

Influențarea deciziei de rutare la BGP

Daniel Valentin Teodorescu
Grupa 441A - Diferență

Cuprins

I.	Introducere.....	3
II.	Atributele BGP si Algoritmul Best Path.....	3
III.	Influențarea rutelor cu exteriorul.....	4
	3.1 Influențarea BGP Weight.....	5
	3.2 Setarea Weight utilizând Route Map.....	5
	3.3 Setarea Weight utilizând comanda Neighbour Weight.....	7
	3.4 Setarea atributului Local Preference.....	7
	3.5 Rute IP bazate pe BGP Best Path	8
	3.6 BGP și comanda maximum-paths.....	9
	3.7 Influențarea rutelor interioare prin intermediul MED.....	10
	3.8 Mărirea valorii lungimii AŞ_Path utilizând AŞ_Path prepend.....	11
IV.	Încheiere.....	12

Influențarea deciziei de rutare la BGP

I. Introducere

Border Gateway Protocol (BGP), cel mai complex protocol de rutare, a fost inventat pentru a înlocui EGP (primul protocol de tip EGP).

BGP face cunoscute, învață, și alege cele mai bune căi în interiorul Internetului.

Când doi Furnizori de Servicii de Internet se interconectează, utilizează de obicei BGP pentru a interschimba informații.

Când o Companie dorește să se conecteze la Internet, ea poate utiliza rute statice sau BGP. BGP nu este necesar în toate cazurile.

BGP folosește algoritmul cea mai bună cale (Best Path).

BGP stabilește mai întâi o relație de vecinătate cu alte routere înainte de a schimba informații legate de topologie cu alte routere.

Border Gateway Protocol este un protocol de rutare unic, deoarece, spre deosebire de celelalte protocoale de rutare, stabilește și menține conexiuni între ruterele vecine folosind protocolul TCP. În cazul ruterele aflate în AS-uri diferite, o conexiune BGP poate fi stabilită doar dacă ruterele sunt direct conectate. Legătura se realizează pe portul TCP 179, fiind menținută prin mesaje periodice de 19 octeți (intervalul implicit este de 60 de secunde).

Cu toate că BGP este proiectat pentru a fi un EGP, el vine ca un protocol suport pentru două subprotocoale: eBGP și iBGP.

BGP este un protocol de rutare foarte robust și scalabil (funcționează corect indiferent de mărimea rețelei), cum este dovedit de faptul că BGP este protocolul de rutare folosit în Internet. Numărul de rute, din tabelele de rutare Internet BGP, sunt peste 90000. Pentru a putea atinge scalabilitatea la acest nivel, BGP folosește multe atribute ale rutelor.

BGP utilizează atribute "Path Attribute" (PA) pentru o serie descopuri. Atributele definesc informațiile despre o cale, o rută prin rețea. Câteva atribute BGP definesc informația necesară în alegerea celei mai bune rute BGP, utilizând algoritmul Best Path; exista și atribute folosite pentru alte scopuri, nu numai pentru Best Path.

II. Atributele BGP si Algoritmul Best Path

Fiecare atribut BGP definește câte o caracteristică diferită a unei rute prin rețea.

BGP folosește atributul AS_Path (Autonomous System_Path) ca unealtă principală de prevenire a buclelor; de asemenea BGP folosește atributul AS_Path pentru a calcula AS_Path length.

BGP definește de asemenea adresa IP următoare (Next_Hop) a unei căi drept atribut. Tabelul 1 listează o serie de atribute și trasatura BGP Weight (care nu este atribut) utilizate în implementarea algoritmului Best Path.

Tabelul 1. Atribute BGP utilizate in algoritmul Best Path

Atribut	Descriere
---------	-----------

Next_Hop	Listeaza adresa IP next hop utilizata pentru a ajunge la un prefix
Local_Pref	Valoare numerica ,0 – 2 ³² -1, setata si comunicata intr-un singur AS cu scopul de a influenta alegerea celei mai bune cai pentru toate routerele din AS.
AS_Path (length)	Numarul de ASN din atributul AS_Path
Origin	Valoare ce implica faptul ca ruta a fost injectata in BGP; I(IGP), E(EGP), sau ?(informatii incomplete)
MED (Multi Exit Discriminator)	Setat si anuntat de routere într-un AS, cu impact asupra deciziilor BGP. O valoare mai mica este mai buna.
Weight	O valoare numerica; 0 – 2 ¹⁶ -1, setata de un router cand primeste updates; influenteaza ruta pentru un anumit prefix.

Termenul de algoritm Best Path în BGP se referă la faptul ca BGP examinează într-un singur router toate rutele posibile din tabela sa BGP, pentru un singur prefix, alegând o singură rută drept cea mai bună rută.

Algoritmul Best Path urmărește pașii prezentați pe scurt în tabelul 2.

Tabelul 2. Procesul de decizie din BGP

Pas	Descriere scurta	Care valoare este mai buna
0	Next_hop accesibil?	Daca nu, nu se poate utiliza aceasta ruta
1	Weight	Valoarea mai mare
2	Local_Pref	Valoarea mai mare
3	Rute injectate local	Mai bune decat iBGP/eBGP
4	AS_Path length	Valoarea mai mica
5	Origin	I este preferat lui E; E este preferat lui ?
6	MED	Mai mica
7	Neighbour Type	eBGP preferat lui iBGP
8	IGP metric to next hop	Valoarea mai mica

Dacă un router nu a determinat cea mai bună cale la finalul pasului 8, acesta va mai urmări câțiva pași pentru a rupe egalitatea. Pașii adiționali sunt:

- Cea mai veche rută cunoscută eBGP
- Cel mai mic RID al unui vecin BGP
- Cea mai mică adresă IP a unui vecin

De îndată ce unul din cei trei pași de mai sus determina ruta cea mai bună, comparația rutelor se oprește.

O parte din pașii BGP Best Path oferă inginerului un instrument de influențare a alegerii celei mai bune căi, pe când ceilalți pași au alte scopuri, adesea fiind doar o reacție adversă a unor trăsături BGP. Așadar, când un inginer începe construirea unui plan de implementare BGP, trebuie luat în calcul doar un subset al celor 8 pași de bază, și anume:

- Weight (pasul 1)
- Local_Pref (pasul 2)

- AS_Path length (pasul 4)
- MED (pasul 6)

III. Influențarea rutelor cu exteriorul

Există trei factori de decizie diferiți care influențează rutele cu exteriorul ale unei companii: factorul Weight, atributul Local_Pref, și lungimea AS_Path.

Influențarea BGP Weight

Un router Cisco poate utiliza BGP Weight pentru a influența alegerea rutelor externe a acestui router. Pentru aceasta, atunci când un router primește un BGP Update, acel router poate seta parametrul Weight fie selectiv, per rută, folosind o hartă a rutelor (a route map), sau pentru toate rutele învățate de la un singur vecin. Algoritmul 'Best Path' examinează apoi valoarea Weight a diferitelor rute, alegând ruta cu cea mai mare valoare pentru Weight.

Setările pentru Weight configurate pe un singur router pot influența doar acel router, deoarece parametrul Weight nu poate fi comunicat routerelor vecine. Așadar pentru a folosi Weight, un router trebuie să fie configurat să examineze updatările primite pentru a seta Weight. Mesajele Update nu suportă un câmp în care să fie comunicată setarea parametrului Weight. Weight este o trăsătură proprie a Cisco; nu este un atribut.

Tabelul 3 sumarizează principalele trăsături ale Weight.

Tabelul 3. Principalele trăsături ale Weight

Trasatura	Descriere
Este un atribut PA?	Nu; este o trăsătură proprietate a Cisco
Scop	Identifica cea mai bună ruta pentru un anumit router
Gama	0 – 65536 ($2^{16}-1$)
Care valoare este mai bună	Valorile mai mari sunt mai bune
Valoare Implicită	0 pentru rute învățate; 32768 pt rute injectate local
Definirea de noi valori implicite	Nu e suportată
Configurare	neighbour route-map (per prefix) neighbour weight (toate rutele învățate de la acest vecin)

Exemplu

În figura 1 este ilustrat un mod de a seta Weight. Figura prezintă o singură companie și un singur router al companiei și doi furnizori de servicii. Cele 3 routere din figura îndeplinesc următoarele:

- Routerele C1 și I1-1 folosesc adrese IP de loopback (11.11.11.11 și 1.1.1.1) pentru relația lor de vecinătate.
- Routerele C1 și I3-1 folosesc adrese IP de interfață pentru relația lor de vecinătate.

- Niciunul dintre cele 3 routere nu au încercat să schimbe vre-o setare care să impacteze alegerea celei mai bune cai.

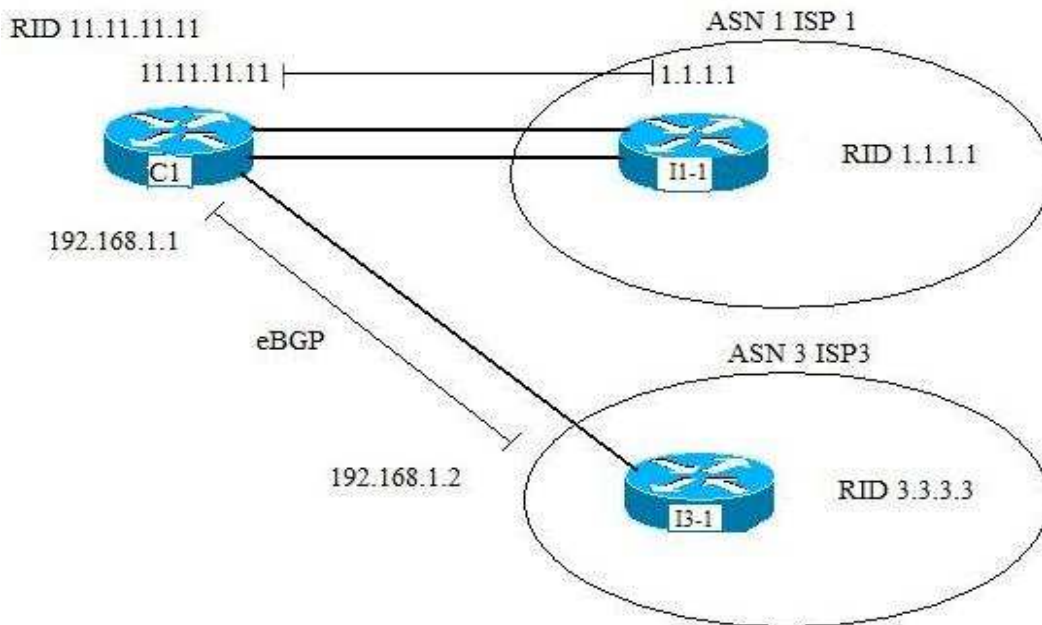


Figura 1 Exemplu BGP Weight

În continuare routerile ISP anunță fiecare rute BGP pentru aceleași 5 prefixe. În figură 2 sunt ilustrate 5 asemenea prefixe anunțate de cele două routere ISP către routerul C1.

Cea mai bună cale pentru 4 din cele 5 prefixe este evidentă: Prefixele 181.0.0.0/8 și 182.0.0.0/8 au AS_Path mai scurt prin ISP3, iar prefixele 184.0.0.0/8 și 185.0.0.0/8 au AS_Path mai scurt prin ISP1. Doar 183.0.0.0/8 ridică întrebări deoarece distanța AS_Path pentru cele două rute este aceeași.

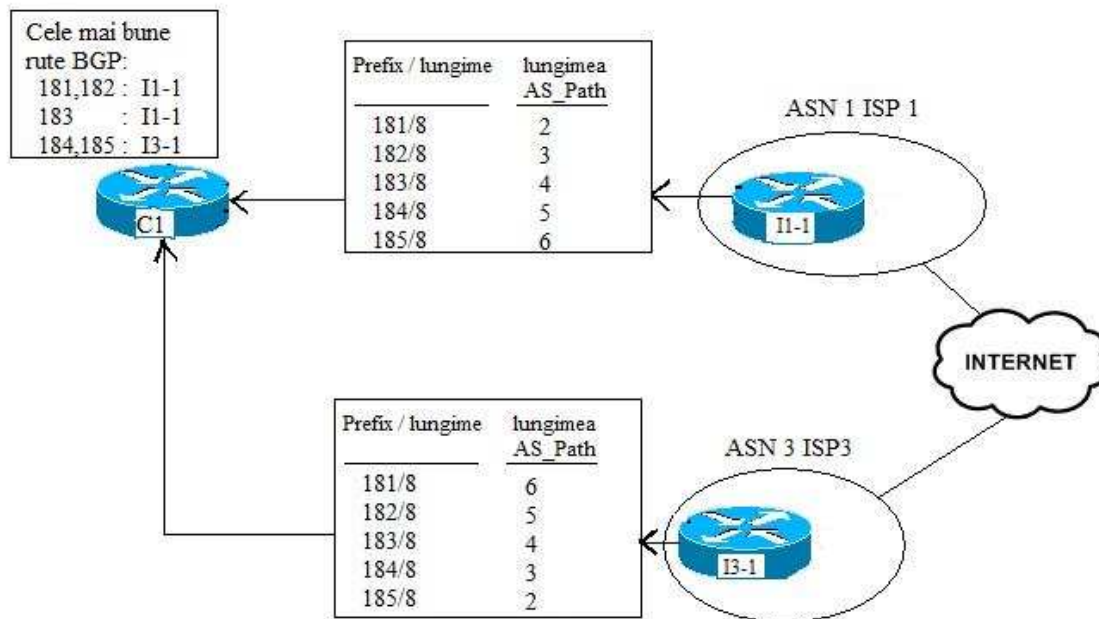


Figura 2 Prefixe si AS_Path Lengths

Mai întâi considerăm algoritmul Best Path în routerul C1 pentru prefixul 181.0.0.0/8. C1 cunoaște două rute BGP pentru 181.0.0.0/8. Cea mai bună cale are adresa IP 1.1.1.1 (I1-1) drept next hop. Următorii pași au fost urmați pentru a lua această decizie:

- Pasul 0. Fiecare Next_hop este accesibil;
- Pasul 1. Weight are aceeași valoare pentru ambele căi, și anume 0;
- Pasul 2. Local_Pref nu este setat pentru niciuna;
- Pasul 3. Niciuna dintre rute nu este injectată local;
- Pasul 4. AS_Path length are o valoare mai mică pentru ruta care trece prin I1-1.

Apoi considerăm exemplul rutei către 183.0.0.0/8. Au fost urmați pașii următori pentru a ajunge la o decizie:

- Pasul 0. Fiecare Next_hop este accesibil;
 - Pasul 1. Weight are aceeași valoare pentru ambele căi, și anume 0;
 - Pasul 2. Local_Pref nu este setat pentru niciuna;
 - Pasul 3. Niciuna dintre rute nu este injectată local;
 - Pasul 4. AS_Path length are valoarea 4.
 - Pasul 5. Ambele au Cod de Origine: i;
 - Pasul 6. MED are aceeași valoare pentru ambele căi, și anume 0;
 - Pasul 7. Tipul de vecin este eBGP pentru ambele rute;
 - Pasul 8. Metrica IGP nu se aplică (ambele rute sunt statice);
 - Pasul 9. Ruta învățată de la I1-1 este cea mai veche (este listată ultima la comanda show ip bgp).
- Dacă routerul I1-1 va fi restartat, cea mai veche ruta va fi cea prin I3-1.

Setarea Weight utilizând Route Map

Subcomanda **neighbour neighbour_ip route map** în BGP îi spune unui router să aplice route map tuturor Update-urilor primite de la vecinii listați.

Hărțile rutelor BGP (route maps) pot fi folosite de asemenea pentru a schimba atributele rutelor prin intermediul comenzii set. Route Map poate conține permit classes ce pot determina ca anumite rute să nu fie filtrate. În aceleași clase route map, includerea comenzilor precum **set weight 100** și **set local-preference 200** pot fi folosite pentru a seta parametrii Weight sau Local_Pref ai unei rute.

Setarea Weight utilizând comanda Neighbour Weight

Alternativ, Weight poate fi setat pentru toate rutele învățate de la un vecin folosind comanda **neighbour weight**.

În figură 3 este ilustrată această configurație adăugată routerului C1, setând Weight la 60 pentru toate rutele învățate de la I1-1 (1.1.1.1). Drept rezultat ruta aleasă de C1 pentru a ajunge la 181.0.0.0/8 va alege ruta prin I1-1 (1.1.1.1).

```
E1# conf t
Enter configuration commands, one per line. End with CNTL/Z.
E1(config)# router bgp 11
E1(config-router)# neighbor 1.1.1.1 weight 60
E1(config-router)#^Z
E1# clear ip bgp 1.1.1.1 soft
```



```

E1# show ip bgp 176.0.0.0/4 longer-prefixes
BGP table version is 54, local router ID is 128.107.9.1
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
*> 181.0.0.0/8      1.1.1.1           0             60 1 1811 i
*                   192.168.1.2       0             50 3 2 50 51 52 1811 i
*> 182.0.0.0/8      1.1.1.1           0             60 1 2 1822 i
*                   192.168.1.2       0             0 3 2 50 51 1822 i
*> 183.0.0.0/8      1.1.1.1           0             60 1 2 50 1833 i
*                   192.168.1.2       0             0 3 2 50 1833 i
*> 184.0.0.0/8      1.1.1.1           0             60 1 2 50 51 1844 i
*                   192.168.1.2       0             0 3 2 1844 i
*> 185.0.0.0/8      1.1.1.1           0             60 1 2 50 51 52 1855 i
*                   192.168.1.2       0             0 3 1855 i

```

Figura 3. Setarea Weight la valoarea 60 pentru toate rutele invatate dela I1-1

Setarea atributului Local Preference

Atributul Local_Pref oferă rutelor din interiorul unui sistem autonom AS o valoare pe care poate fi setată per rută, anunțată tuturor routerelor iBGP din interiorul AS, astfel încât toate routerele din AS să se pună de acord în legătură cu routerul cel mai potrivit ca exit point pentru pachetele destinate acelui prefix. Local_Pref poate fi setat de către routere pe măsură ce primesc rute eBGP, folosind route map interior (inbound route map). Routerele apoi anunță Local_Pref în notificările iBGP (iBGP updates). Drept rezultat, toate routerele din același AS pot să facă aceeași alegere în legătură cu cea mai bună rută, punându-se de acord ce router să folosească pentru a ieși din AS pentru fiecare prefix.

Pentru a seta Local_Pref, un router poate utiliza **neighbour neighbour_ip route-map** în BGP. De obicei un router folosește această comandă cu direcțiile interne pentru rutele primite de la routerele eBGP vecine. Apoi, fără a fi necesară vre-o configurare adițională, routerul anunță Local_Pref tuturor routerelor eBGP vecine. De reținut că BGP nu include atributul Local_Pref în eBGP updates, ci în iBGP updates.

Rute IP bazate pe BGP Best Path

O parte din complexitatea protocolului BGP se observă la funcțiile BGP create de atributele BGP, incluzând utilizarea lor în algoritmul Best Path. După ce algoritmul a trecut peste această complexitate și a ales cea mai bună rută pentru un prefix, routerul

încearcă să adauge această rută la tabela de rutare IP. În loc să adauge direct această rută în table, BGP trimite această rută altui process: Managerul Tabelei de Rutare IOS (IOS RTM).

IOS RTM alege cea mai bună ruta din mai multe surse. De exemplu rutele pot fi învățate de IGP,BGP, sau pot fi chiar rute statice. IOS adună cea mai bună astfel de rută pentru fiecare prefix și o trimite către funcția RTM. RTM va alege apoi cea mai bună rută. Figura 4 ilustrează ideea generală.

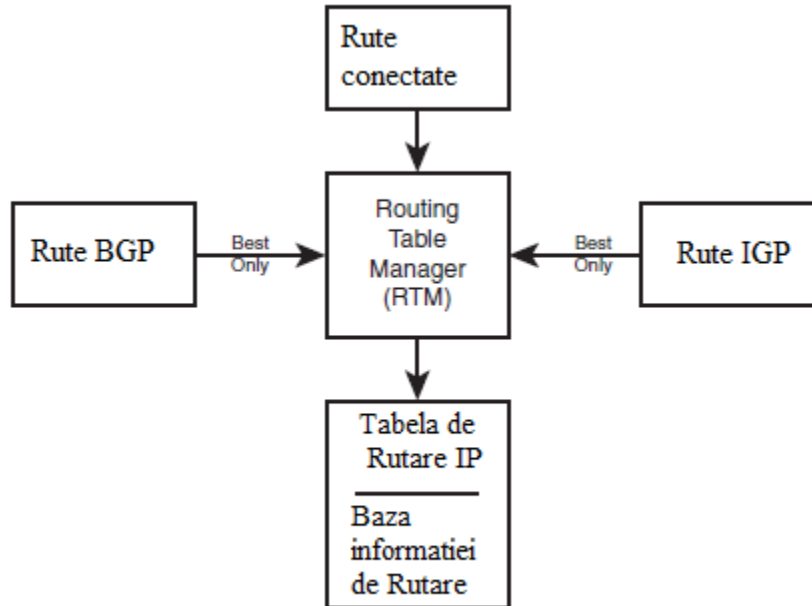


Figura 4. Conceptul de Routing Table Manager

Printre alte sarcini, RTM utilizează conceptul de distanta administrativă (AD) pentru a alege cea mai bună ruta dintre aceste surse diferite. eBGP are distanta de 20, iBGP are distanta administrativa 200, rutele statice au distanta 1, rutele conectate au distanta 0, rutele inaccesibile(unreachable) au distanta 255, rutele învățate cu RIP au distanta administrativa 120, etc.

În cele mai multe cazuri, un ruter al unei Companii nu se va întâlni cu situația în care un prefix învățat cu BGP este de asemenea învățat cu IGP, sau învățat cu o ruta conectată (dimpotrivă, aceste probleme au loc mai des când sunt implementate VPN uri MPLS cu BGP/IGP redistribuit). Totuși aceste situații au loc, iar comanda **show ip bgp rib-failures** poate fi folositoare. Această comandă listează rutele pe care BGP le-a ales drept cele mai bune, dar pe care RTM nu le-a plasat în RIB(Routing Information Base), care este doar un alt nume pentru tabela de rutare IP.

BGP și comanda maximum-paths

La fel ca și protocoalele IGP, BGP suporta subcomanda, **maximum-paths number-of-paths** însă BGP, însă BGP utilizează o logică diferită de cea a IGP. Spre deosebire de protocoalele IGP, BGP trebuie să aleagă o rută și numai una drept cea mai bună rută pentru fiecare prefix.

Dacă algoritmul Best Path nu alege o rută până la pasul 8, rutele care sunt la egalitate până în acest moment vor fi adăugate în tabela de rutare, până la numărul maxim definit de subcomanda BGP **maximum-paths number-of-paths**.

Mărirea valorii lungimii AŞ_Path utilizând AŞ_Path prefixat(prepend).

Pasul 4 din algoritmul Best Path examinează lungimea atributului AŞ_Path, care pare a fi evidentă: pur și simplu aduni numărul de ASN-uri listate în AŞ_Path. Însă anumite trăsături BGP impactează de asemenea și modul de calculare al AŞ_Path. Unealta AŞ_Path prepered oferă inginerilor o modalitate de a mari AŞ_Path length adăugând ASN-uri la AŞ_Path, și totodată neinfluențând rolul atributului AŞ_Path de a preveni buclele. Dacă se dorește ca traficul să nu treacă tot printr-un anumit nod, se poate modifica AŞ_Path primit de la I3-1 să fie cu 2ASN-uri mai mare decât AŞ_Path primit de la I1-1.

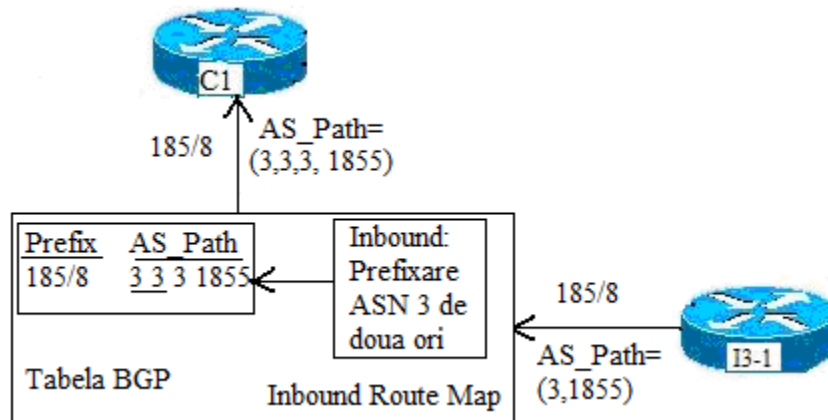


Figura 5. Prefixarea lui AS_Path cu doua ASN-uri

Configurarea în sine necesita doar puțină muncă adițională în comparație cu alte exemple. Așa cum se vede în figura, routerul Ci ar putea utiliza o inbound route map, utilizând comanda **set as-path prepend 3 3** pentru a adăuga 2 noi ASN-uri, făcând AŞ_Path length egală cu 4.

De reținut că atunci când utilizați AŞ_Path prefixat, nu prefixați orice ASN. BGP folosește în continuare AŞ_Path pentru evitarea buclelor, așadar cel mai bine este să utilizați un ASN aflat deja în AŞ_Path(de exemplu ASN 3 în cazul de mai sus) , sau ASN-ul local.

Deși este prezentată aici ca o unealtă de influențarea a rutelor cu exteriorul (outbound), AŞ_Path prefixat poate fi folosit pentru a influența și rutele interioare.

Influențarea rutelor interioare ale unei companii prin intermediul MED

O companie are un control destul de bun asupra rutelor sale IP cu exteriorul. Inginerii pot configura BGP să seteze și să reacționeze la Weight, Local_Pref, și AŞ_Path length, manipulând pe fiecare să aleagă o rută cu exteriorul diferită, sau un router prin care să forwadeze pachetele prin internet.

O companie are mai puțin control asupra rutelor interioare: rutele pentru pachetele ce se întorc spre companie. Mai întâi aceste rute exista pe routere pe care compania nu le deține. Chiar dacă un furnizor poate fi convins de inginerii companiei să facă în așa fel încât rutele spre companie să urmeze o anumită cale, probleme tehnice ar putea împiedica implementarea acestui proiect. În particular, dacă gama de adrese IP publice ale companiei este sumarizata, companiile care folosesc adrese din acea gama ar putea avea obiective rivale, așadar nicio politică nu se poate aplica pentru a influența cea mai bună ruta.

Exista totuși câteva modalități de a permite același control asupra ultimului hop ASN dintre ISP și compania client. Una dintre modalități se numește MED (Multi-Exit Discriminator).

Într-un design dual-homed, cel puțin două link-uri exista între ISP și Companie. Companie poate să anunțe Furnizorului o valoare (MED) care îi spune furnizorului care cale spre Companie este mai bună. Drept rezultat un ISP poate deosebi între multiplele puncte de ieșire din ISP spre acea Companie.

Deoarece MED permite ASN-ului Companiei să comunice doar ASN-ul vecin, inginerii utilizează de obicei MED când anunta spațiul de adrese IP publice ale unei Companii.

```
route-map set-med-to-I1-1 permit 10
  match ip address prefix only-public
  set metric 10
!
route-map set-med-to-I1-4 permit 10
  match ip address prefix only-public
  set metric 20
!
ip prefix-list only-public permit 128.107.0.0/19
router bgp 11
  neighbor 1.1.1.1 route-map set-med-I1-1 out
  neighbor 192.168.1.2 route-map set-med-I1-4 out
```

MED utilizeaza logica mai-mic-este-mai-bine.

Comanda este **set metric**, si nu **set med**, care nu exista.

Atat configuratia cat si outputul comenzii **show ip bgp** se refera la med ca si metrica. Chiar si comanda **show ip route** listeaza valoarea MED in paranteze drept metrica pentru ruta BGP.

IV. Încheiere

Din 1994, versiunea patru a protocolului este folosită în Internet, toate versiunile anterioare fiind considerate depășite. Cel mai important progres al versiunii 4 a fost suportul pentru CIDR și folosirea agregării rutelor pentru a reduce dimensiunea tabelelor de rutare. Din ianuarie 2006, BGPv4 este standardizat prin RFC 4271, care a trecut prin zeci de versiuni preliminare, bazate pe versiunea de BGP din RFC 1771. RFC 4271 a corectat unele erori, a clarificat ambiguitățile și a apropiat standardul de practicile curente din industrie.

Una din cele mai mari probleme întâmpinate de BGP, dar și de întreaga infrastructură Internet, provine de la creșterea rapidă a tabelii globale de rutare (care conține toate rutele din Internet). Pe măsură ce cerințele de memorie și de putere de procesare cresc, ruterele mai vechi nu mai fac față, utilitatea lor scăzând considerabil. Mult mai important, căutarea într-o tabelă de mari dimensiuni durează mai mult și provoacă instabilitate în cazul schimbărilor importante de topologie.

Până la sfârșitul anului 2001, creșterea tabelii de rutare era exponențială, ceea ce crea amenințarea unor probleme grave de conexiune. Pentru a evita acest lucru s-au luat o serie de măsuri între ISP-uri, printre care utilizarea CIDR și agregarea rutelor. Acest lucru a provocat o creștere liniară până în 2004, când creșterea a devenit din nou exponențială datorită cererii de conexiuni redundante din partea rețelelor de mici dimensiuni. În ianuarie 2009, tabela de rutare globală avea aproximativ 300.000 intrări.

BGP este un protocol foarte popular în ziua de azi; folosește algoritmul Best Path și oferă o mare flexibilitate în modul în care routerele aleg cea mai bună rută BGP.

Lista de abrevieri

ASN=Autonomous System Number
AS= Autonomous System
BGP=Border Gateway Protocol
IGP=Interior Gateway Protocol
IOS=Internetwork Operating System
IP=Internet Protocol
EGP=Exterior Gateway Protocol
MED=Multi-Exit Discriminator
RIB=Routing Information Base
RID=Router ID
RTM=Routing Table Manager

Bibliografie:

1. CCNP ROUTE 642-902, Wendell Odom, Cisco Press, Indianapolis, USA, Jan 2010

2. Cisco IOS CookBook - Publisher: O'Reilly Pub Date: December 01, 2006

3. Internet Routing Architectures, 2nd Edition, Sam Halabi with Danny

McPherson

4. Buiding reliable networks with BGP - Iljutsch Van Beijunum, September 2002

5. <http://www.wikipedia.org/>

6. http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094934.shtml