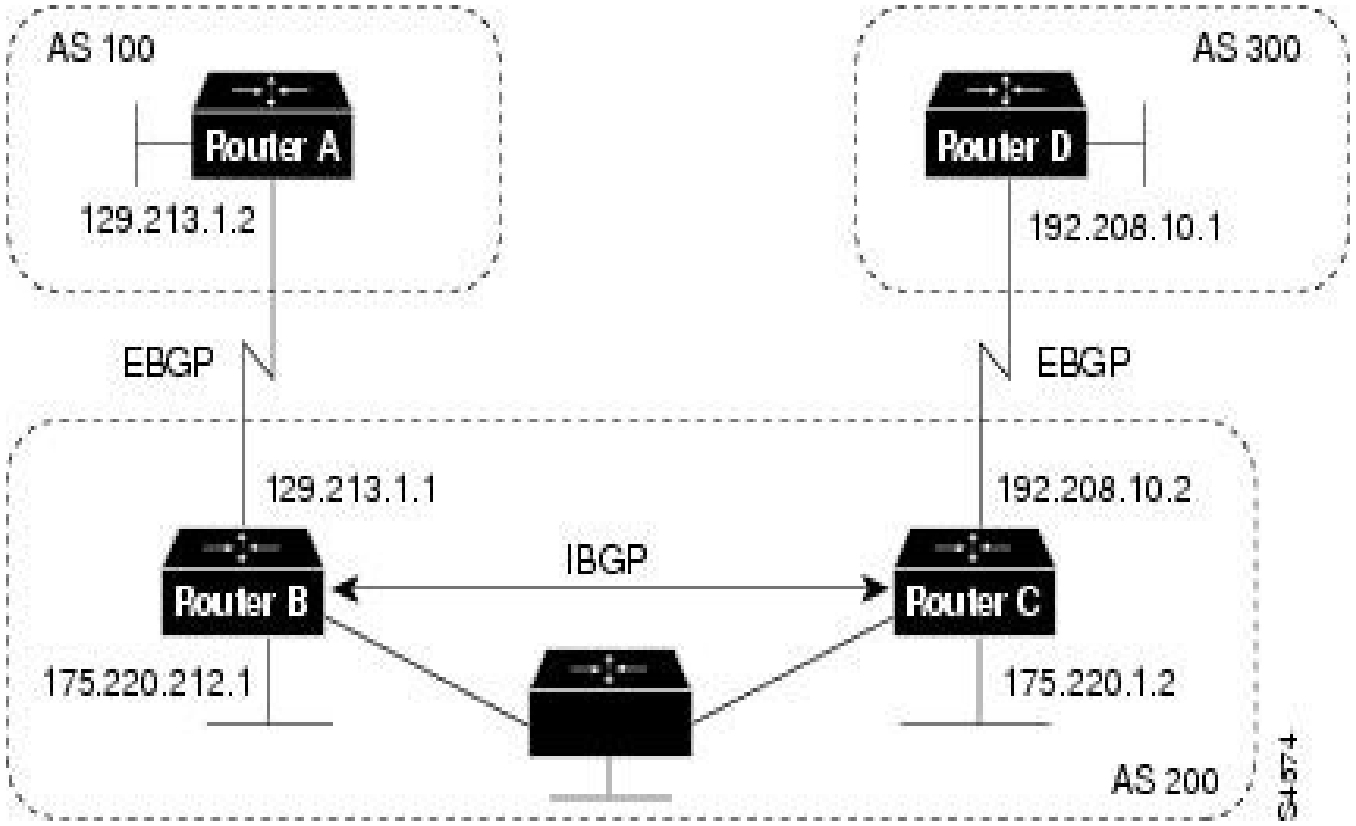
iBGP и eBGP

Когато BGP работи в рамките на AS, третира се като вътрешен (iBGP *Interior Border Gateway Protocol*).

Когато работи между ASs, нарича се външен (eBGP *Exterior Border Gateway Protocol*).

Маршрутизаторите на границата на дадена AS, които обменят информация с друга AS, се наричат гранични (border или edge).

iBGP и eBGP


AS 100
Router A
129.213.1.2
AS 300
Router D
192.208.10.1
EBGP
EBGP
129.213.1.1
192.208.10.2
IBGP
Router B
Router C
175.220.212.1
175.220.1.2
AS 200
3454

iBGP и eBGP. Конфигурации.

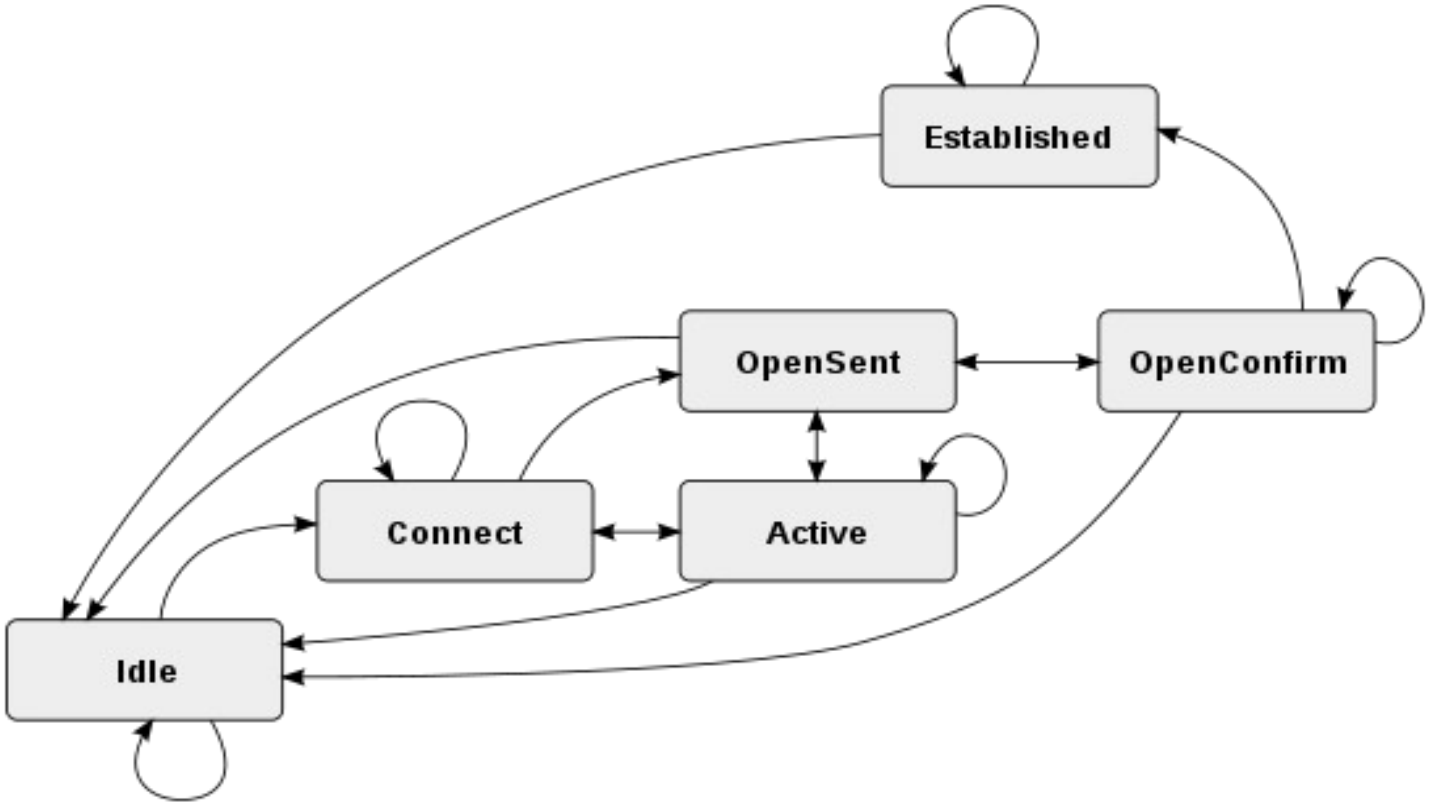
Router B:

```
router bgp 200
neighbor 129.213.1.2 remote-as 100
neighbor 175.220.1.2 remote-as 200
```

Router C:

```
router bgp 200
neighbor 175.220.212.1 remote-as 200
neighbor 192.208.10.1 remote-as 300
```

Принцип на работа на **BGP**.
Схема на състоянията.



Състояния при установяване на BGP сесия

За да установи сесия с партньор (peer), BGP преминава през 6 състояния, описани с краен автомат (finite state machine - FSM).

Това са: Idle, Connect, Active, OpenSent, OpenConfirm и Established.

В BGP е дефинирана променлива на състоянието, която определя в кое от шестте състояния се намира сесията.

При преход от едно състояние в друго се генерират стандартни съобщения.

Idle, Connect и ...

Първоначално BGP маршрутизаторът е в състояние “Idle”. Инициализира всички ресурси, отказва всички входящи опити за установяване на BGP свързаност и инициира TCP сесия със съседа си.

Второто състояние е “Connect”. Маршрутизаторът изчаква да се установи TCP сесия.

Ако е успешно, преминава в "OpenSent".

Ако не, преминава в състояние "Active", докато се нулира таймера ConnectRetry.

Active, OpenSent, Established

В състояние "Active" маршрутизаторътът нулира таймера ConnectRetry, след което се връща в състояние "Connect".

След "OpenSent" маршрутизаторът изпраща съобщение Open и чака за подобно в отговор.

Разменят се съобщения Keepalive и след успех рутерът влиза в състояние “Established”.

Готов е да изпраща и получава от съседа си съобщения Keepalive, Update и Notification.

Обмен на маршрутна информация

BGP съседите си обменят пълната маршрутна информация след установяване на TCP сесия между тях.

Или част от маршрутната таблица, зависи от споразумението между страните, политики, филтри и т.н.

При промени в маршрутната таблица, BGP маршрутизаторите изпращат на съседите си само променените маршрути.

NLRI

Не изпращат периодични обновления (routing updates).

Рекламират (advertise) само оптималния път до дадена дестинация.

В BGP описанието на маршрут до дадена дестинация се нарича Network Layer Reachability Information (NLRI).

NLRI включва префикса на дестинацията и дължината му, пътят през автономните системи и следващия възел, както и допълнителна информация - атрибути.

NLRI

bgpd@border-lozenetz# sh ip bgp

...

```
Network  Next Hop    Metric LocPrf
  Weight    Path
*>1.9.0.0/16 194.141.252.21  0 6802 20965
  3549 4788 i
*              62.44.96.234 50 0 8717 8928
  4788       i
```

!Избран е маршрут ***>**, защото **LocPrf=100** (default), макар AS-PATH да е по-дълъг.

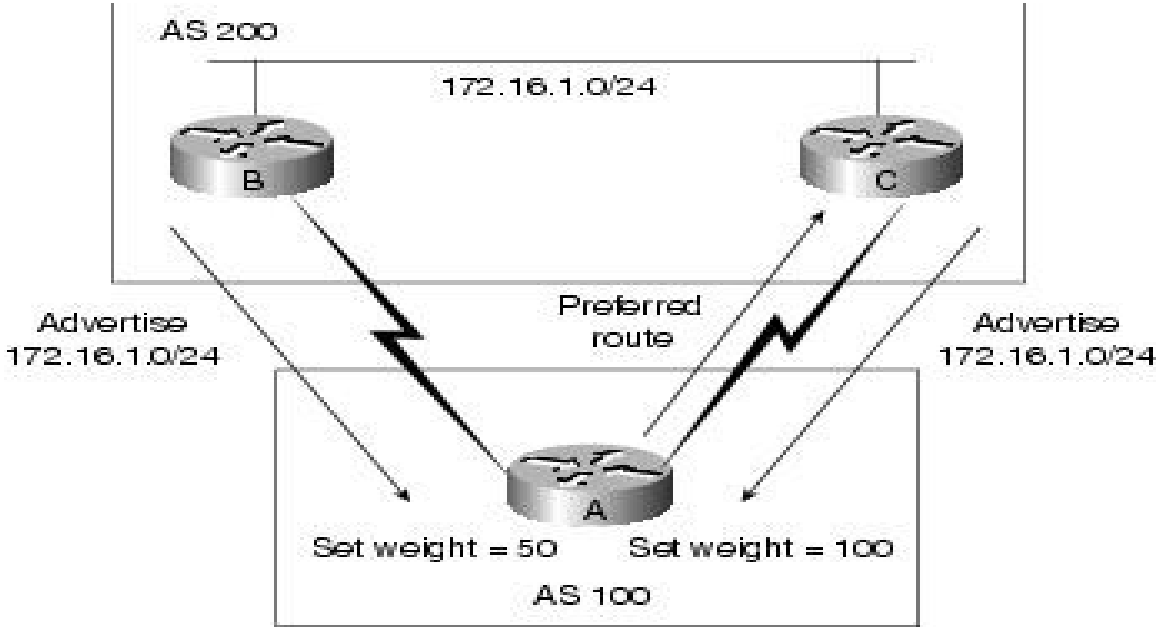
Избор на маршрут

BGP не носи със себе си “политики”, а по-скоро информация, с чиято помощ BGP рутерите вземат “политически” решения, съгласно наложени правила, определени чрез атрибути.

BGP attributes

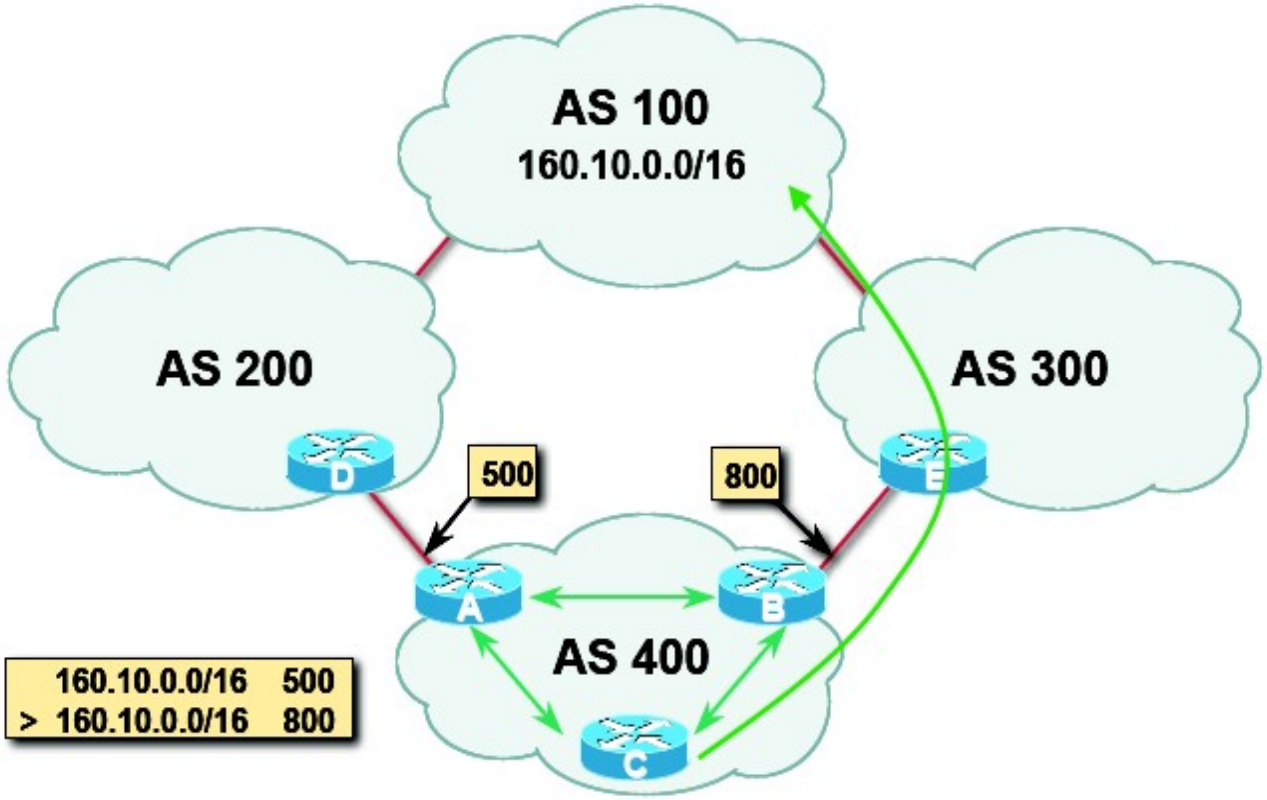
- --Weight
- --Local preference
- --Multi-exit discriminator
- --Origin
- --AS_path
- --Next hop
- --Community

Weight



Weight е специфичен за Cisco и е локален за рутера. Не се рекламира на съседите. Предпочита се маршрут с най-голяма стойност на weight.

Local Preference



Local Preference

Локален за AS – нетранзитивен

local preference = 100, когато е научена от съседна AS

Влияе на избора на път за изходящия трафик

Път с най-висок local preference печели

Local Preference. Конфиг.

Конфигурация на Router B:

```
router bgp 400
neighbor 120.5.1.1 remote-as 300
neighbor 120.5.1.1 route-map local-pref
  in
!
route-map local-pref permit 10
match ip address prefix-list MATCH
set local-preference 800
!
ip prefix-list MATCH permit 160.10.0.0/16
```

Origin

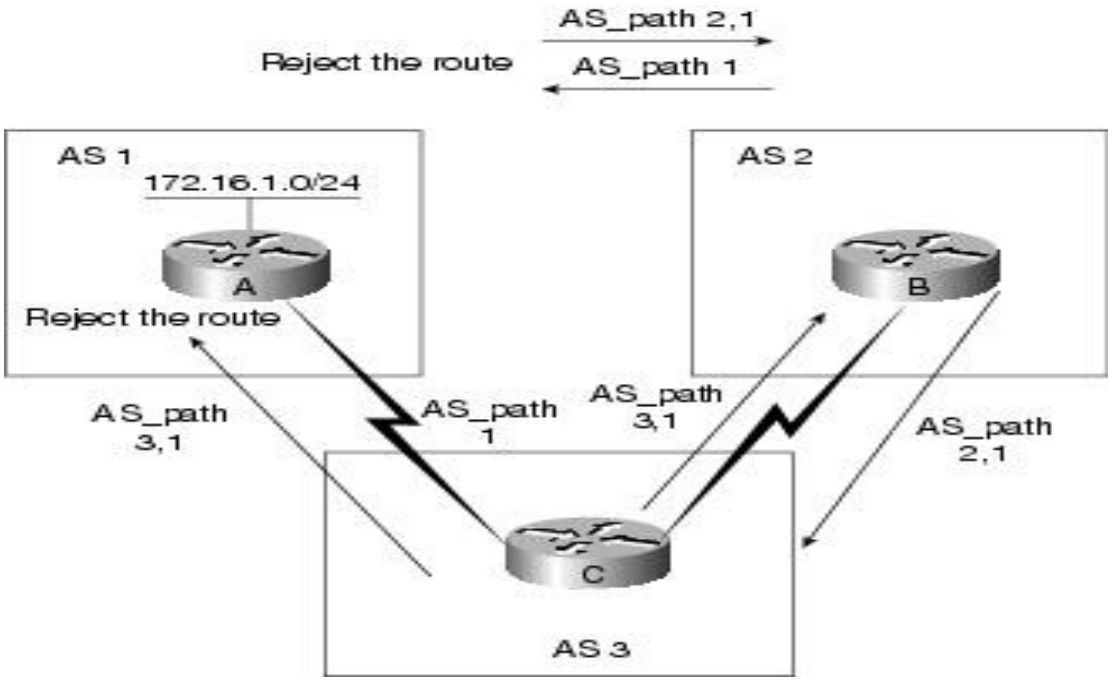
Как BGP научава за конкретен маршрут. Три възможни стойности:

- **IGP**—Маршрутът е вътрешен за AS-източник. Когато е в резултат на `router BGP` командата network.
- **EGP**—Маршрутът е научен чрез eBGP.
- **Incomplete**—Произходът (origin) на маршрута е неизвестен или научен по друг начин. Напр. разпространен (redistributed) в BGP.

Команда network

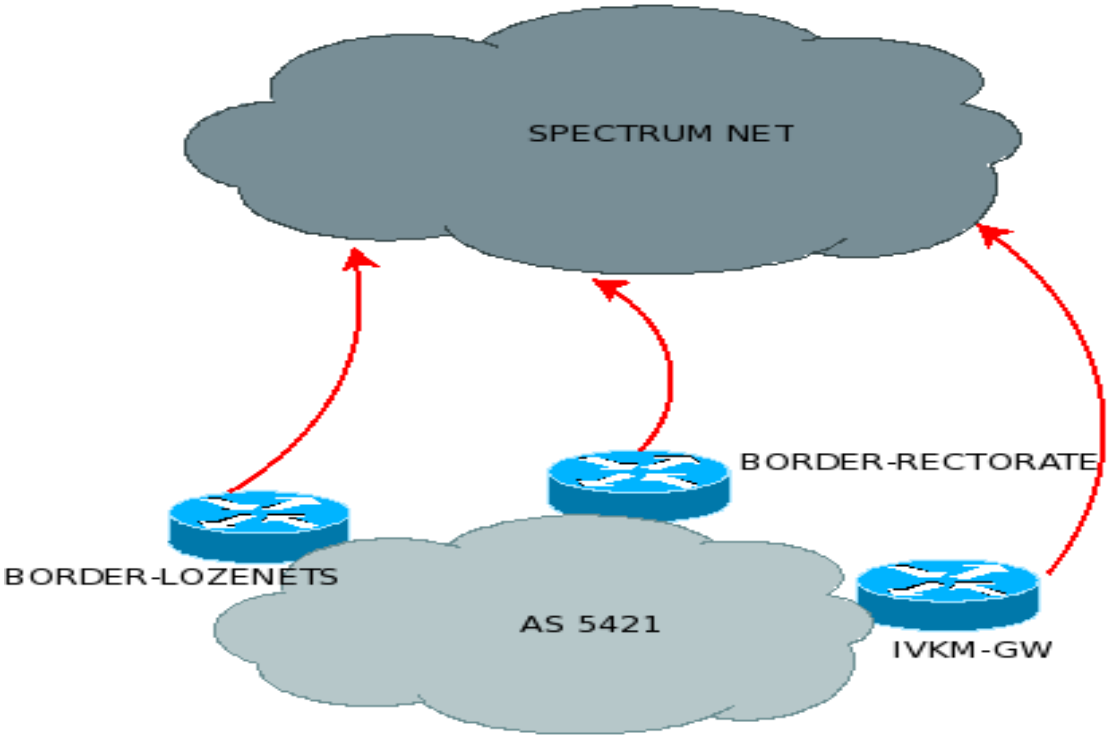
```
bgpd@border-lozenetz# sh run
!
router bgp 5421
 bgp router-id 62.44.127.21
 network 62.44.96.0/19
 network 62.44.96.208/30
 network 62.44.96.232/30
 network 62.44.96.248/30
!...
```

AS_path



Когато реклама на маршрут прминава през авт. система, нейният AS No. се добавя във верижен списък от номера на AS.

AS PATH prepend



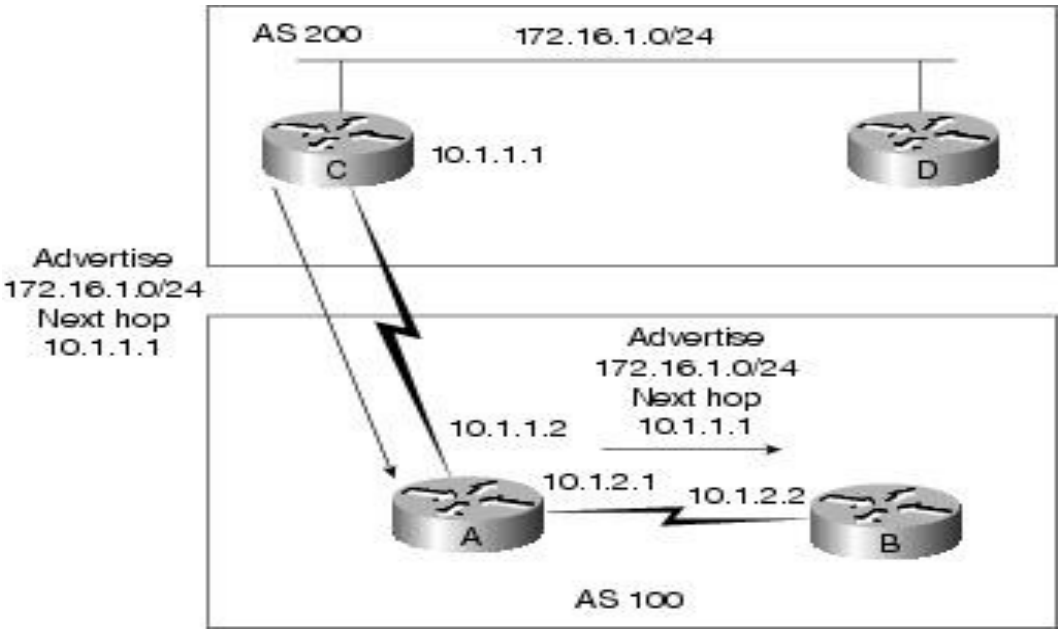
Караме специфичните префикси в Ректорат да излизат през border-rectorate

Препендване на AS

Border-rectorate:

```
route-map BIOM_BORDER_EXPORT_IPV4 permit 10
 match ip address prefix-list
  SU_SUPERBLOCK_IPV4 !62.44.96.0/19
 set as-path prepend 5421
!
route-map BIOM_BORDER_EXPORT_IPV4 permit 20
 match ip address prefix-list
  SPECIFIC_EXPORT_IPV4 !62.44.105.0/24
!62.44.110.0/23 и 62.44.112.0/21
route-map BIOM_BORDER_EXPORT_IPV4 deny 100
```

Next-Hop



IP адресът, чрез който се достига рекламиращият рутер.

За eBGP съседи - IP адреса на връзката между тях.

За iBGP, eBGP next-hop се пренася през локалната AS.

Показване на Origin, Next Hop...

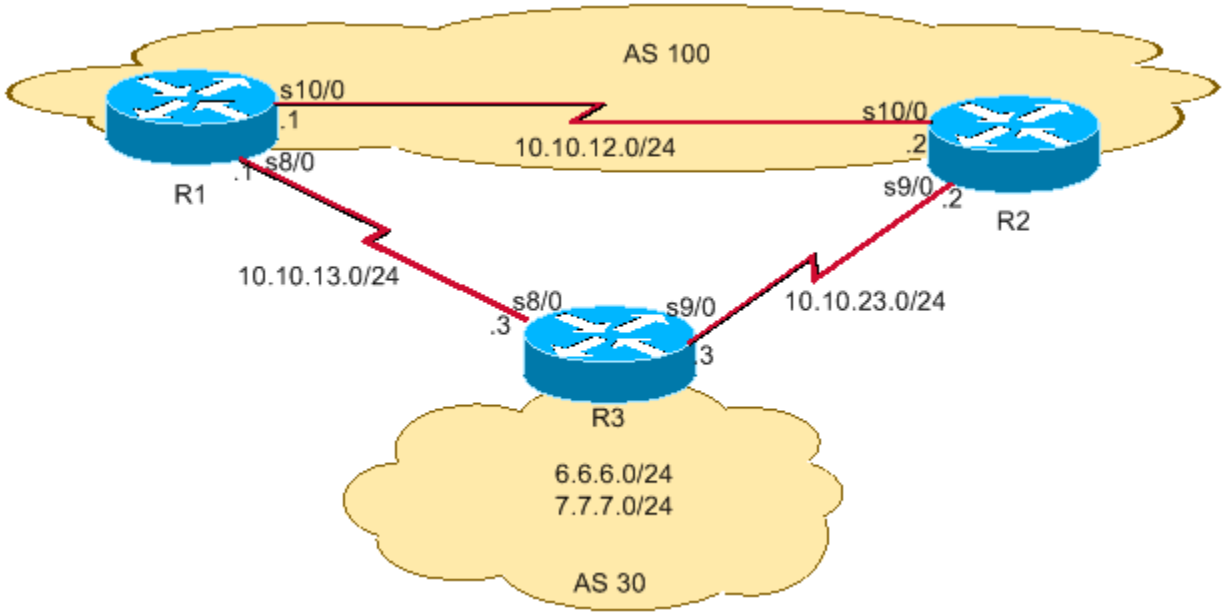
```
bgpd@border-lozenetz# sh ip bgp 2.0.0.0
BGP routing table entry for 2.0.0.0/16
Paths: (1 available, best #1, table
   Default-IP-Routing-Table)
  Advertised to non peer-group peers:
  62.44.127.2 62.44.127.11 62.44.127.15 62.44.127.16
   62.44.127.19 62.44.127.23 62.44.127.43 62.44.127.51
   62.44.127.52 62.44.127.61 62.44.127.70 62.44.127.71
   62.44.127.72 62.44.127.73
  6802 20965 559 30132 12654
    194.141.252.21 from 194.141.252.21 (194.141.252.13)
      Origin IGP, localpref 100, valid, external, best
      Community: 6802:1

      Last update: Sun Dec 13
```

Community

Групиране на дестинации (communities), към които се прилагат решения за маршрутизация.

Рутерите на провайдерите, които трансферират нашия трафик, използват тези communities, за да прилагат конкретни политики за маршрутизация (напр., local preference) към тази група от дестинации.

Community: Пример



Community: Пример - R3

```
access-list 101 permit ip 6.6.6.0/24
access-list 102 permit ip 7.7.7.0/24
…
route-map Peer-R1 permit 10
 match ip address 101
 set community 100:300
!
route-map Peer-R1 permit 20
 match ip address 102
 set community 100:250
```

Community: Пример - R1

```
ip community-list 1 permit 100:300
ip community-list 2 permit 100:250
!
route-map Peer-R3 permit 10
 match community 1
 set local-preference 130
!
route-map Peer-R3 permit 20
 match community 2
 set local-preference 125
```

BGP Peer Groups

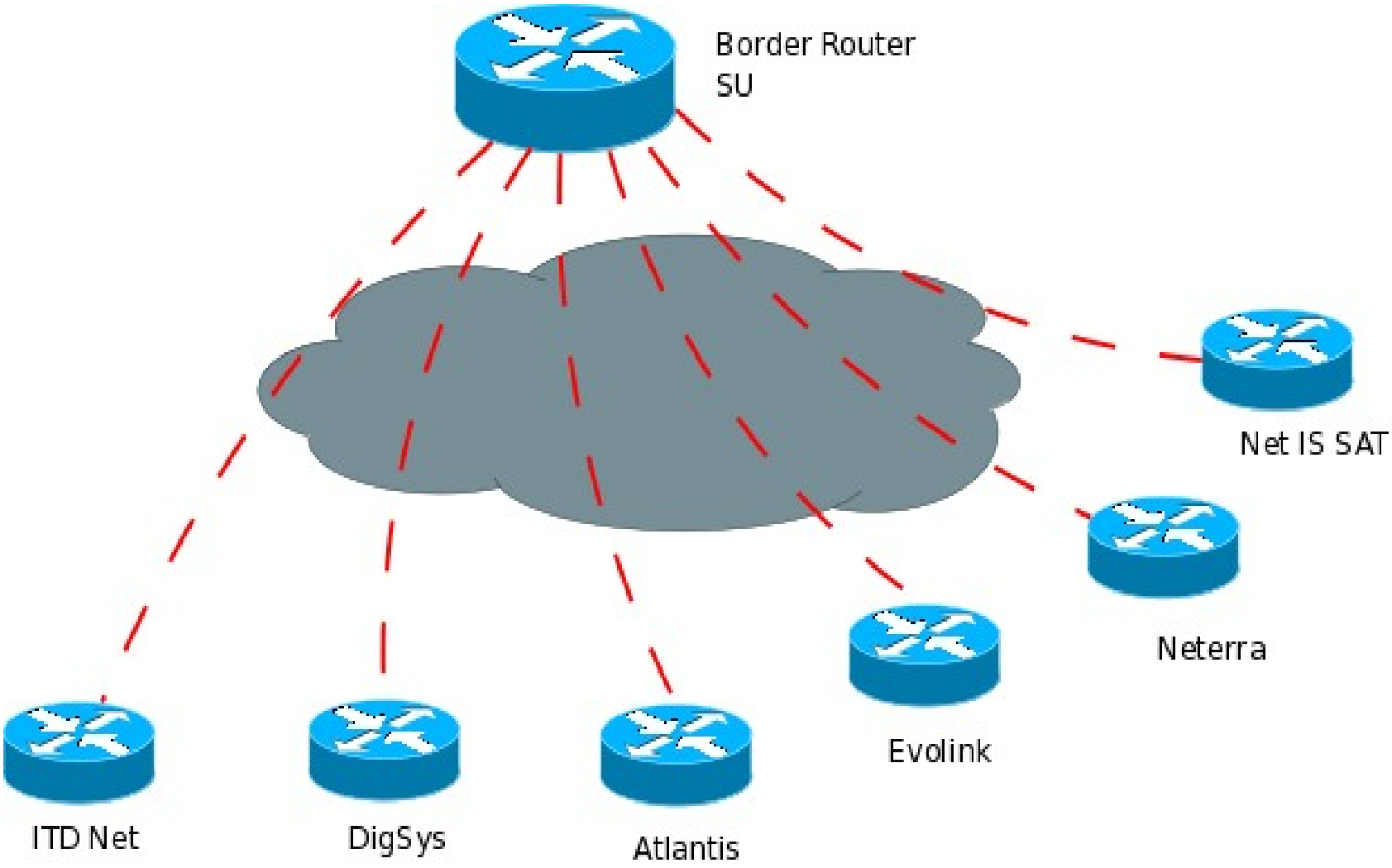
BGP peer group представлява група от BGP съседи, които споделят обща политика, определена от маршрутни карти и филтри - route maps, distribution lists.

Вместо политиката да се прилага на всеки съсед поотделно, тя се прилага върху цялата група.

Членовете на групата наследяват всички конфигурации на групата.

AS 5421 има peering споразумения с основните ISP да й подават само собствените си префикси.

Peering партньори на AS 5421



BGP Peer Groups. Конфигурация.

```
neighbor PEERING_DOUBLE_IPV4 peer-group
 neighbor PEERING_DOUBLE_IPV4 activate
 neighbor PEERING_DOUBLE_IPV4
   soft-reconfiguration inbound
 neighbor PEERING_DOUBLE_IPV4
   maximum-prefix 50000
 neighbor PEERING_DOUBLE_IPV4 route-map
   PEERING_DOUBLE_IMPORT_IPV4 in
 neighbor PEERING_DOUBLE_IPV4 route-map
   PEERING_DOUBLE_EXPORT_IPV4 out
```

BGP Peer Groups. Конфигурация.

neighbor 62.44.108.70 remote-as 9070

neighbor 62.44.108.70 peer-group PEERING_DOUBLE_IPV4

neighbor 62.44.108.70 description ITDNET_IPV4

Свързаност и научаване на маршрути в **BGP**

По принцип всички BGP маршрутизатори в дадена AS трябва да бъдат конфигурирани да “говорят” всеки с всеки (full mesh).

При това положение броят на връзките нараства квадратично с увеличаване на броя на рутерите.

BGP има две решения на това неудобство: рефлекторна схема (route reflectors - RFC 4456) и конфедерации (confederations - RFC 5065).

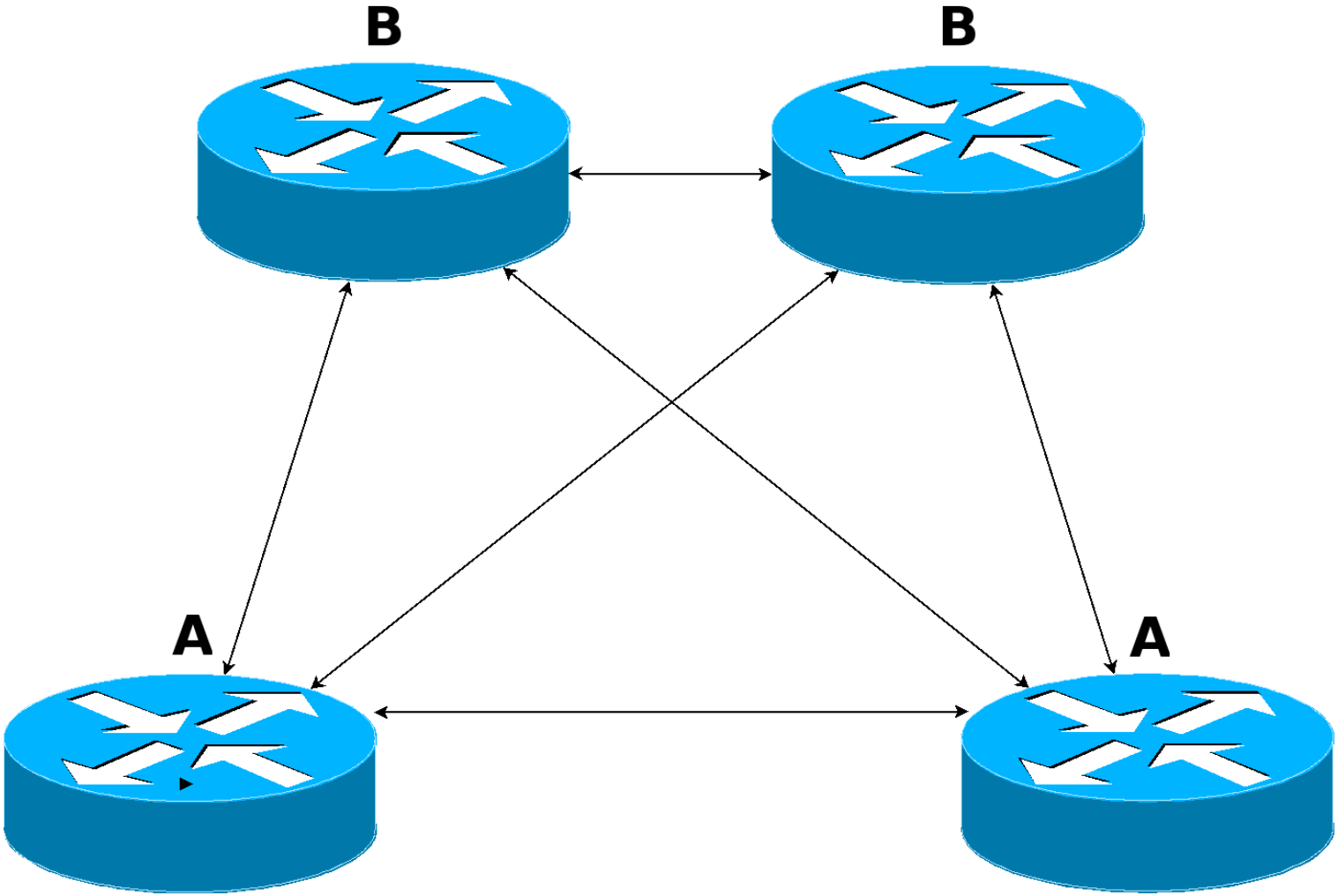
Рефлектори

В автономна система с iBGP трябва да има свързаност към всички iBGP peers (съседи), т.е “всеки с всеки” - full mesh.

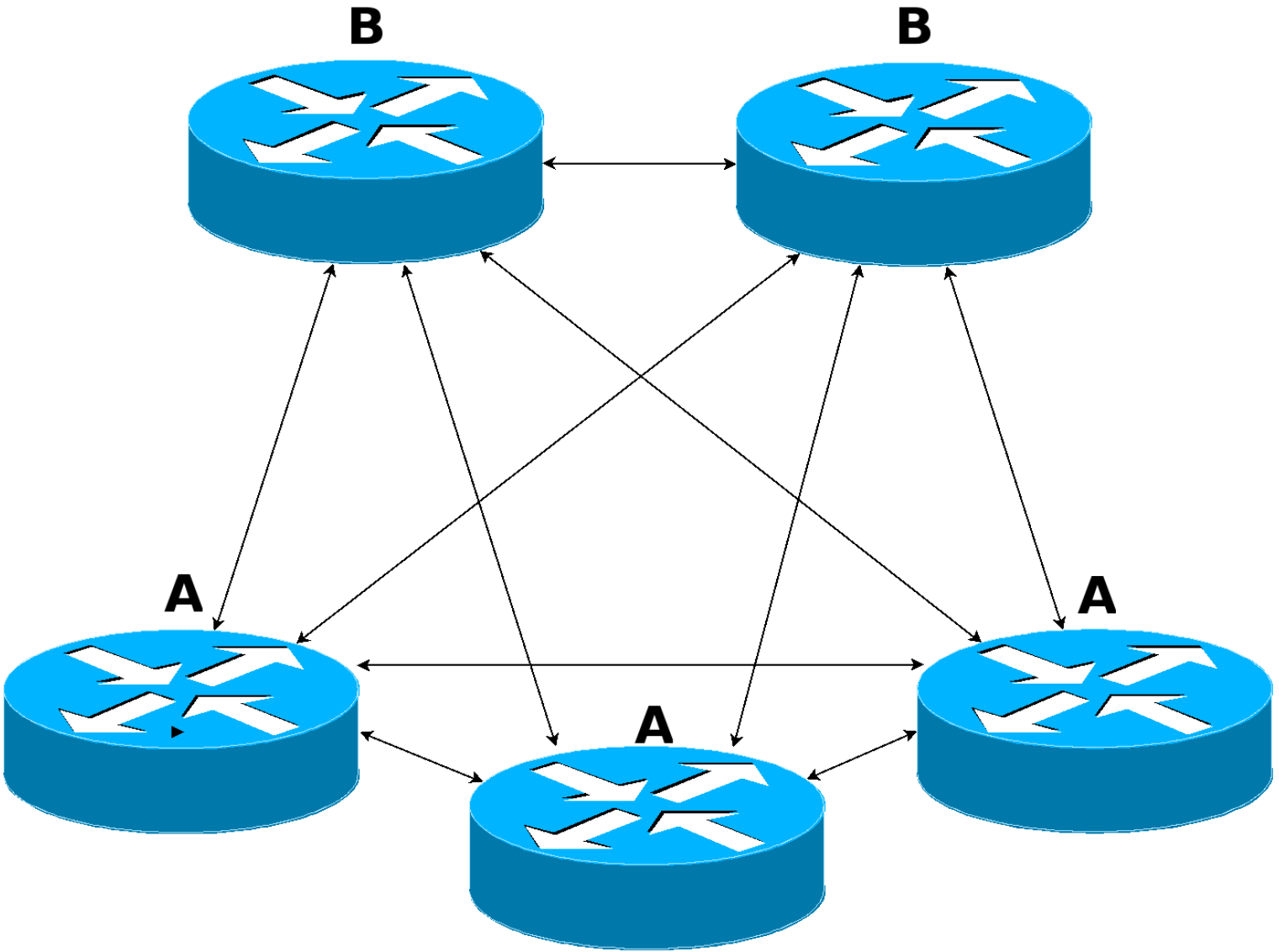
С помощта на рефлекторна схема се редуцира броя на iBGP съседите и от там натоварването на процесори и комуникационни канали.

Един рутер (или два за резервираност) става рефлекторен сървър, а другите – рефлекторни клиенти.

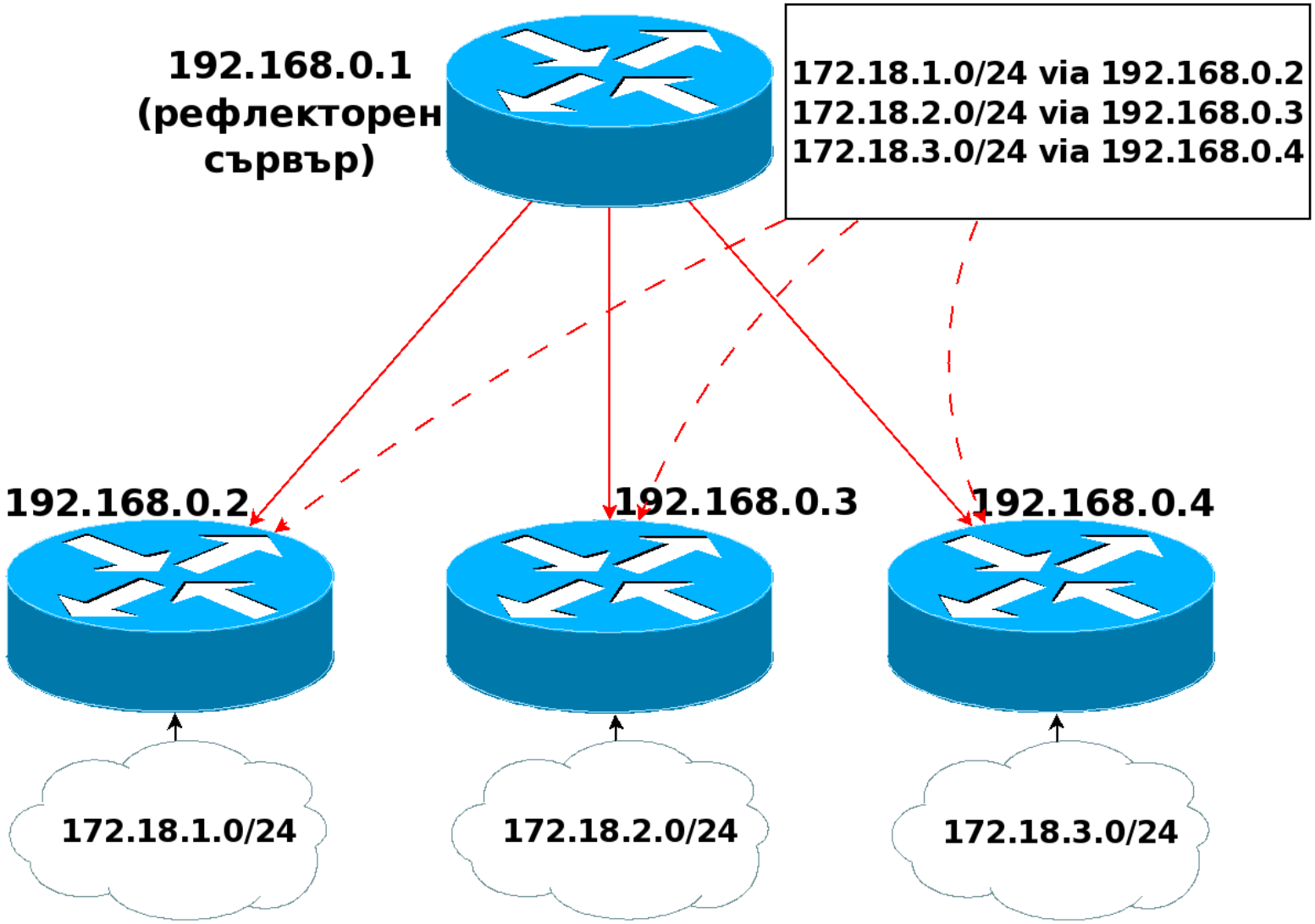
Full-mesh: 4 маршрутизатора => 6 вътрешни BGP сесии



5 маршрутизатора => 10 вътрешни BGP сесии



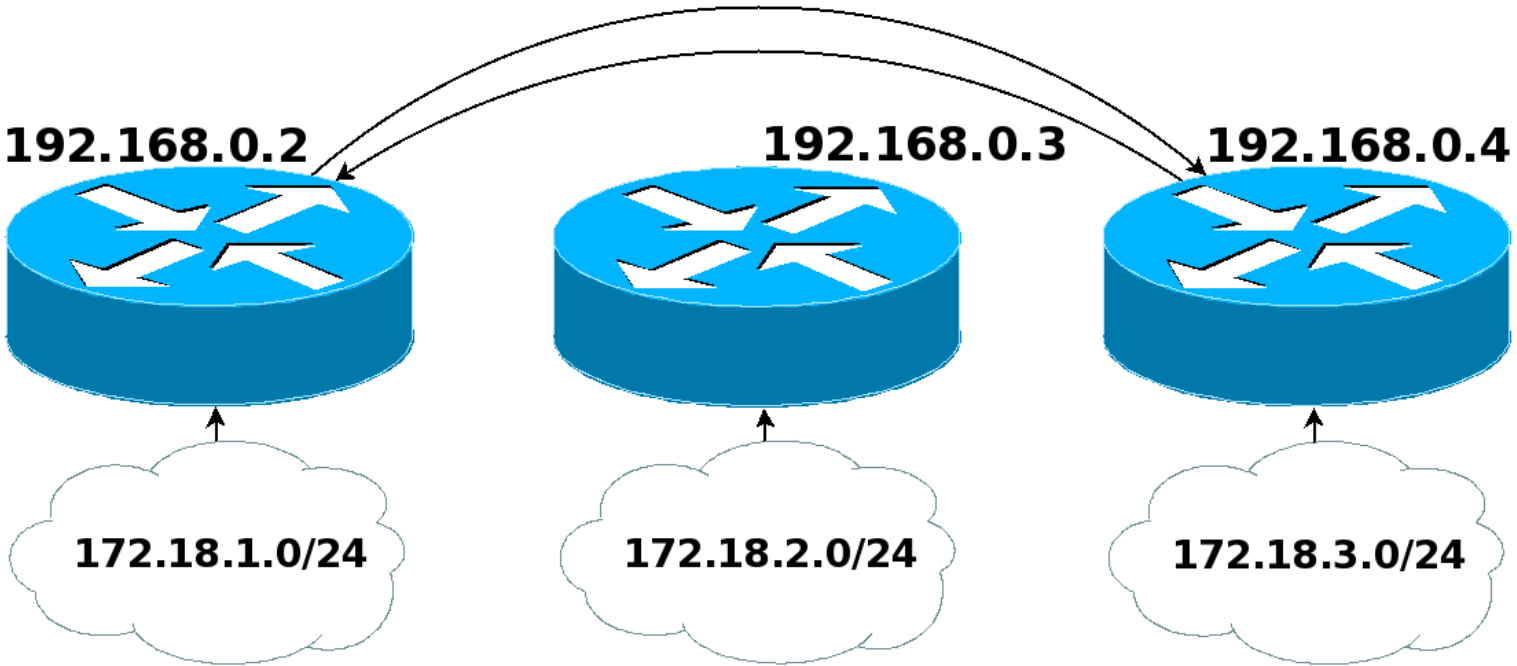
Рефлекторният сървър при решаване на задачата



Обмен на трафик между клиентите

```
172.18.1.0/24 via 192.168.0.2
172.18.2.0/24 via 192.168.0.3
172.18.3.0/24 via 192.168.0.4
```

ОБМЕН НА ТРАФИК ПО НАЙ-КРАТКИЯ ПЪТ - ДИРЕКТНО МЕЖДУ МАРШРУТИЗАТОРИТЕ В ЕТЕРНЕТ СЕГМЕНТА, БЕЗ УЧАСТИЕ НА РЕФЛЕКТОРНИЯ СЪРВЪР ПРЯКО В МАРШРУТИЗАЦИЯТА



Големина на маршрутната таблица

Един от основните проблеми пред BGP, респ. Internet, е растежа на глобалната таблица с маршрутите.

Не всички рутери са в състояние да я поемат (RAM, CPU) и ефективно да обработват трафика.

И, още по-важно, колкото е по-голяма таблицата, толкова по-бавно се стабилизира (конвергира).

В момента броят на префиксите в Глобалната мрежа е над 300 000.

Всички префикси подавани от Интернет провайдер

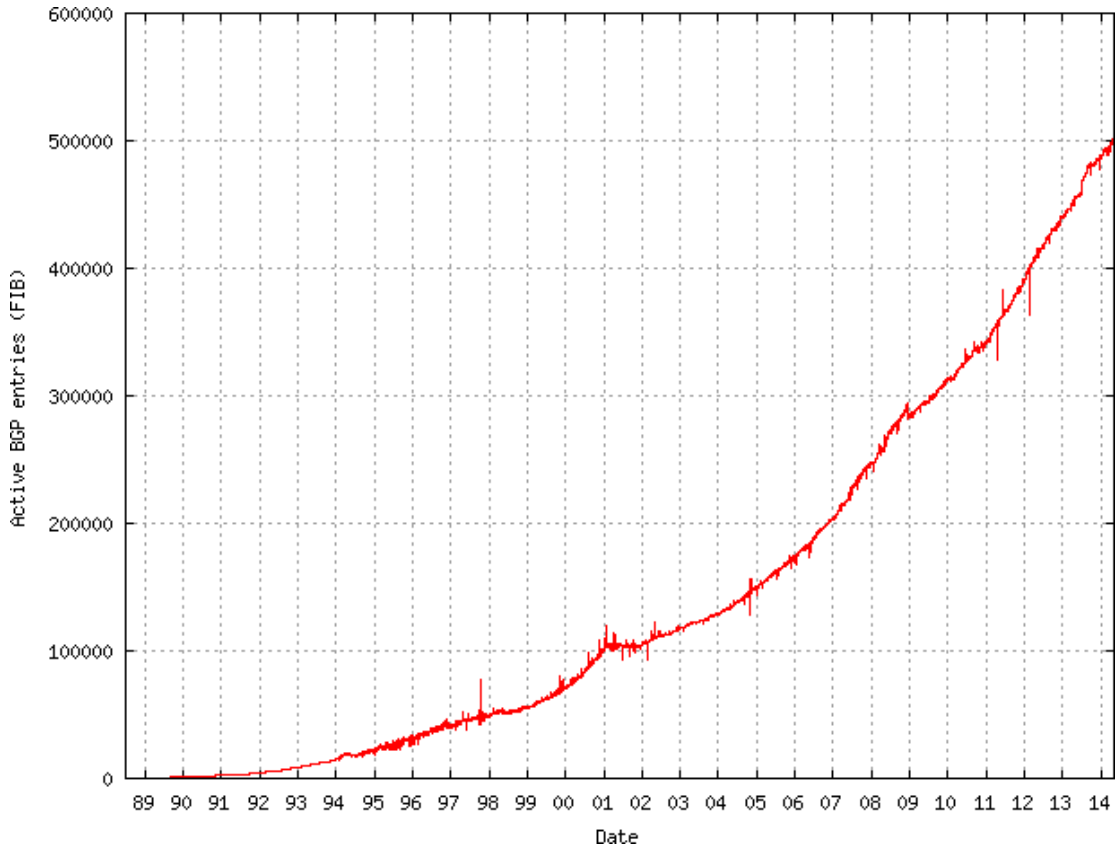
bgpd@border-lozenetz# sh ip bgp summary

BGP router identifier 62.44.127.21, local AS number 5421

RIB entries 725021, using 66 MiB of memory

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
62.44.96.234	4	8717	1382432	23634		0	0	0	01w1d06h 400218
194.141.252.21	4	6802	755845	12280		0	0	0	6d21h04m 402138

Глобалната IPv4 таблица достига 500k маршрута



Агрегиране и/или сумаризиране на маршрути

172.16.0.0/16 (summary)
172.16.0.0/18
172.16.64.0/18
172.16.128.0/18

!172.16.192.0/18
празно

или
172.16.0.0/17 !aggregated
172.16.128.0/18

Агрегиране и/или сумаризиране на маршрути

Да приемем, че на AS1 е присвоено адресно пространство 172.16.0.0/16 (summary).

AS1 иска да анонсира по-специфични маршрути: 172.16.0.0/18, 172.16.64.0/18 и 172.16.128.0/18.

Префиксът 172.16.192.0/18 не съдържа никакви хостове и AS1 не го анонсира.

При това положение AS1 ще анонсира 4 маршрута:172.16.0.0/16, 172.16.0.0/18, 172.16.64.0/18 и 172.16.128.0/18.

Агрегиране и/или сумаризиране на маршрути

Тези 4 маршрута ще бъдат видяни от AS2.

Въпрос на политика е дали да ги копира 4-те или или да запише само сумаризирания (summary), 172.16.0.0/16.

Ако AS2 иска да изпрати данни към 172.16.192.0/18, те ще се отправят по маршрут 172.16.0.0/16.

Граничният маршрутизатор на AS1 или ще изхвърли пакета, или ще го върне като “unreachable” в зависимост от конфигурацията.

Агрегиране и/или сумаризиране на маршрути

Ако AS1 реши да не анонсира маршрут 172.16.0.0/16 (т.е да не сумаризира) и остави 172.16.0.0/18, 172.16.64.0/18 и 172.16.128.0/18, в таблицата й ще има три маршрута.

AS2 ще вижда тези три маршрута в зависимост от политиката си или ще запише в паметта и трите, или ще агрегира префиксите 172.16.0.0/18 и 172.16.64.0/18 на 172.16.0.0/17.

Тогава в паметта на граничния маршрутизатор на AS2 ще се съхраняват само два маршрута: 172.16.0.0/17 и 172.16.128.0/18.

Агрегиране и/или сумаризиране на маршрути

Ако AS2 иска да изпрати данни към 172.16.192.0/18, те ще бъдат изхвърлени на нейната граница или към маршрутизаторите в AS2 ще бъде изпратено съобщение “unreachable” (а не към AS1), защото 172.16.192.0/18 няма да е в маршрутната таблица.

Извод: За намаляване на редовете в маршрутната таблица, да прилагаме:

Агрегация без сумаризация

Сигурността на BGP сесиите

На глобално ниво. Ние не можем да предвидим маршрутите, които ни се подават, дали са точно те. Идват отдалеч. Известни са случаи на подвеждане (Напр., Китай, Пакистан).

Решение: Resource Public Key Infrastructure (RPKI).

RPKI

От 1.01.11 RIR би трябвало да добавят слой на криптиране, така че ISPs и др. да могат да доказват, че са оторизирани да маршрутизират трафик за дадена AS.

Но засега само APNIC ще го правят. Има и скептицизъм, преработка на софтуер и др.

Сигурност. BGP и TCP.

На локално ниво: **IPSec AH** (*в рамките на AS5421*)

В BGP - удостоверяване чрез парола. Но след установяване на TCP сесия с всичките й предимства и недостатъци.

С помощта на IPSec - Authentication Header - защитата на най-ниското възможно ниво - OSI L3.

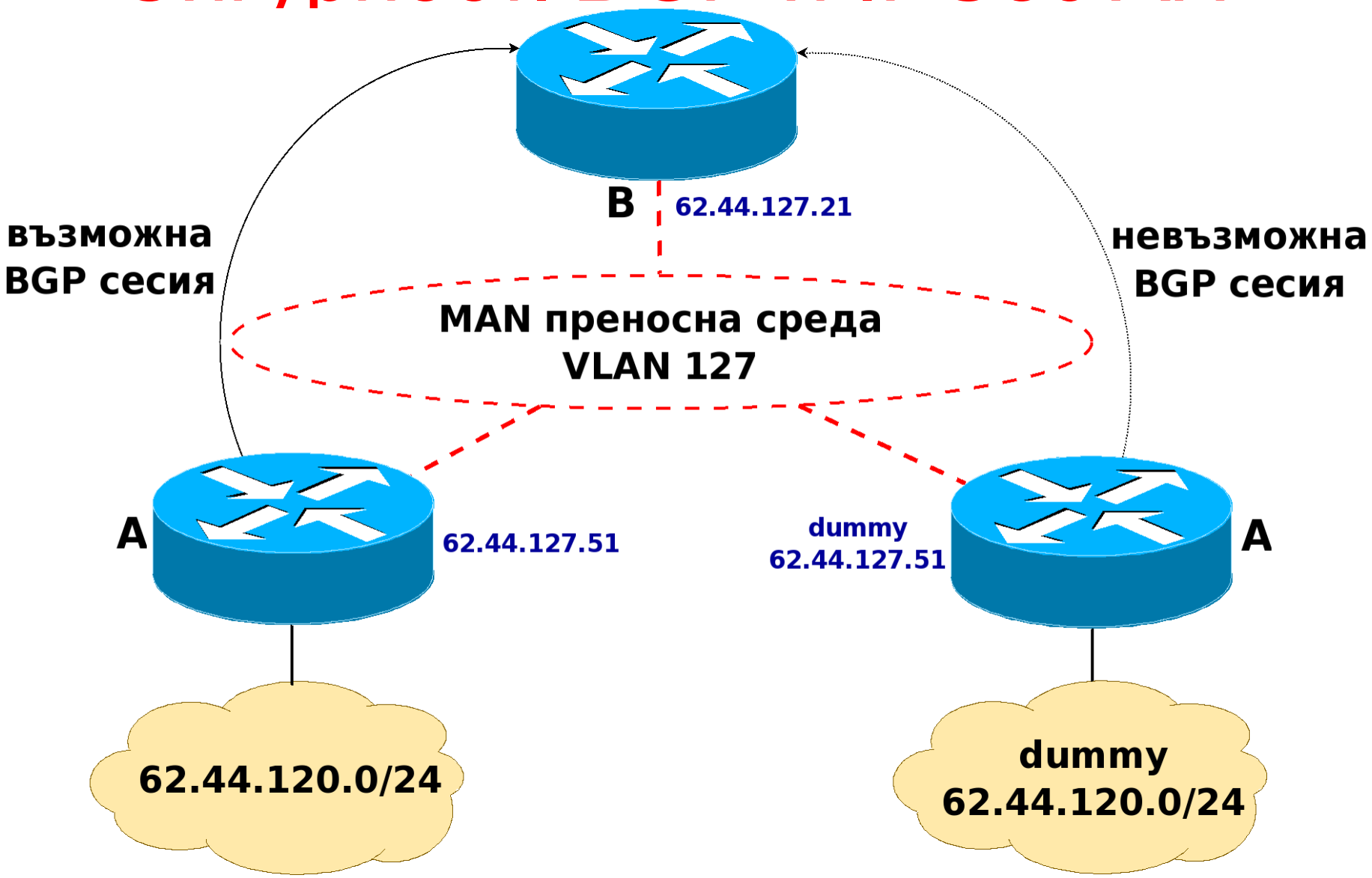
IPSec AH защита

Удостоверяваме страните в една TCP сесия с максимално ниво на сигурност - проверка на електронен подпис.

Така се защитаваме, например, от IP spoofing.

Следващия слайд:

Сигурност. BGP и IPSec AH



BGP и IPv6

BGP4/4+ с “multi-protocol extensions” поддържа едновременно IPv4 и IPv6 (RFC 4760).

```
[root@border-lozenets ~]# less /etc/quagga/bgpd.conf
!
router bgp 5421
 bgp router-id 62.44.127.21
 no bgp default ipv4-unicast
 network 62.44.96.0/19
...
```

BGP и IPv6

...

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fff6 remote-as 9070

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fff6 description ITD_BORDER_IPV6

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fffa remote-as 8262

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fffa description EVOLINK_BORDER_IPV6

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fffe remote-as 8717

neighbor 2001:67c:20d0:fffe:ffff:ffff:ffff:fffe description SPNET_BORDER_IPV6

BGP и IPv6

```
neighbor 2001:67c:20d0:ffff::3 remote-as 5421
 neighbor 2001:67c:20d0:ffff::3 description
    ivkm-gw.uni-sofia.bg/IPv6
 neighbor 2001:4b58:acad:252::25 remote-as 6802
  neighbor 2001:4b58:acad:252::25 description
    BIOM_BORDER_IPV6
!
 address-family ipv6
 network 2001:67c:20d0::/47
 network 2001:67c:20d0::/48
 network 2001:67c:20d0:fffe:ffff:ffff:ffff:fff4/126
```

Глобална IPv6 таблица

```
[root@border-lozenets ~]# vtysh -c "sh ipv6 bgp
   sum"
...
2001:67c:20d0:fffe:ffff:ffff:ffff:fffe
             4  8717  115137   23681        0    0    0
   01w1d06h     8772
2001:4b58:acad:252::25
             4  6802  175781   12304        0    0    0
   6d21h22m     8620
```

Глобална IPv6 таблица (предвиждания)