

Border Gateway Protocol (BGP)

BGP är ett routingprotokoll som används för transport av prefix över Internet. Det har även stöd för många adressfamiljer (AFI), vilket gör att det är en populär bärare av olika routingdata.

Det finns intern BGP (iBGP) och extern BGP (eBGP). En grupp routrar tillhörande samma organisation, eller funktion inom en organisation, grupperas enligt Autonomous Systems (AS). En BGP session mellan routrar med samma AS-nummer är extern BGP och sessioner mellan routrar med samma AS-nummer är intern BGP. En router kan enbart ha 1 AS-nummer. Extern BGP har administrative distance (AD) 20 och intern BGP har AD 200.

Routrar etablerar kommunikation med varandra via TCP 179. Routrar behöver alltså ej vara direktanslutna med varandra för att skapa grannskap. Routinginformation byts ut via Network Layer Reachability Information (NLRI) paket som innehåller prefix och dess path attributes.

Path Attributes

Pre-checks innan path selection algoritm

1. Next hop reachability (NEXT_HOP). Det måste finnas en väg till next-hop i routingtabllen för att routen ska installeras i RIB.
2. iBGP synchronization. Den här funktionen är avstängd per default i samtliga moderna implementationer per default. Men om aktiverad måste det finnas en exakt matchande route från ett IGP (eller statisk route) för att BGP routen ska installeras. i RIB Om OSPF används måste OSPF och BGP router IDs (RID) matcha för att routen ska anses synkroniserad.
3. Prebestpath cost-community. Om pre-bestpath point of insertion (POI) är med i ett prefix via extended community så kommer den att föredras över andra path attributes.

Path Attributes

Vid val mellan samma prefix finns en lång lista som BGP utvärderar från toppen till botten. Vid första path attribute som vinner så installeras den routen till RIB.

Vissa path attributes måste finnas och stödjas medan vissa inte måste det. Följande varianter av path attributes finns:

- **Well-Known Mandatory.** Måste stödjas i samtliga implementationer av BGP och måste vara med i varje NLRI. Exempel på detta är ORIGIN, AS_PATH och NEXT_HOP
 - **Well-Known Discretionary.** Måste stödjas i samtliga implementationer av BGP men måste ej vara med i samtliga NLRI. Exempel på detta är LOCAL_PREF.
 - **Optional Transitive Attributes.** Måste ej stödjas av samtliga BGP implementationer. Dessa värden följer med i NLRI paket till andra BGP-sessioner.
 - **Optional Non-Transitive Attributes.** Måste ej stödjas av samtliga BGP implementationer. Dessa värden stannar lokalt och skickas ej vidare via andra BGP sessioner.
1. **Weight.** Optional non-transitive. Cisco propriätär som konfigureras per router och är ej transitive, dvs att värdet stannar lokalt på routern och annonseras ej till andra. Högst weight vinner. Per default får varje lokalt injicerad route till BGP weight 32768.
 2. **Local-preference.** Well-Known Discretionary. Högre vinner, defaultvärde är 100 per prefix. Värdet sprider sig inom ett AS men ej utanför ett AS. Används ofta för att tilldela prefix, eller samtliga prefix från en viss granne, ett högre värde för att influera routingväg för utgående trafik.
 3. **Accumulated IGP (AIGP).** Tillåter BGP att lägga till IGP metric till BGP next-hop med remote AS metric värde.
 4. **Locally originated.** Rötter som har skapats lokalt vinner över andra. Lite redundant i Cisco-världen då weight slår denna.
 5. **AS-path.** Varje gång ett paket lämnar ett AS läggs AS-numret till i NLRI:n. Den väg med minst AS-nummer vinner. Det går att prependa, alltså lägga till sitt eget (eller annat) AS-nummer flera gånger för annonserade prefix.
 6. **Origin.** Mandatory, transitive, well known. Hackordningen är IGP > EGP > Unknown. IGP är för rötter som har injicerats i BGP med `network` kommandot, EGP är legacy och unknown är för rötter redistribuerade in i BGP.
 7. **Multi-exit discriminator (MED).** Optional, non-transitive. Används när en router har flera anslutningar till ett annat AS för att påverka utgående trafik.
 8. **Neighbor type.** eBGP föredras över iBGP.
 9. **IGP metric till BGP next-hop.** Bästa väg till next-hop vinner.
 10. **IGP cost-community.** Optional, non-transitive. Om IGP är punkten för pre-bestpath point of insertion (POI), vilket har annonserats via extended communities, kommer prefixet att vinna.
 11. **Multipath.**
 12. **Äldsta routen (endast eBGP).** Den route som är äldst enligt `show bgp afi safi x.x.x.x` vinner. Tanken är att den mest stabila routen är bäst.
 13. **Always compare RID.** Om always compare rid är konfigurerat kommer routen som kommer från granne med lägst router ID att vinna.
 14. **Kortast cluster-length (endast iBGP).** NLRI med kortast klusterlista (från route-reflektorer) vinner.
 15. **Lägst peer IP.** Sista valet. Den granne med lägst IP, enligt TCP-sessionerna, vinner.

Accumulated IGP (AIGP)

AIGP utvärdes efter Local Preference. Mellan AS kommer ett AS att skriva AIGP värdet till en NLRI genom att kopiera IGP metricen.

AIGP läggs på ett prefix genom en route-map. I exemplet nedan används network-statement för att injicera nätverket i BGP. Det går att definiera AIGP metric manuellt men i exemplet nedanför så tas metricen från den IGP som används. AIGP behövs aktiveras per granne.

```
route-map AKTIVERA_AIGP permit 10
  set aigp-metric igp-metric

router bgp 1337
  address-family ipv4
    network 1.3.3.7 mask 255.255.255.255 route-map AKTIVERA_AIGP
  neighbor 7.3.3.1 aigp
```

Avaktivera eller ändra hänsyn till PAs

Det går att konfigurera BGP så att den lokala uträkningen struntar i vissa path attributes (PA) för att hitta bästa rutten eller ändrar hur den använder PA:t. Här finns några utvalda kommandon:

```
! Aktivera kontroll av router ID
bgp bestpath compare-routerid

! Strunta i AS-path
bgp bestpath as-path ignore

! Strunta i IGP metric för next-hop
bgp bestpath igp-metric ignore

! Strunta i cost-community
bgp bestpath cost-community ignore

! Jämför alltid MED
bgp always-compare-med

! Finns ingen MED, anse den som sämst
bgp bestpath med missing-as-worst
```

Graceful Restart

BGP kallar Non Stop Forwarding (NSF) för Graceful Restart (GR). GRs funktion är att tillåta forwarding och upprätthållande av grannskap under en switchover. BGP gör detta genom att använda en End-of-RIB (EoR) markerare. EoR signalerar att BGP UPDATE meddelanden kommer att avta. Istället för att vänta på fler uppdatering kan en BGP router direkt köra best-path val när samtliga peers har indikerat EoR.

Per default har IOS-XE GR avstängt. GR aktiveras i BGP-processen med kommandot `bgp graceful-restart`. Vid aktivering måste samtliga grannar tas ned hårt, ej soft reset. `clear bgp ipv4 unicast *`. GR kan avaktiveras mot en specifik granne med `neighbor 1.3.3.7 ha-mode graceful-restart disable`.

BGP GR specificerar två timers, restart-time och stalepath-time. restart-time börjar räkna ned när en TCP session misslyckas mellan peers under en RP (route-processor) switchover. Timern kontrollerar hur länge BGP väntar på ett BGP OPEN meddelande från grannen som utför RP switchover. När en router påbörjar RP switchover så markeras samtliga rötter från den grannen som "stale" men trafik kan fortfarande forwardas genom den. stalepath-time kontrollerar hur länge trafik kan skyfflas till grannen innan de rensas. Vanligtvis är stalepath-time-timern längre än restart-time timern. Värdena konfigureras med `bgp graceful-restart restart-time x` och `bgp graceful-restart stalepath-time x` där x är sekunder.

RPKI

RPKI (Resource Public Key Infrastructure) är en säkerhetsfunktion för BGP-annonseringar. Routrar som kollar efter RPKI kontrollerar att ett ASN får annonsera ett nätblock. Stämmer inte AS-numret med nätblocker kommer inte routen att installeras. Kombinationen nät och ASN inom RPKI kallas för ROA (Route Origin Authorizations).

Det finns olika varianter av RPKI. Hosted RPKI innebär att någon annan, ex. LIR, ansvarar för certifikatshanteringen. ROA administreras via leverantörens portal.

Kontroll av RPKI går att utföra här: <https://rpki.cloudflare.com/?view=validator>

Additional Paths (add-path)

Per default annonserar BGP enbart den bästa vägen till ett prefix. Det går att annonsera samtliga vägar från en Route-Reflector. Se exempelkonfiguration:

```
router bgp 1337
  template peer-policy IPV4
    advertise diverse-path backup
  address-family ipv4
    bgp additional-paths select backup
    bgp additional-paths install
```

För icke RRs går annonsering och/eller mottagning av additional paths att aktivera och avaktivera med kommandon i exemplet nedanför. Det går att aktivera för en hel adressfamilj eller per granne/peer-template.

```
router bgp 1337
  address-family ipv4
    maximum-paths ibgp 4
    bgp additional-paths receive
    bgp additional-paths send
    bgp additional-paths select best 3
    neighbor 1.3.3.7 additional-paths disable
    neighbor 7.3.3.1 advertise additional-paths best 2
```

För att förtydliga är add-path enbart aktuellt för iBGP.

RT-Filter

I en iBGP topologi används oftast route-reflectors (RR). En RR kommer att annonsera samtliga NLRler till samtliga RR-klienter. En router kommer ex. enbart att installera VPNv4 rötter om Route Targets (RT) för en viss VPNv4 route i viss VRF finns definierad lokalt. Det gör att en router som inte har samtliga RTs installera, som kanske inte heller ska eller behöver ha alla RTs installerade, kommer att slänga många rötter. Med RT-filter kan man filtrera vad som ska annonseras från en RR.

```
! Aktivering av RT-filter
router bgp 1337
  address-family rtfiler unicast
    neighbor 1.3.3.7 activate
    neighbor 1.3.3.7 send-community extended
```

Verifiering av rtfiler med `show bgp rtfiler unicast all { detail }`.

DMZ Link Bandwidth

DMZ Link bandwidth är en teknik som tillåter unequal cost load sharing i BGP. DMZ länkar är eBGP länkar som ansluter olika AS, även kända som transit links. Information om bandbredd skickas inom ett extended community som är non-transitive.

```
router bgp 1337
address-family ipv4
  bgp dmzlink-bw
  maximum-paths ibgp 4
  neighbor 7.3.3.1 dmzlink-bw
```

Multicast Source Discovery Protocol (MSDP)

MSDP är en teknik för att stödja multicast-transport över flera AS. Grannar ska vara aktiverade i adressfamiljerna `address-family ipv4 multicast` och/eller `address-family ipv6 multicast`. Inom ett AS måste multicast såklart vara konfigurerat, PIM, RPs und so weiter. Protokoll såsom Bootstrap Router och AutoRP ska ej annonseras till andra AS.

I MSDP är tanken att RPs i ett AS ansluter mot RPs i andra AS. Konfigureras med `ip msdp peer 1.3.3.7 connect-source Loopback0 remote-as 1337`. Verifiering med `show ip msdp summary`. Specifikt peer kan kontrolleras med `show ip msdp peer 1.3.3.7`. TCP 639 används för MSDP-kommunikationen. MD5 autentisering konfigureras med `ip msdp password peer 1.3.3.7 LOOOSENORD`. Timers konfigureras med `ip msdp keepalive 1.3.3.7 15 35`, där 15 är keep-alive och 35 är dead-timer.

Autentisering

IOS-XE verkar ha en gräns på 25 tecken för password authentication.

Revision #1

Created 26 April 2025 12:03:03 by Jehrlander

Updated 26 April 2025 12:03:08 by Jehrlander