Маршрутен протокол BGP

Понятие за автономна система (AS)

Автномните системи (**autonomous system - AS**) са **основните градивни блокове, на които се разделя интернет пространството**.

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), както и EGP - BGP.

Понятие за AS. ASN.

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

Идентификатор на AS е 16-битов номер (ASN - Autonomous System Number) (0 до 65 535) до 2007 г.

2007 г. са въведени 32-битови ASNs, разширяващи адресацията от

0 до 4 294 967 295.

Към момента има разпределени над 100 000 автономни системи, като над 60 000 от тях са активни.

ASN – 16- и 32-битови. Присвояване на ASNs.

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.

ASN се използва при обмен на маршрутизираща информация със съседни AS.

ASNs, подобно на IP адресите се разпределят от IANA към RIR и организации.

(http://iana.org/assignments/as-numbers/as-numbers.xml)

Присвояване на ASNs. Кога ни трябва AS.

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

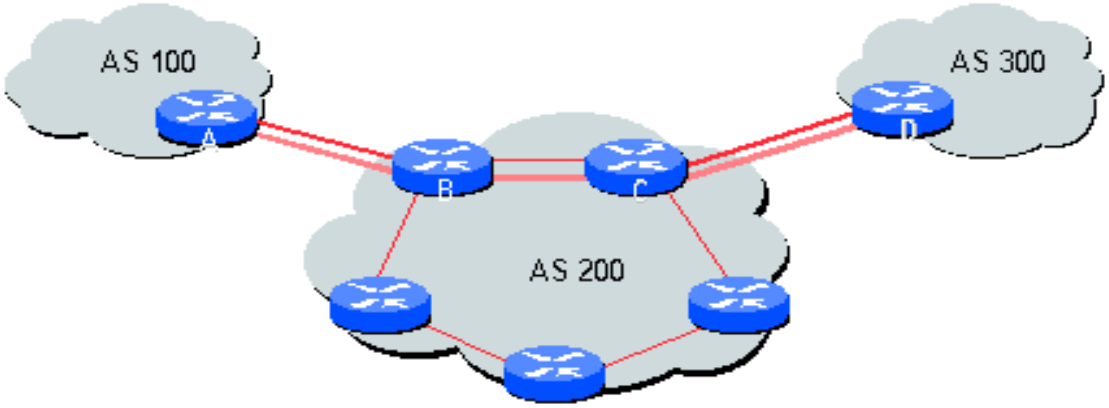
Single-homed site, единствен или множество префикси, свързан към един единствен доставчик (т.е. една AS).

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).

Multi-homed AS



Може да бъде **транзитна**, ако принадлежи на интернет провайдер (ISP) и е част от пътя до даден адрес.

Може и да е **stub**, ако е последна в пътя до даден адрес. Такава е **AS 5421** на СУ.

Single-homed AS задължително са stub.

Border Gateway Protocol (**BGP**)

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

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

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

BGP. Версии и стандарти.

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.

Последната и все още актуална версия BGP-4 е утвърдена през 2006 г. в RFC 4271.

Поддържа CIDR и обединяване (агрегация) на маршрути, с което се намалява размера на маршрутните таблици.

BGP. Версии, стандарти, алгоритми.

Развитие за протокола е въвеждането на възможност за пренос на информация за протоколи различни от IPv4 в (RFC4760).

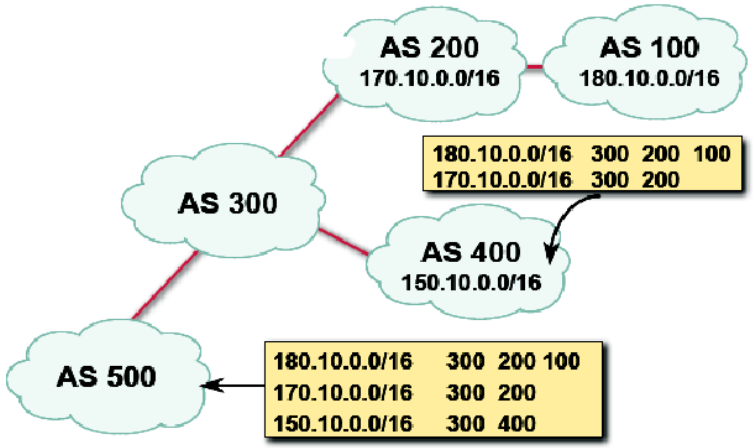
Нарича се MP-BGP (Multi Protocol BGP) и може да пренася multicast, MPLS и IPv6.

BGP4+ е условно наименование на употребата на MP-BGP специално за пренос на IPv6 префикси.

MP-BGP (респ. BGP4+) съвпада с BGP4 по същество.

BGP прилага вектора на пътищата (Path-vector routing).

Вектор на пътищата



Подобен на DV. Рутерите, които са на границата на AS - border routers - говорители (speakers).

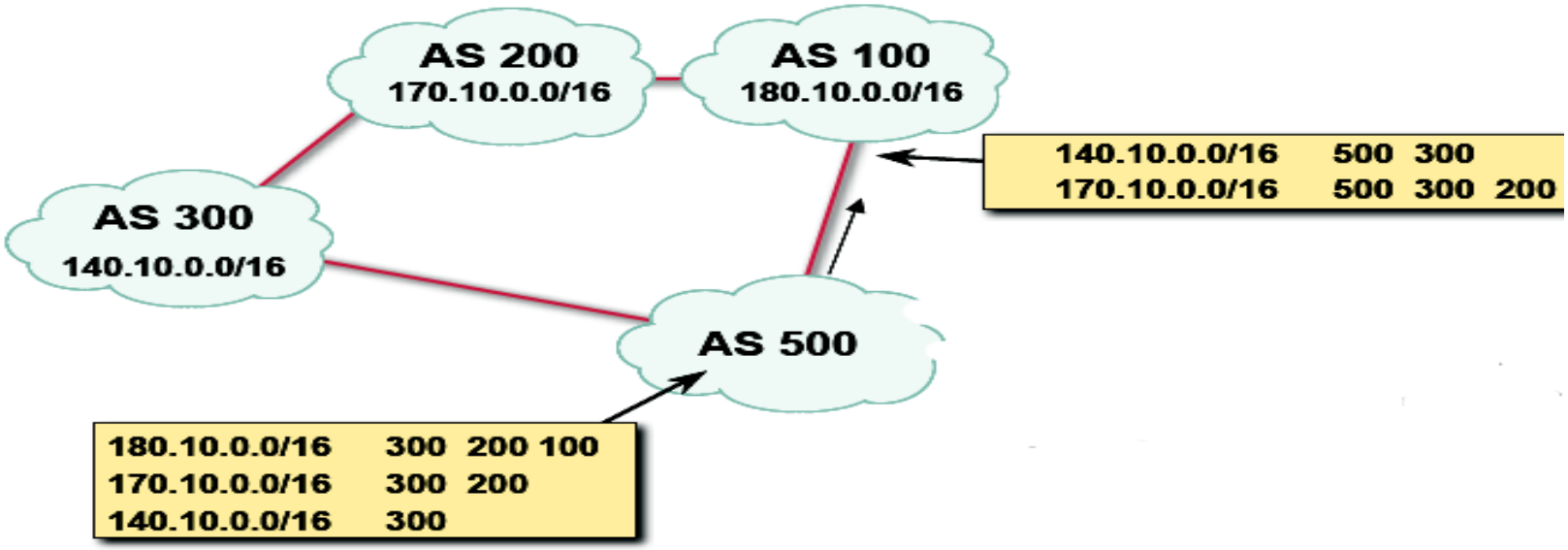
Те рекламират дестинациите, до които имат пътища, на speaker-те в съседни AS.

border routers рекламират маршрутите до дестинациите като "адрес и описание на пътя през ASs" до даден адрес. Затова алгоритъмът е path-vector routing.

Рутерите получават вектор, който съдържа пътища до определен брой дестинации.

Пътят се съдържа в специален "атрибут на пътя", описващ последователността от автономни системи до дестинацията.

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.

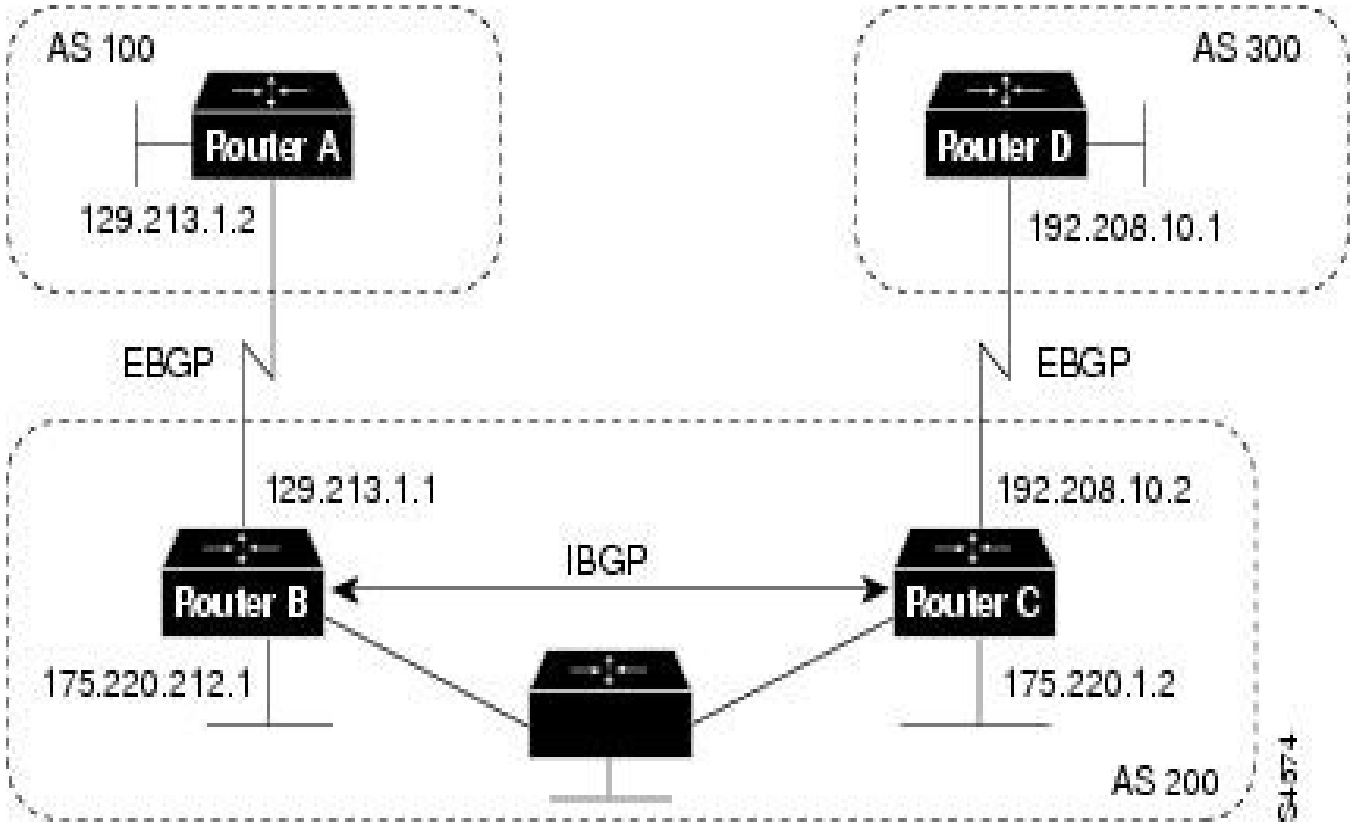
iBGP и eBGP

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
34574

iBGP и eBGP. Конфигурации.

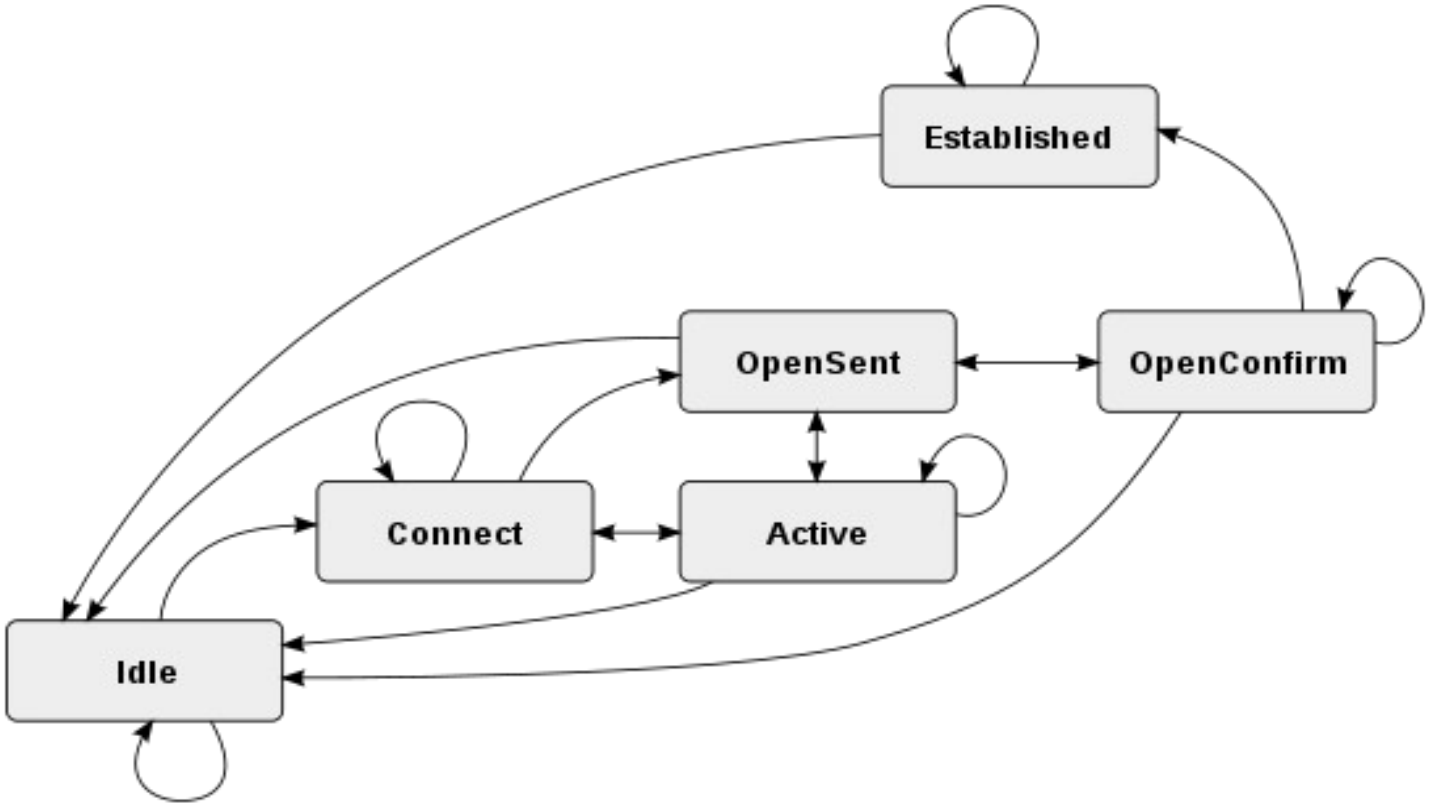
Router B:

```
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.

BGP инициира TCP сесия в състояние Connect.

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.

Състояния при установяване на BGP сесия

След успешно преминаване през началното и междинните състояния рутерът влиза в състояние “Established”.

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

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

Обмен на маршрутна информация

...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 ***>**

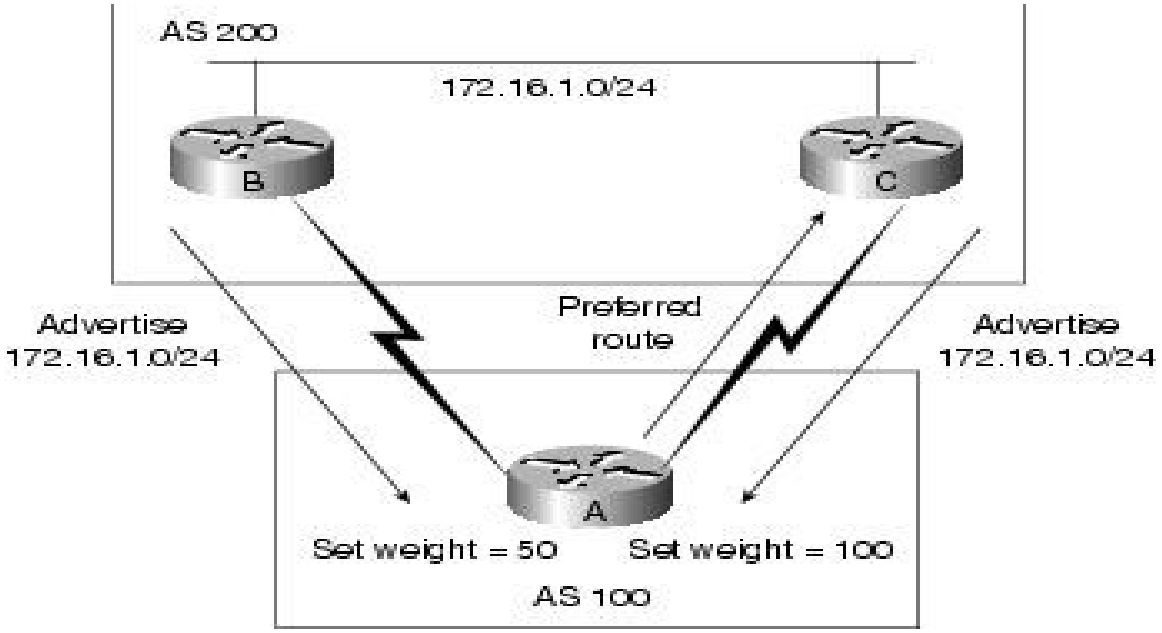
Избор на маршрут

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

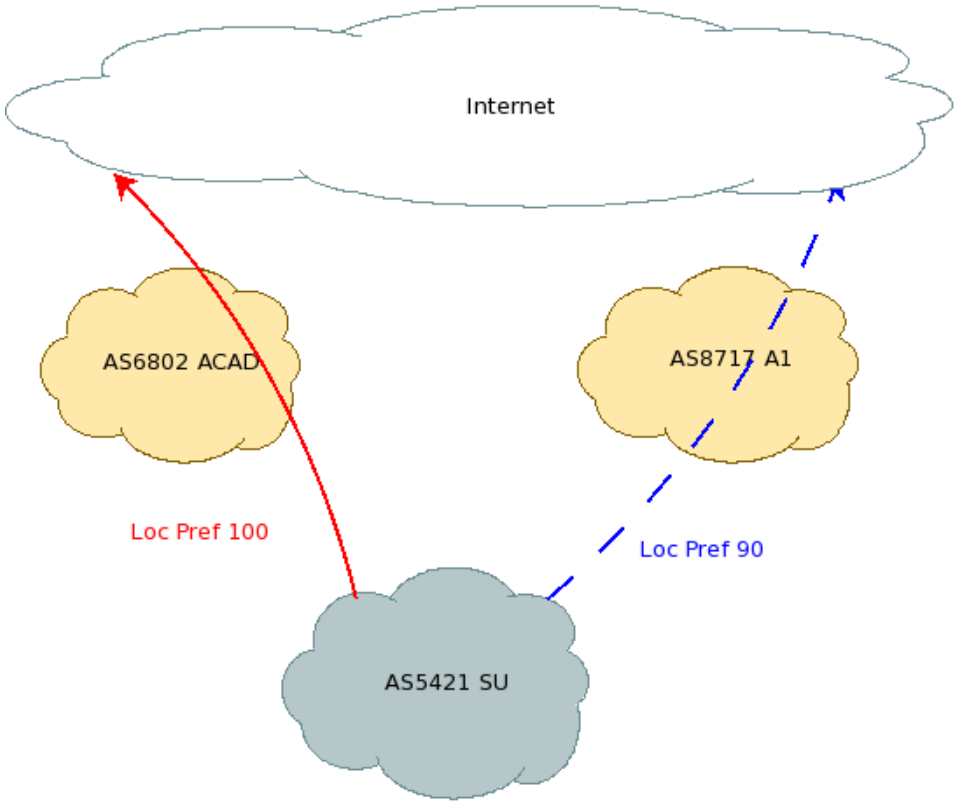
Локален за AS – нетранзитивен

local preference = 100 по подразбиране

Влияе на избора на път за изходящия трафик

Път с най-висок local preference печели

Local Preference в СУнет



Local Preference в СУнет

!Избран е маршрут #1, защото LocPrf=100 (default), макар AS-PATH да е по-дълъг:

```
[root@border-lozenets ~]# vtysh -c "sh ip bgp 1.9.0.0"
BGP routing table entry for 1.9.0.0/16
Paths: (2 available, best #1, table Default-IP-Routing-Table)
…
6802 21320 6939 4788
    194.141.252.21 from 194.141.252.21 (194.141.252.11)
      Origin IGP, localpref 100, valid, external, best
…
8717 6939 4788
    62.44.96.234 from 62.44.96.234 (10.251.203.41)
      Origin IGP, localpref 90, valid, external
```

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.

Кои IP мрежи ще рекламираме. Команда **network**...

(http://docs.frrouting.org/en/latest/bgp.html)

```
router bgp 1
 address-family ipv4 unicast
  network 10.0.0.0/8
!
 address-family ipv6 unicast
  network 2001:0DB8:5009::/64
 exit-address-family
```

IPv4 мрежата 10.0.0.0/8, респ. IPv6 - 2001:0DB8:5009::/64 ще бъдат рекламирани на всички съседи.

…или aggregate-address

```
!
address-family ipv4 unicast
  aggregate-address 62.44.96.0/19
  aggregate-address 62.44.96.0/24
  aggregate-address 62.44.97.0/24
  aggregate-address 62.44.103.0/24
  Aggregate-address 62.44.109.0/24
!
 address-family ipv6 unicast
  aggregate-address 2001:67c:20d0::/47
  aggregate-address 2001:67c:20d0::/48
```

AS-PATH PREPEND

Изкуствено удължаваме пътя през автономните системи и така решаваме откъде да минава трафика.

Напр., да минава през ACAD, а не през A1 (SPNET):

```
!
route-map SPNET_EXPORT_IPV4 permit 10
 match ip address prefix-list SU_SUPERBLOCK_IPV4
 set as-path prepend 5421 5421 5421 5421 5421 5421 5421 5421 5421 5421
!
route-map SPNET_BORDER_EXPORT_IPV6 permit 10
 match ipv6 address prefix-list LOZENETS_IPV6_PREFIXES
 set as-path prepend 5421 5421
```

Показване на Origin и др. атрибути за даден маршрут

```
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
```

BGP Peer Groups

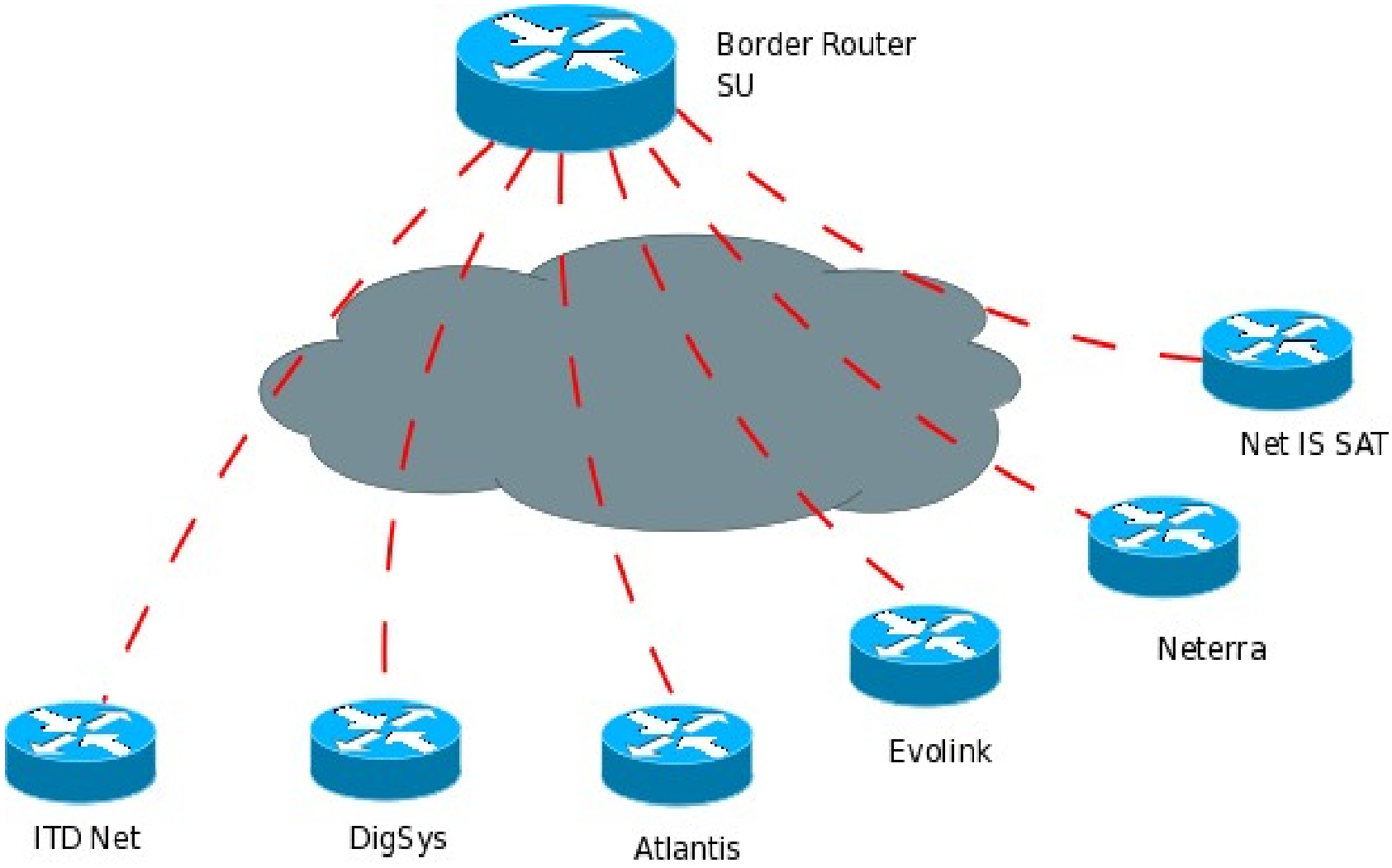
BGP peer group представлява група от BGP съседи, които споделят обща политика, определена от маршрутни карти и филтри - route maps, distribution lists.

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

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 стигна до **768k Day** и го надхвърли.

Моментно състояние на глобалната BGPv4 таблица

[root@border-lozenets ~]# vtysh -c "sh bgp ipv4 sum"

```
BGP router identifier 62.44.127.21, local AS number
   5421

RIB entries 1477132, using 135 MiB of memory

Peers 61, using 272 KiB of memory

Peer groups 8, using 256 bytes of memory

Neighbor         V    AS MsgRcvd MsgSent   TblVer  InQ
   OutQ Up/Down  State/PfxRcd

62.44.96.234     4  8717 53747647   56250        0    0
      0 02w1d11h   815449

194.141.252.11   4  6802 22311524  110938        0    0
      0 01w4d19h   824099
```

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

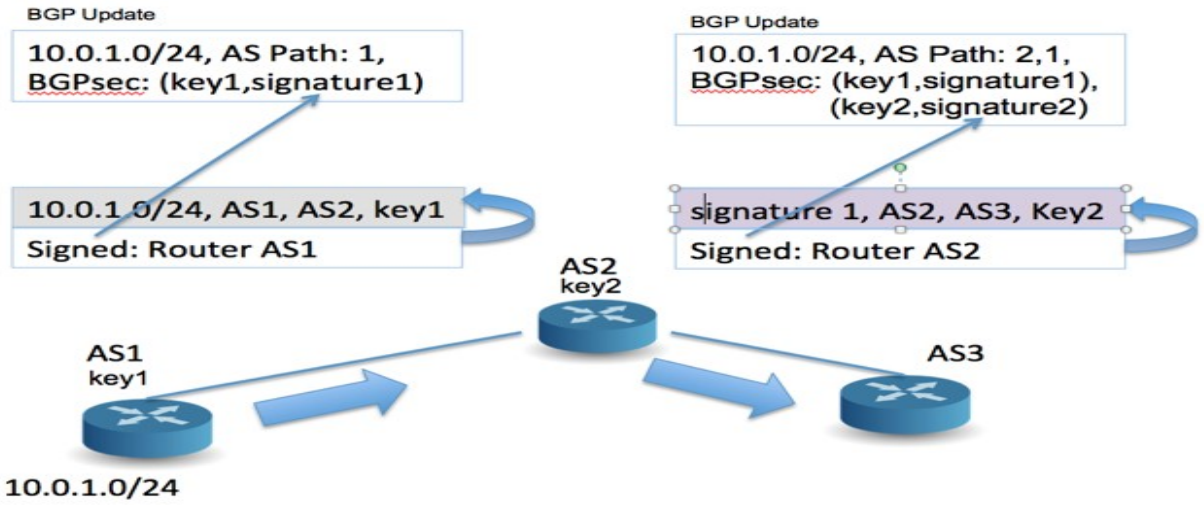
Сигурността на BGP сесиите

BGP има един недостатък, свързан със **сигурността**. Ние не можем да предвидим маршрутите, които ни се подават, от кого всъщност са подадени.

Постоянно има случаи на отклоняване на трафик през “екзотични” дестинации - Китай, Пакистан и др.

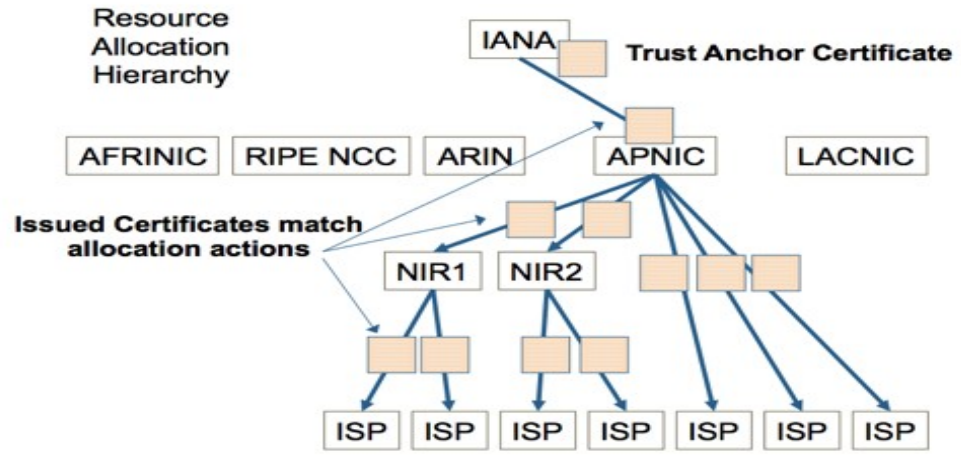
Решения на проблемите със сигурността на BGP4 е разширението му **BGPsec** (RFC8205), както и RPKI.

BGPsec



При BGPsec (RFC8205) всеки граничен маршрутизатор добавя X.509 подписан обект за AS, на която изпраща даден анонс. Този обект включва криптографско проверимо копие на пътя от автономни системи. Това решение, макар и пълно, създава много проблеми от имплементационна гледна точка. Добавянето на криптографски операции предполага обновяване на хардуера, за да не пострада производителността.

Resource Public Key Infrastructure (RPKI)



RPKI, дефиниран в RFC 6480, е предложение за X.509 йерархична инфраструктура, която да даде възможност на крайните AS да придобият публично проверими сертификати, с които да подписват съдържание.

RPKI въвежда децентрализирана X.509 PKI (Public Key Infrastructure) система, със CA (Certificate Authority) коренни сертификати (root certificates) на петте RIR организации. За целите на RPKI, тези 5 корена се наричат TA (Trust Anchor).

BGP и IPv6

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

```
router bgp 1
 no bgp default ipv4-unicast
 neighbor 10.10.10.1 remote-as 2
 neighbor 2001:0DB8::1 remote-as 3
 address-family ipv4 unicast
  neighbor 10.10.10.1 activate
  network 192.168.1.0/24
 exit-address-family
 address-family ipv6 unicast
  neighbor 2001:0DB8::1 activate
  network 2001:0DB8:5009::/64
 exit-address-family
```

Глобална IPv6 таблица (рекламирана на СУ)

[root@border-lozenets ~]# vtysh -c "sh bgp ipv6 sum"

```
BGP router identifier 62.44.127.21, local AS
  number 5421

RIB entries 164573, using 15 MiB of memory

...

Neighbor          V    AS MsgRcvd MsgSent   TblVer
  InQ OutQ Up/Down  State/PfxRcd

2001:67c:20d0:fffe:ffff:ffff:ffff:fffe
                  4  8717 8968128   56873        0
    0    0 02w1d22h    85279

2001:4b58:acad:252::11
                  4  6802 3860101  112119        0
    0    0 01w5d06h    99398
```

Растеж на глобалната IPv6 таблица