

## Border Gateway Protocol RFC 1771

BGP پروتکلی پیچیده است که در هسته اینترنت جهت Routing از آن استفاده می شود. وظیفه این پروتکل اتصال و Routing بین Autonomous System ها و شبکه های بزرگ IP است. بطور مثال ارتش برای اتصال شبکه های نیروهای زمینی، هوایی و دریایی خود - که هر کدام شبکه ای بزرگ IP دارند - از BGP بعنوان Exterior Routing Protocol استفاده میکند.

BGP حامل اطلاعات Autonomous هاست تا از این طریق شبکه های IP (در درون Autonomous) را پیدا کند. از مشخصات کلی این پروتکل میتوان به موارد زیر اشاره کرد:

- یک پروتکل Path Vector است تا Distance Vector.
- VLSM و CIDR را پشتیبانی میکند.
- ارتباط با همسایگان را روی پورت TCP 179 برقرار میکند و با Keep-alive آن را نگه میدارد.
- Routing Table مستقل خود را دارد.
- Metric در BGP، Attribute نامیده می شود.
- نسبت به پروتکل های Interior کاملاً متفاوت بوده و برای ارتباط بین AS هاست. (بعبارتی Interior Routing Protocol است.)
- انعطاف پذیر (Flexibility) بوده و با Policy-based Routing حرفه ای همراه است. به هر حال Routing در اینترنت بصورت Hop-by-hop است و سیاست های شما در طول مسیر بوسیله روترهای میان راه - که کنترلی رویشان ندارید - ممکن است بکار گرفته نشود.
- همانند RFC 1819 جهت اختصاص IP برای استفاده Private، AS Number نیز طبق RFC 1930 از 64512 به انتها (یعنی تا 65535) برای مصارف Private در نظر گرفته شده است.

RFC 1771 (March 1995)

Interdomain Routing Protocol – Exterior Routing Protocol

Path-Vector

Uses TCP port 179.

BGP4 Implements CIDR (CIDR RFC 1519).

Classless supported (VLSM), Auto-Summary (IANA Classful), and Manual.

Administrative Distance = 20 (eBGP) and 200 (iBGP).

Attributes (as metric for routing decisions).

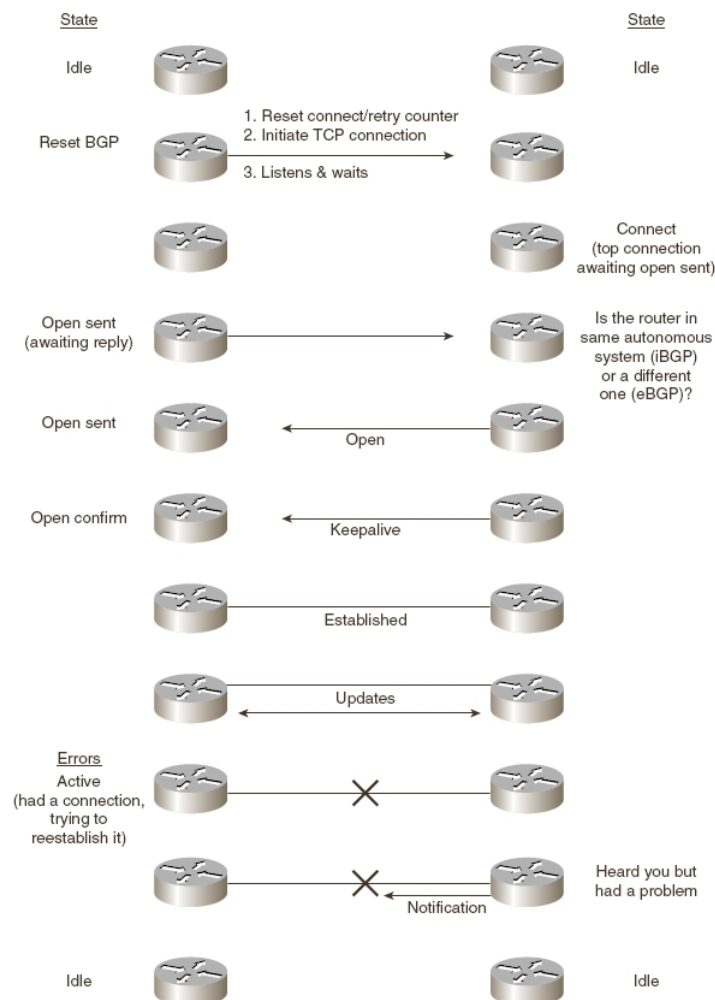
Authentication supported.

Partial updates supported.

## پیام های BGP

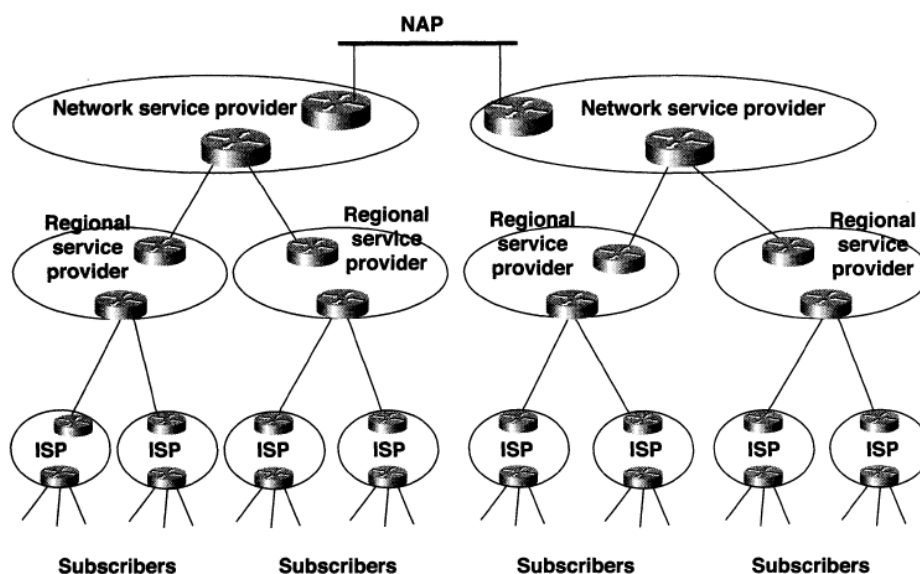
در BGP چهار گونه پیام رد و بدل میشود:

- Open Messages: برای ایجاد Connection بین Peer ها استفاده میشود.
- Keepalives: برای نگهداری Connection بین Peer ها بصورت تناوبی استفاده میشود. اگر Timer آن صفر ست شده باشد، Keepalive ارسال نمیشود.
- Update Messages: حاوی اطلاعات مسیر و Attribute های مسیر است. به ازای هر مسیر یک Update ارسال میگردد.
- Notification: برای اطلاع همسایه از مشکل (Error) و در زمان قطع ارتباط ارسال میشود.



## معماری اینترنت:

مشترکین به ISP منطقه خود، وصل میشوند که به آن، Point of Presence - POP گفته میشود. ISP های کوچکتر به ISP ای بزرگتر متصل میشوند مثل اتصال ISP های ایران به مخابرات که Regional Service Provider نامیده میشود، از سوی دیگر ISP های منطقه ای به Network Service Provider ها متصل میشوند (نظیر Sprint، C&W، MCI/WorldCom (UUNET) و غیره) که به هر کدام از این مراتب به ترتیب، Tier 3، Tier 2، و Tier 1 Provider اطلاق میشوند. در بالاترین مرحله NAP قرار گرفته است.



NAP یا Network Access Point، Provider ها را با سرعت بالا (مثلا در Ethernet Switch) به یکدیگر متصل میکند تا Route و Data را بین خود ردوبدل کنند. مثل CIX (Commercial Internet Exchange) یا FIX (Federal Internet Exchange) که نمونه هایی از NAP های اولیه اند که NSFnet (National Science Foundation) در 1994 براساس آن تجربه پیاده سازی NAP های خود را اجرا کرد.

در داخل NAP، یک یا چند Unix Server بعنوان Route Server معرفی شده و Service Provider ها تنها با Route Server، Peer شده و به زبان BGP صحبت میکنند. Database استفاده شده در این Route Server منطبق با RIPE-181 است. Route Server هیچ گونه بسته ای را Forward نمیکند و تنها Routing را تنظیم میکند. (از آنجا که Next-hop درون همان شبکه قابل دسترسی است.)

بر اساس توافق Route های 19/ به بالا (یعنی بزرگتر یا مساوی 19/ - نظیر 16/ و غیره) در Provider ها فیلتر نمیشوند، از آنجا که یک آدرس Provider همیشه به سمت Provider ارسال میشود برای داشتن آدرس خود باید درخواست Provider Independent ارائه دهیم که طبق ضوابط خاص و در صورتیکه Provider قادر به فراهم آوردن آدرس های مورد نیاز ما نباشد، بصورت حداقل آدرس مورد نیاز فعلی، به ما اختصاص داده میشود.

## eBGP و iBGP

ارتباط بین دو Peer در BGP بوسیله یک TCP Connection برقرار میشود. دو همسایه در دو AS متفاوت از طریق eBGP یا External BGP و در صورتیکه داخل یک AS باشند، ارتباط دو همسایه از طریق iBGP یا Internal BGP برقرار میشود.

برای اینکه Route ها بین دو همسایه داخل یک AS بهتر قابل درک باشد، استفاده از iBGP پیشنهاد میشود در غیر این صورت اطلاعات BGP درون یک IGP Redistribute شده و در سمت دیگر از IGP درون BGP مجدداً Redistribute میشود. استفاده از iBGP (در مقابل Redistribution بواسطه یک IGP) بهتر است چرا که Attribute ها بین دو همسایه با صحت منتقل میشوند. استفاده از iBGP بخصوص در Transit AS ها پیشنهاد میشود. ارتباط میان همسایگان iBGP باید Full-Mesh باشد چون تغییرات دریافتی از یک همسایه از طریق iBGP بواسطه قانون Split-Horizon برای بقیه همسایه ها ارسال نخواهد شد.

## BGP Router ID

در BGP هر روتر باید یک RID داشته باشد. Router ID بر اساس بالاترین IP در روتر انتخاب میشود مگر آنکه از طریق تنظیمات RID را برای روتر مشخص کنیم.

در صورتیکه از آدرس Loopback استفاده کنیم از این آدرس بعنوان RID استفاده خواهد شد.

## Route Dampening

مکانیزمی است که ناپایداری حاصل از Route Flapping را به حداقل میرساند. با شمارش تعداد دفعات موفق نبودن یک Route و احتساب جریمه برای آن، BGP مسیرها را به دو دسته خوش و ناخوش تقسیم میکند. اگر مسیری ill-behaved یا ناخوش شناخته شد، BGP آنرا به بقیه advertise نمیکند، مگر پس از مدتی که از صحت و سالم بودن آن مطمئن شد.

# Cisco in Persian

## Attribute های مسیر

Attribute ها متغیرهایی هستند که برای انتخاب بهترین مسیر توسط BGP بکار گرفته و استفاده میشوند و اطلاعاتی را که بر اساس آن بهترین مسیر انتخاب میشود Attribute مینامیم.

هر Update حاوی Route و Attribute های آن است، BGP برای انتخاب بین Route ها و انتخاب مسیر Attribute ها را باهم می سنجد.

بر اساس RFC، Attribute ها به دو دسته تقسیم میشوند که آن دو نیز از دو بخش تشکیل شده اند:

### • Well-known

- **Mandatory**: این Attribute همیشه همراه Route است، همه آنرا میشناسند و در هر Update ی وجود دارد.
- **Discretionary**: اگر بکارگرفته شده باشد، بقیه روترها باید آنرا بکار میگیرند. مثل Atomic Aggregate و Local Preference

## Cisco in Persian

### • Optional

- **Transitive**: اگر قابل تشخیص نبودند، تغییر نداده آنرا به بقیه ارسال میکند. مثل Community و Aggregator که در تعیین مسیر تأثیری ندارند.
- **Non-transitive**: در طول مسیر از آنها استفاده نمیشود و فقط بین دو Peer است. مثل MED، Originator ID و Cluster List.

*Path Attributes\**

| <b>Attribute</b>         | <b>Class</b>             |
|--------------------------|--------------------------|
| ORIGIN                   | Well-known mandatory     |
| AS_PATH                  | Well-known mandatory     |
| NEXT_HOP                 | Well-known mandatory     |
| LOCAL_PREF               | Well-known discretionary |
| ATOMIC_AGGREGATE         | Well-known discretionary |
| AGGREGATOR               | Optional transitive      |
| COMMUNITY                | Optional transitive      |
| MULTI_EXIT_DISC<br>(MED) | Optional nontransitive   |
| ORIGINATOR_ID            | Optional nontransitive   |
| CLUSTER_LIST             | Optional nontransitive   |

Cisco in Persian

**(Well-known – Mandatory) ORIGIN**

مشخص کننده نوع Update است. وقتی چند مسیر به یک مقصد وجود داشته باشد، Origin یکی از فاکتورهای مهم در انتخاب مسیر است:

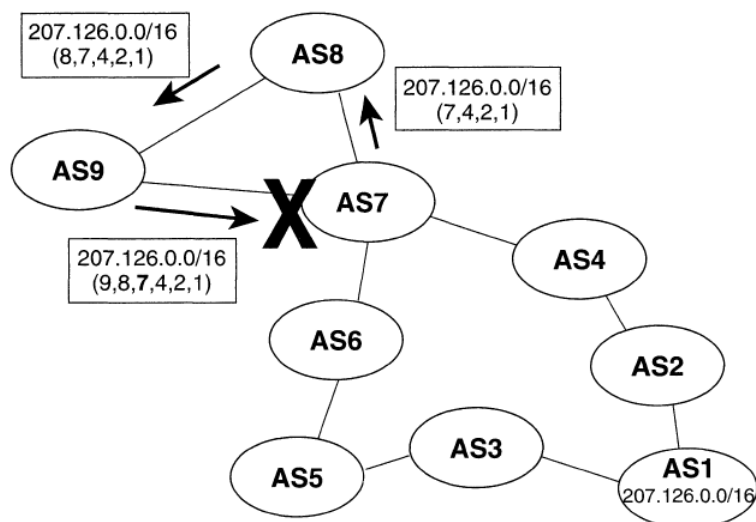
- IGP: توسط یک IGP در داخل AS، BGP مسیر را یاد گرفته است. (بطور مثال با دستور Network، شبکه به BGP معرفی شده است).
- EGP: بواسطه یک پروتکل EGP این مسیر را یاد گرفته است.
- Incomplete: مشخص نیست از IGP یا EGP گرفته شده مثل Route ی که درون BGP، Redistribute شده باشد.

**(Well-known – Mandatory) AS\_PATH**

حاوی لیست AS های بین راه است. وقتی یک Route توسط BGP به بیرون Advertise میشود، BGP، AS خود را به این لیست اضافه میکند. بدین طریق از وقوع Loop جلوگیری شده و مسیری که خود AS درون آن ذکر شده از AS نباید عبور کند.

## Cisco in Persian

*If a BGP Router Sees Its Own AS Number in the AS\_PATH of a Route from Another AS, It Rejects the Update*

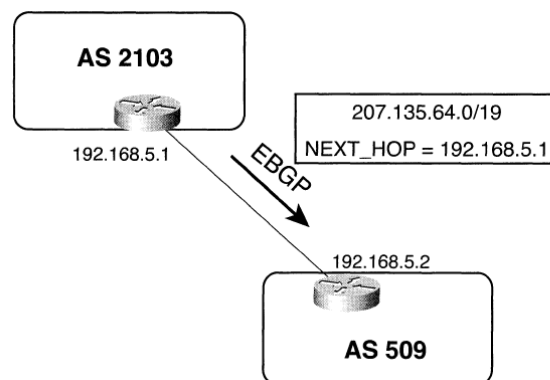




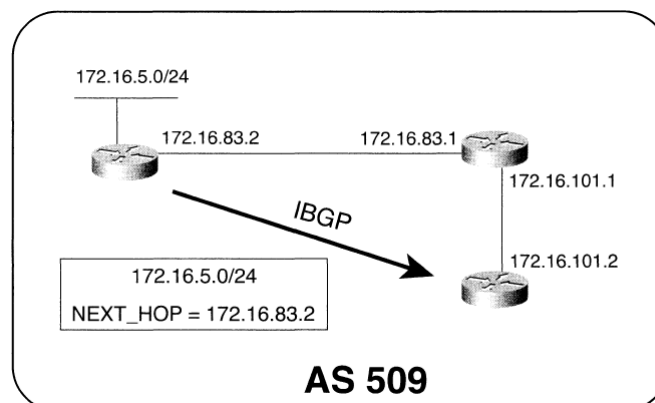
**(Well-known – Mandatory) NEXT\_HOP**

آدرس روتر بعدی در امتداد مسیر است و لزوماً آدرس همسایه ای که این Route را Advertise کرده نیست.

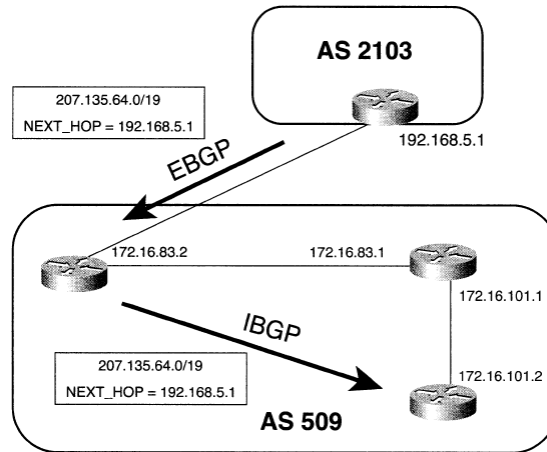
- اگر دو همسایه در دو AS متفاوت باشند NEXT\_HOP همیشه، آدرس Interface همسایه ای است که مسیر را تبلیغ کرده است.



- اگر دو همسایه در یک AS باشند، NEXT\_HOP آدرس روتری است که همسایه Route را از او یاد گرفته است.



- اگر دو همسایه داخل یک AS باشند و همسایه مسیری به خارج را Advertise کند، آدرس NEXT\_HOP External Peer است. که در شکل صفحه بعد نشان داده شده است.



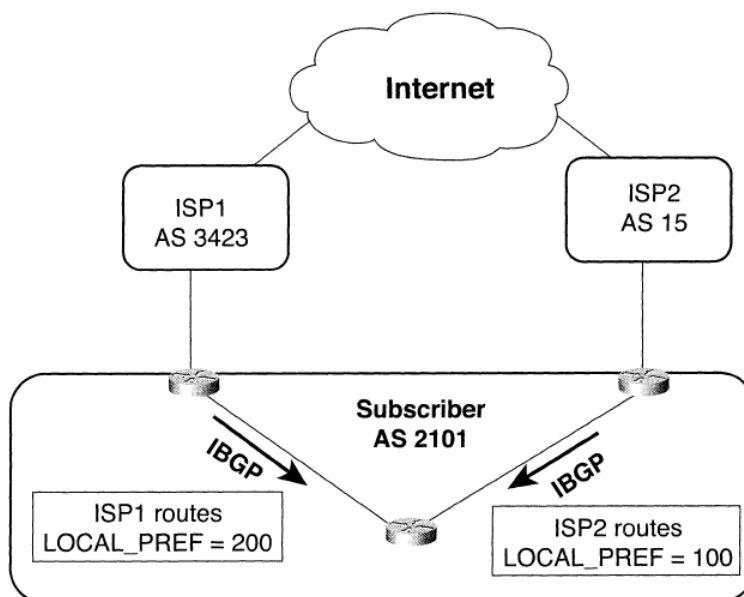
با توجه به شکل، برای یک روتر داخل AS، جهت ورود به شبکه Advertise شده، NEXT\_HOP روتری در AS 2103 است که روتر داخلی ممکن است مسیری به آن روتر نداشته باشند و در نتیجه به AS مذکور وصل نشوند.

برای این که Border Router آدرس NEXT\_HOP را تغییر داده و آدرس خود را قرار دهد به تنظیمات next-hop-self نیاز داریم.

## Cisco in Persian

### (Well-known – Discretionary) LOCAL\_PREF

بین iBGP ها استفاده میشود و به AS دیگری اعلام نمیگردد. اگر یک iBGP چند مسیر برای مقصدی داشت بر اساس Local Preference بالاتر، مسیر را انتخاب میکند. (Default=100)

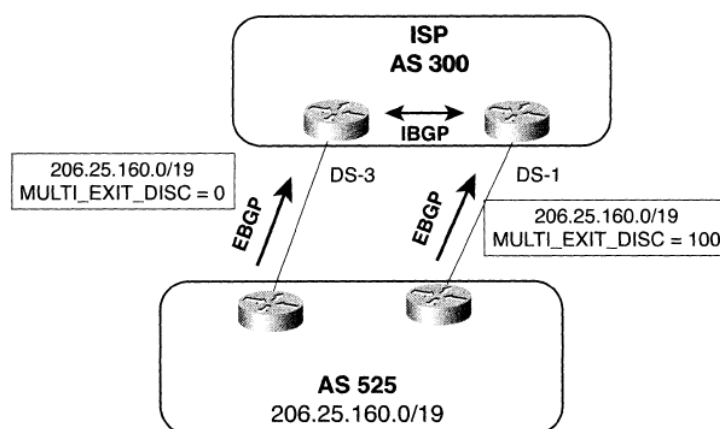


**(Optional – Nontransitive) MULTI\_EXIT\_DISC**

از آنجا که Local Preference روی ترافیک خروجی از AS تاثیرگذار است، برای ترافیک ورودی و ایجاد اولویت از MED استفاده میکنیم.

در واقع برای AS مقابل خود Preferred Ingress Point (ورودی برگزیده) را مشخص میکنیم. اما بر خلاف Local Preference که بالاترین عدد انتخاب میشود، از آنجا که MED حکم Metric را دارد، کوچکترین MED بعنوان بهترین انتخاب میشود.

*The Lower MED Associated with Routes Passed Over the DS-3 Link Causes the ISP to Prefer This Link*



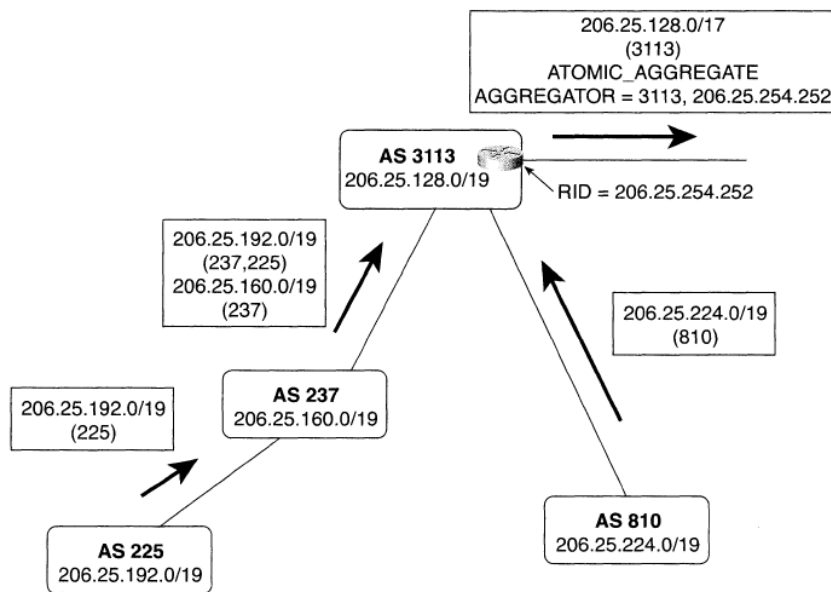
باید توجه داشت که MED برای AS سومی که به ISP (در شکل بالا) متصل است ارسال نمیشود و بین دو AS بصورت Directly connected در انتخاب لینک ورودی به AS 525 تاثیرگذار خواهد بود.

(Well-known – Discretionary) ATOMIC\_AGGREGATE  
(Optional – Transitive) AGGREGATOR

آدرس هایی از AS های Downstream را با هم Aggregate کرده ایم، با توجه به اینکه برای Advertise کردن این شبکه Aggregate شده نمیتوان AS\_PATH مشخصی ارائه داد، AS\_PATH از جایی که Aggregate شده، معرفی میگردد. از آنجا که به بقیه باید اعلام کنیم که این آدرس Aggregate شده است، از ATOMIC\_AGGREGATE استفاده میکنیم که به روتر های Downstream، اخطار دهیم که مسیر ذکر شده، خلاصه شده است و شامل کل AS\_PATH نیست. این Attribute به Route چسبیده و هر جا که روتری آنرا Advertise میکند همراه با ATOMIC\_AGGREGATE attribute اعلام خواهد شد.

AS و IP Aggregate کننده را بعنوان AGGREGATOR میتوانیم مشخص کنیم.

*The ATOMIC\_AGGREGATE Attribute Indicates That a Loss of Path Information Has Occurred, and the AGGREGATOR Attribute Indicates Where the Aggregation Occurred*



## AS\_SET

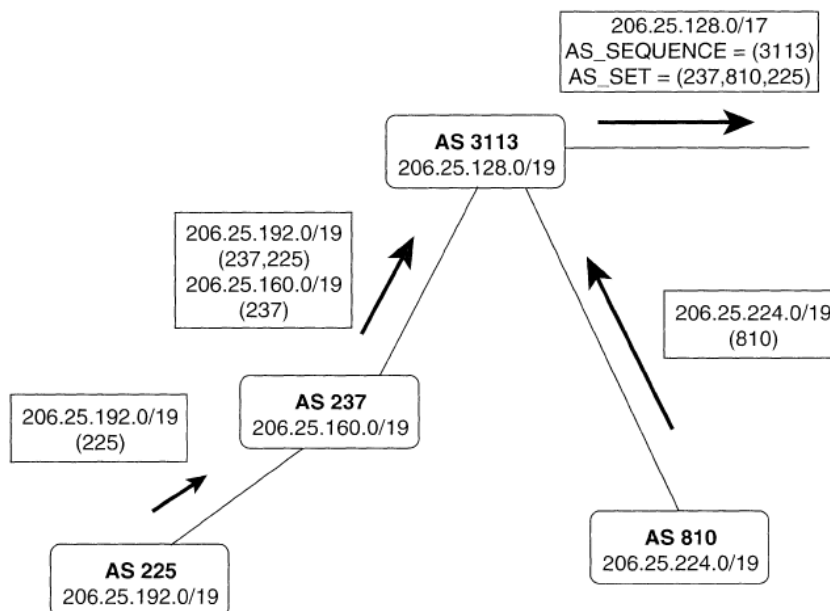
در مورد AS\_PATH، چهار نوع فرمت ارائه میشود که دو نوع آن مربوط به Confederation است و دو نوع دیگر در زیر آمده است:

- **AS\_SEQUENCE**: مسیر را به "ترتیب" AS های طی شده مشخص میکند.
- **AS\_SET**: لیست AS های طی شده بصورت نامرتب است. (ترتیب قابل ارائه نبوده)

همان طور که اشاره شد، از فواید مهم AS\_PATH جلوگیری از Loop است و ترتیب AS ها اصلا در این رابطه اهمیتی ندارد. وقتی Aggregate صورت میگیرد، توسط AS\_SET لیست تمام AS ها مثل AS\_SEQUENCE در AS\_PATH قرار میگیرد که AS های دیگر میفهمند که ممکن است لیست مرتب نباشد.

وقتی AS\_SET همراه با AS\_PATH ذکر میشود، دیگر از ATOMIC\_AGGREGATE همراه Route نمیتوان استفاده کرد. استفاده از AS\_SET بجای ATOMIC\_AGGREGATE برای مقابله با Loop پیشنهاد میشود.

*Including an AS\_SET in the AS\_PATH of an Aggregate Route Restores the Loop Avoidance That Was Lost in the Aggregation*



**(Optional – Transitive) COMMUNITY**

برای ساده کردن اعمال یک Policy بکار برده میشود. بطور مثال تمام Customer های یک ISP با یک Community مشخص شده و تنظیمات تکراری از قبیل MED و LOCAL\_PREF برای Community اعمال میگردد تا برای هر کدام از آنها. Community - 70 در AS 326 بصورت 326:70 مشخص میگردد. Community های معروف و تعریف شده شامل موارد زیرند:

- **INETRNET**: هر مسیر بصورت پیش فرض عضو این انجمن است و Advertise میشود.
- **NO\_EXPORT – 0xFFFFFFFF01**: مسیرهایی که با این مقدار دریافت میشوند به eBGP، Advertise نمیشود.
- **NO\_ADVERTISE – 0xFFFFFFFF02**: مسیرهایی که با این مقدار دریافت میشوند، هیچگاه advertise نخواهند شد. چه به eBGP و چه به iBGP.
- **LOCAL\_AS – 0xFFFFFFFF03**: مسیرهایی که با این مقدار دریافت میشوند به eBGP یا یک Confederation Peer در AS دیگر اعلام نمیشوند.

**(Optional – Nontransitive) ORIGINATOR\_ID**  
**(Optional – Nontransitive) CLUSTER\_LIST**

در Route Reflector ها استفاده شده و ORIGINATOR\_ID توسط Route Reflector استفاده میشود و مشخص کننده فرستنده Update در داخل AS است. اگر روتر Originator، Update ی با این ID دریافت کرد به Loop شدن شبکه پی میبرد.

CLUSTER\_LIST مشخص کننده، Route Reflection هایی است که طی شده و اگر در Update دریافتی Cluster ID خود Reflector دیده شود متوجه Loop خواهیم شد.

**(Cisco-Defined) Weight**

این attribute در داخل روتر تنها برای ایجاد اولویت بکار میرود و به بقیه ارسال نمیگردد. عددی است بین 0 تا 65535 که به Route اطلاق میشود و هر چه بزرگتر باشد، مسیر اولویت بالاتری دارد. مسیر های دریافتی بصورت پیش فرض وزنی برابر صفر دارند، مسیرهایی که خود روتر ایجاد کرده (در واقع شبکه های Directly-Connected) برایش 32768 وزن خواهد داشت پس بالاترین اولویت را دارند.

### انتخاب مسیر

وقتی چند مسیر برای یک مقصد (Destination) وجود داشته باشد بترتیب مقادیر زیر کارساز خواهند بود:

Weight (Highest)

Local Preference (Highest)

مسیری که خودت Advertise کردی را انتخاب کن.

AS PATH (Lowest)

ORIGIN (Lowest)

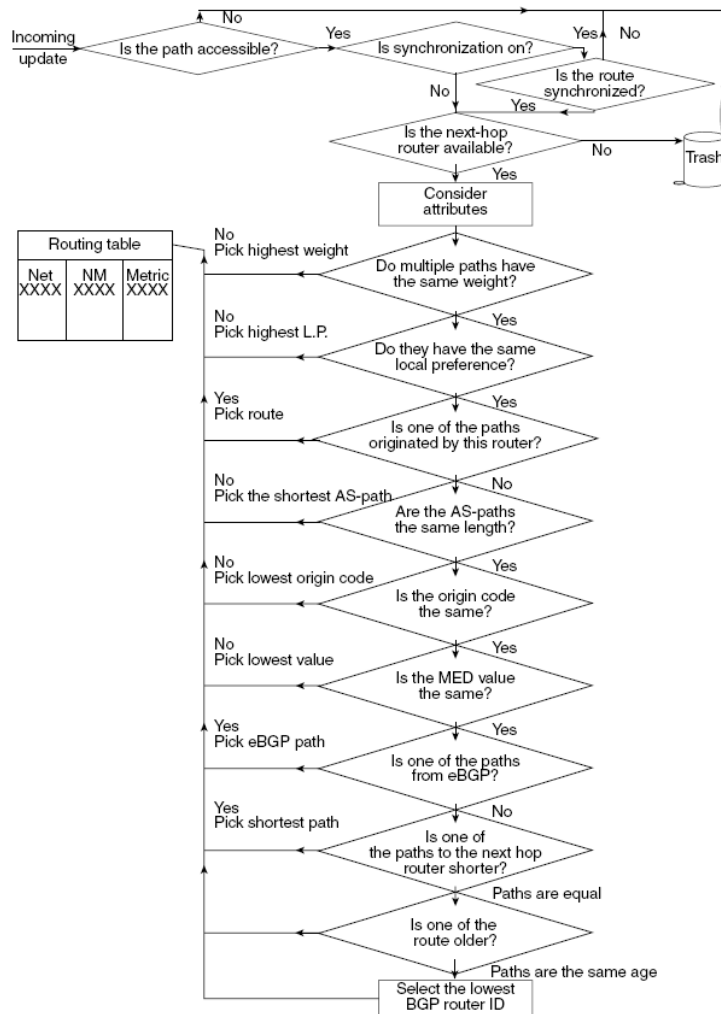
MED (Lowest)

مسیر eBGP نسبت به iBGP بهتر است.

بهترین مسیر IGP با بهترین Metric را انتخاب کن.

قدیمیترین مسیر را انتخاب کن.

کوچکترین RID را انتخاب کن.



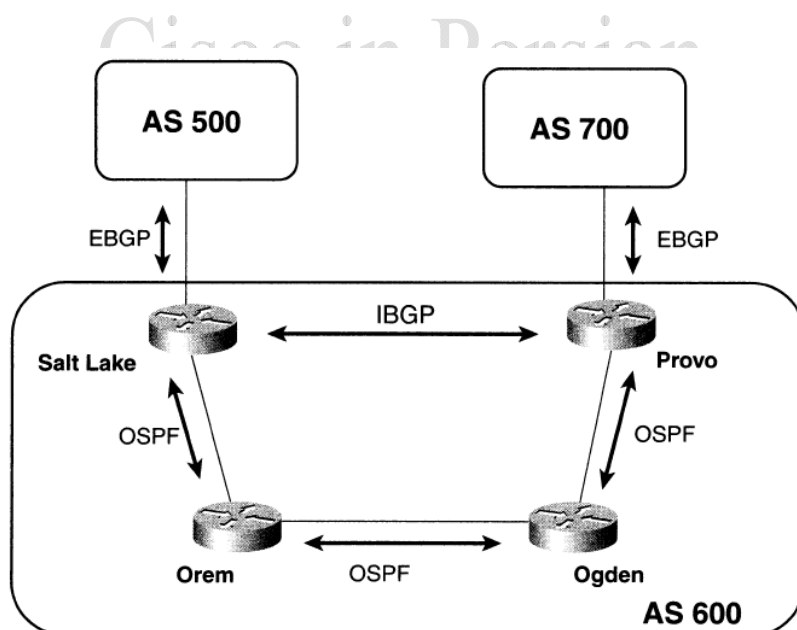
## Synchroni zati on

در شبکه ای که تماماً از BGP است (مثل یک Service Provider) چرا فقط از iBGP بعنوان IGP بهره نبریم؟

از آنجا که iBGP برای ایجاد ارتباط با همسایگانش نیازمند به ارتباط Full-Mesh است، در شبکه ای با تعداد روتر بالا نمیتوان این کار را میسر کرد. به هر حال استفاده از یک IGP توانمند برای ارتباط BGP درون AS پیشنهاد میشود تا ارتباط میان iBGP ها نیز در شبکه برقرار گردد. در ضمن یک بسته که از یک eBGP وارد شده تا بواسطه eBGP دیگر از Transit AS خارج شود نیز باید بوسیله IGP بدرستی Route شود (بخاطر NEXT\_HOP).

هدف از داشتن پیش نیازی نظیر Full-Mesh بودن iBGP ها، جلوگیری از وقوع Loop بین آنهاست. در ارتباط iBGP از AS\_PATH که یک مکانیزم تدافعی برای جلوگیری از Loop است نمیتواند بهره برد زیرا که شماره AS وقتی در AS\_PATH ثبت میشود که از AS توسط eBGP به خارج advertise میشود.

در این شکل، از iBGP بعنوان IGP استفاده نمیشود و OSPF برای این کار بهره برداری شده است.



با توجه به اینکه Provo یک AS بشماره 500 را به AS700 اطلاع میدهد، بسته هایی از 700 به مقصد 500 بدست Provo میرسند که با توجه به مسیر Provo به Salt lake از طریق OSPF به دست Ogden میرسند و در اینجاست که Ogden نمیداند که Salt lake به AS500 مربوط است زیرا OSPF در این باره چیزی نمیداند و BGP در OSPF، Redistribute نشده است.



قانون Synchronization میگوید که اگر یک Route از یک iBGP گرفته شد برای ثبت IGP و تبلیغ آن در BGP، ابتدا Route باید از طریق IGP نیز شناخته شده باشد.

هدف این است که:

از آنجا که تبلیغ یک شبکه بمعنای "تضمین در صحت رسیدن بسته ها بدست آن شبکه است" پس از وجود مسیر مطمئن به روتر NEXT\_HOP در درون AS (IGP) نیز باید مطمئن شویم.

این قانون برای شبکه های Transit که BGP درون IGP Redistribute میشود مفید است و در سیسکو بصورت Default، فعال است (در IOS های جدید دیگر Default نیست). هر چند که از طریق configuration قابلیت غیرفعال کردن آنرا نیز داریم.

در سه حالت میتوانیم Synchronization را غیر فعال کنیم که در بسیاری از موارد این کار به سود ما خواهد بود:

- اگر همه Router ها در AS، BGP میفهمند.
- اگر روترها درون AS، تماما Mesh باشند.
- وقتی AS، Transit نباشد.

## Cisco in Persian

چرا از Synchronization استفاده کنیم؟

- برای جلوگیری از ارسال ترافیک اضافی و Unreachable
- برای جلوگیری از Black-hole در شبکه ها و ارتباط بین AS ها.

باید توجه داشت که با فعال بودن Synchronization ممکن است یک route یادگیری شده از همسایه در IGP وارد نگردد (در زمانی که از IGP استفاده نمیشود) و مشکلاتی برای شبکه Partial-Mesh بواسطه برقراری Synchronization ایجاد شود.

## Route Reflector

قانون BGP Split Horizon میگوید: گرچه روتر های BGP نیازی ندارند که -Directly Connected باشند (مستقیماً روی یک لایه دو به هم متصل باشند) اما ارتباط همسایگی آنها بصورت TCP باید Full-Mesh باشد.

چرا که طبق این قانون: یک Update که از iBGP گرفته شده را به یک همسایه iBGP دیگر نباید فرستاد. آنچه که از iBGP گرفته شود فقط قابل انتقال به همسایه ای eBGP است.

Full-Mesh کردن ارتباط iBGP ها در شبکه های بزرگ نه کار ساده ایست و نه عاقلانه. برای رشد شبکه و رشد پذیری باید از Route-Reflector استفاده کرد که قانون Split-Horizon را بنوعی نقض میکند.

Route-Reflector Update های دریافتی را به Client ها ارسال میکند، همچنین Update های یک Client را نیز به بقیه Client ها و به همسایگان منتقل میکند. (درون Autonomous System)

بدین صورت دیگر نیازی به Peer بودن Client ها بصورت Full-Mesh نیست. Route-Reflector و Client ها با هم یک Cluster را تشکیل میدهند. AS از چندین Cluster میتواند تشکیل شود و هر Cluster حداقل یک Route-Reflector دارد. ارتباط بین Route-Reflector ها باید Full-Mesh باشد.

Route-Reflector در زمان ارسال update از یک Attribute به نام Originator-ID بهره میبرد. (از Router-ID استفاده میکند). برای redundancy و اطمینان بیشتر میتوان از دو یا چند Route-Reflector استفاده کرد که Reflector فرستنده یک update از Cluster-ID attribute استفاده میکند.

### مزایای استفاده از Route-Reflector:

- رشد پذیری و همچنین استفاده از مدل سلسله مراتبی در طراحی شبکه.
- صرفه جوئی در مصرف حافظه، CPU، TCP Connection و ترافیک شبکه.
- سرعت Convergence بالاتر در شبکه.

## BGP Process

بعد از ایجاد هر تغییر در تنظیمات BGP، ارتباط TCP باید reset شود تا تغییرات اعمال گردند. برای این کار نیازی به ریست کردن کل BGP Process نیست و به ازای Neighbor صورت میگیرد.

همچنین میتوان به روتر گفت که update ها را ارسال کند بدون اینکه session را قطع کرده و دوباره ایجاد کند. این soft update بصورت in ویا out میتواند باشد که به ارسال مجدد update های ورودی و یا خروجی اشاره میکند.

```
Router(config-router)# neighbor neighbor soft-configuration inbound  
Router# clear ip bgp neighbor soft [in | out]
```

# Cisco in Persian