Multiprotocol Label Switching(MPLS)

مینا رضایی

بهمن 1394

فهرست

- مقدمه
- بخش اول: مفاهیم پایه
  - MPLS چیست؟
- بخش دوم: MPLS Unicast IP Forwarding
  - MPLS IP Forwarding
  - مروری بر مفهوم و ساختار CEF
  - مروری بر MPLS Unicast IP Forwarding
  - نحوه forwarding پکت ها بر اساس FIB و LFIB
  - ساختار Header و Label, MPLS
  - فیلد MPLS TTL و مفهوم TTL Propagation
  - بررسی Control Plane در MPLS IP Forwarding
  - LDP چیست؟
  - MPLS Label Information Base(LIB)
  - پیاده سازی MPLS Unicast IP Forwarding
  - خلاصه بخش دوم:MPLS Unicast IP Forwarding
- بخش سوم: MPLS L3 VPN
  - VRF چیست؟
  - MPLS VPN Control Plane
  - نحوه عملکرد VRF
  - Route Distinguisher و نحوه ی عملکرد آن
  - Route Target و نحوه عملکرد آن
  - MPLS VPN Data Plane
  - Penultimate Hop Popping
  - پیاده سازی MPLS L3 VPN
  - خلاصه بخش سوم:MPLS L3 VPN

مقدمه:

این مقاله بر اساس فصل Multiprotocol Label Switching کتاب CCIE Routing and Switching v5.0 Official Cert Guide Volume 2, دوره CCIE R&S شرکت INE و داکیومنت های Cisco تهیه شده است.

هدف از این مقاله معرفی مفاهیم پایه ای پروتکل MPLS می باشد. پیش فرض بر آن بوده که حتی الامکان تمام موارد پایه به گونه ای مطرح گردد تا حتی افرادی که با این پروتکل آشنایی چندانی نیز ندارند, به راحتی بتوانند مفاهیم را درک نمایند.

این مقاله به سه بخش اصلی تقسیم می گردد که در ابتدای هر بخش مباحث تئوری مطرح گردیده و نهایتا در گام بعد در بخش هایی که نیازمند پیاده سازی سناریو بوده اند, با پیاده سازی سناریو سعی شده تا مفاهیم تئوری گفته شده در ابتدای هر بخش, پیاده سازی و بررسی گردند و نکات باقی مانده در رابطه با طراحی و پیاده سازی که در طول بخش به آن ها اشاره نشده, مطرح شوند. نهایتا در پایان هر بخش خلاصه ای از آن چه در بخش مربوطه مطرح شده, در قالب جداولی برای یادآوری مفاهیم, تهیه شده است.

در ساختار ISP ها, روترهای ISP نیازی ندارند که از همه ی اطلاعات مربوط به روت های تمام customer ها آگاهی داشته باشند, از طرف دیگر ساختار IP Routing, یک ساختار Destination Base می باشد, به این معنا که تمام دیوایس هایی که در طول مسیر transit قرار دارند, باید از مقصد آگاهی داشته باشند. در نتیجه تمام روترهای داخل SP باید در BGP Table خود اطلاعات route های تمام customer ها را داشته باشند که این غیر ممکن است.

به بیان واضح تر, در ساختار اینترنت برای یک end user فقط این موضوع اهمیت دارد که بتواند به طریقی با یک end user دیگر در سمت دیگر, ارتباط داشته باشد و دلیلی بر این که یک end user از لینک ها و مسیرهای ارتباطی درون cloud SP آگاهی داشته باشد, وجود ندارد. از طرف دیگر برای SP هم فقط Ingress Point و Egress Point اهمیت دارد و داخل کلود SP دیگر آدرس IP مبدا اصلی و مقصد نهایی برای روترهای Core اهمیتی نخواهد داشت.

برای حل این مشکل یک راه استاندارد استفاده از تانل می باشد, یعنی ترافیک مربوط به end-to-end client ها از طریق تانل رد و بدل شوند, و این امر مستلزم آن است که تانلی بین روترهایی که در Ingress Point و Egress Point قرار دارند زده شود, در نتیجه در این حالت فقط این دو روتر که در مرز یا اصطلاحا لبه ساختار SP قرار دارند, احتیاج دارند تا از تمام اطلاعات روتینگ customer ها آگاهی داشته باشند و روترهای درون Cloud SP یا اصطلاحا روترهای Core فقط باید از این امر آگاهی داشته باشند که Ingress Router چه طور باید با Egress Router ارتباط داشته باشد. برای پیاده سازی تانل می توان از انواع مختلف تانل ها استفاده نمود: QinQ, IPinIP, GRE, MPLS VPN و ... از میان این تانل های ذکر شده MPLS استاندارد اصلی محسوب می گردد.

MPLS چیست؟

پروتکل MPLS یا Multiprotocol Label Switching مفهوم متفاوتی را برای Forward پکت ها مطرح می کند, در ساختار MPLS ارسال پکت ها دیگر بر اساس آدرس IP مقصد نخواهد بود, بلکه بر اساس MPLS Label ها این ارسال صورت خواهد گرفت. پس تصمیم گیری مسیریابی می تواند به جای صرفا استفاده از آدرس IP مقصد, بر اساس فاکتورهای دیگر مثلا مباحث: Traffic Engineering, نیازهای QoS و... باشد.

در واقع می توان این گونه تعریف نمود که هنگامی که می گوییم Multiprotocol, یعنی MPLS قادر است Payload های مختلف را Transport کند, حال این Payload ها می توانند لایه 2 یا لایه 3 ای باشند:

لایه 2 همانند: Ethernet, ATM, Frame Relay, PPP, HDLC و ...

لایه 3 مانند: IPV4, IPV6

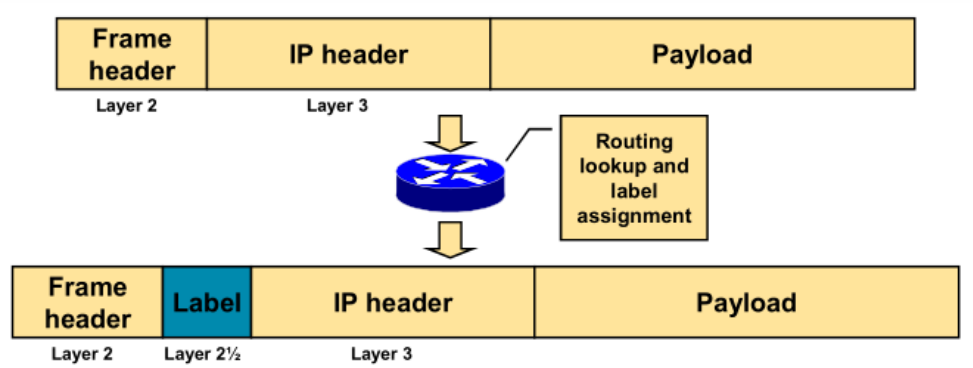
زمانی که از Label Switching صحبت می کنیم یعنی switching ترافیک بین اینترفیس ها بر اساس مقدار Local Label اختصاص داده شده به پکت ها می باشد(در ادامه با مفهوم Local Label آشنا خواهید شد) به نوعی می توان نحوه عملکرد MPLS را به Frame Relay تشبیه کرد, یعنی همانطور که در ساختار Frame Relay ما input/output DLCI ها را داریم, در ساختار MPLS نیز ما با Label ها سر و کار داریم.

اما این که چرا اکثر SP ها از MPLS استفاده می کنند, شاید بتوان مهم ترین دلیل را این امر بر شمرد که در این حالت دیگر نیازی نیست بر روی تمام روترهای Provider(P), پروتکل BGP فعال شود و این روترها نیاز به یک BGP Routing Table کامل از تمام اطلاعات مسیرهای Customer ها داشته باشند, پس در این صورت می توان فضای Table بر روی روترهای Provider(P) را حفظ کرد. هم چنین با استفاده از MPLS می توان خدمات VPN L2/L3 را برای Customer ها فراهم نمود و هم چنین مباحث مربوط به Traffic Engineering و... که در ابتدا به آن ها اشاره گردید.

MPLS اپلیکیشن های متفاوتی را در برمی گیرد, اما آن چه در این مقاله بررسی می گردد دو مفهوم زیر خواهد بود:

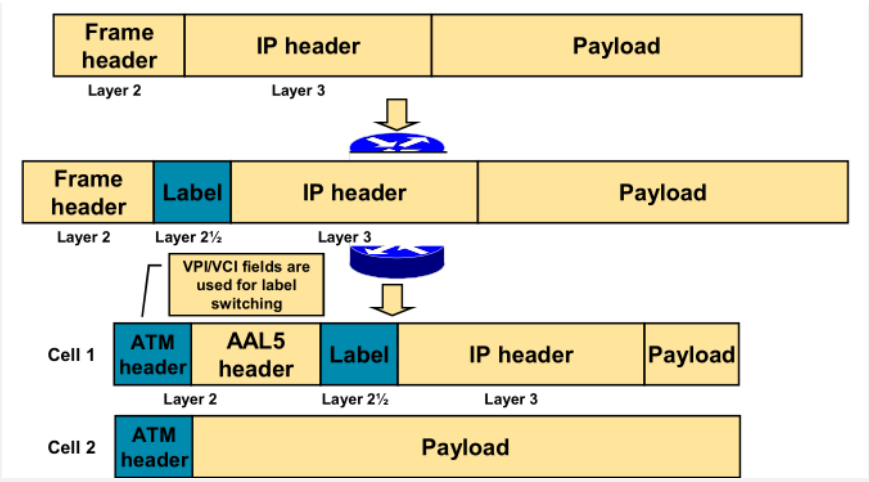
- MPLS unicast IP
- MPLS VPNs

نکته MPLS شامل Frame-mode MPLS و Cell-mode MPLS است که آن چه در این مقاله مورد بررسی قرار می گیرد حالت Frame mode MPLS می باشد. برای درک بهتر حالت Frame-mode شکل زیر را در نظر بگیرید:



تصویر 1 Frame-Mode MPLS

حالت Cell-Mode در ساختارهای ATM مورد استفاده قرار می گیرد و منطق آن به صورت شکل زیر می باشد:



تصوير **2** Cell-Mode MPLS

بخش دوم: MPLS Unicast IP Forwarding

می توان از MPLS برای ارسال ساده پکت ها نیز استفاده کرد(بدون نیاز به مباحث تانلینگ و ...) با استفاده از این روش منطق ارسال پکت ها به جای استفاده از آدرس IP مقصد, بر اساس Label خواهد بود.

شاید MPLS Unicast IP Forwarding به تنهایی چندان کارایی و سودی نداشته باشد اما به بسیاری از کاربردهای مفید و مهم MPLS کمک خواهد نمود مانند: MPLS VPN

MPLS از Control Plane پروتکل ها(مثلا OSPF به همراه LDP) برای یادگیری Label ها و ارتباط دادن این Label ها با Prefix های مقصد و ایجاد یک جدول Forwarding درست, بهره می گیرد. هم چنین MPLS منطق اصلی ارسال پکت ها توسط Data Plane را نیز تغییر می دهد که در ادامه به بررسی این موارد خواهیم پرداخت.

MPLS IP Forwarding:

MPLS الگوی کاملا متفاوتی را از ارسال پکت ها معرفی می کند. در این ساختار Host ها حق Label زدن به پکت ها یا ارسال و دریافت پکت های Label خورده را نخواهند داشت. اما در همین ساختار تعدادی از روترها وظیفه Label زدن به پکت های بدون Label, ارسال و دریافت پکت های Label خورده و نهایتا حذف Label از یک پکت را بر عهده دارند. روترهای MPLS یعنی روترهایی که وظیفه inject(Push), Remove(Pop) و Forward پکت ها را بر عهده دارند, از منطق Forwarding MPLS استفاده می کنند.

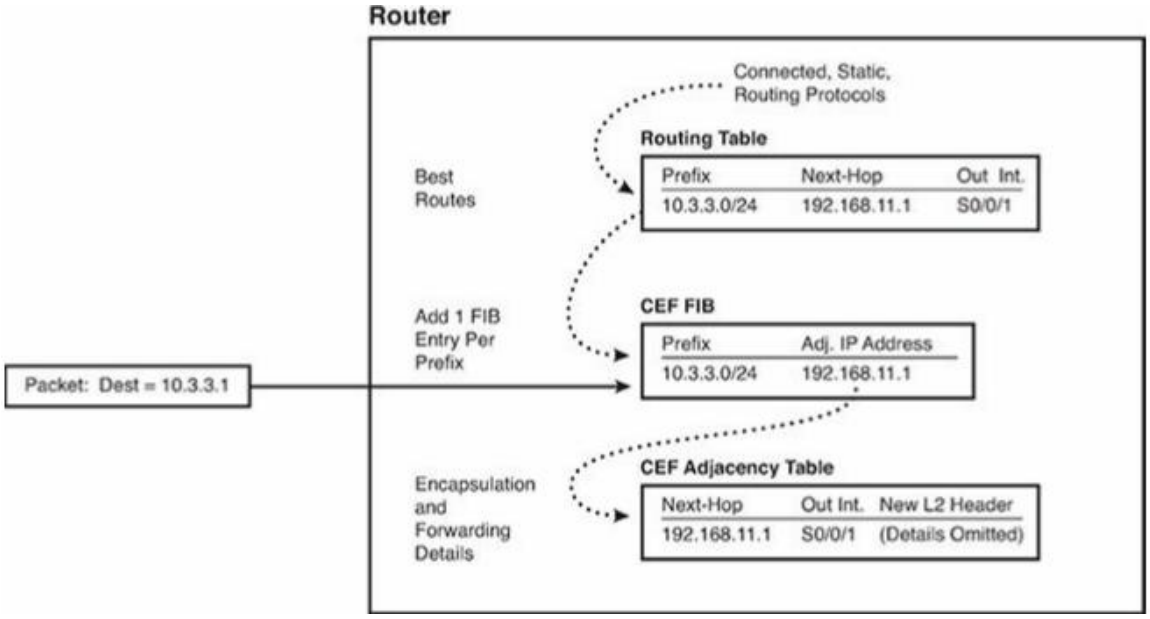
بیاید این مفاهیم را دقیق تر بررسی کنیم و به سراغ منطق فورواردینگ در MPLS برویم. MPLS به منطق و ساختار Cisco Express Forwarding متکی است, پس ابتدا ساختار CEF و نحوه عملکرد آن را بررسی می کنیم سپس به سراغ مفهوم جدیدی که MPLS Label Forwarding Information Base (LFIB) نامیده می شود خواهیم رفت.

مروری بر مفهوم و ساختار CEF:

Control Plane یک روتر با استفاده از اطلاعاتی که از طریق Routing Protocol ها, Static Route ها و روت های connected جمع آوری می کند, Routing Information Base(RIB) یا همان Routing Table را می سازد. حال اگر CEF بر روی روتر فعال باشد, Control Plane یک گام فراتر رفته و CEF Forwarding Information Base(FIB) را می سازد, به ازای هر روتی که در Routing Table قرار دارد, ما یک ورودی برای FIB

خواهیم داشت. ورودی های FIB می شوند اطلاعاتی که برای Forward پکت ها به آن ها احتیاج داریم یعنی: اینترفیس خروجی که از طریق آن به مسیر موردنظر خواهیم رسید و آدرس IP روتری که گام بعدی محسوب می شود(next-hop). هم چنین CEF adjacency table, اطلاعات data-link Header جدیدی که روترها قبل از این که بسته ای را ارسال کنند, آن را به ابتدای پکت اضافه می کنند, نگهداری می کند.

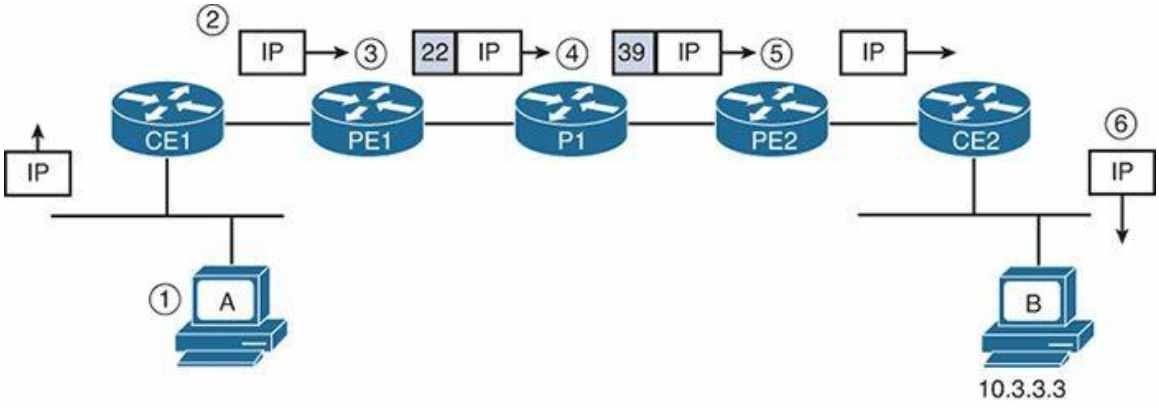
برای Data Plane, یک روتر که CEF بر رویش فعال باشد, آدرس IP مقصد پکت را با اطلاعات موجود در FIB مقایسه می کند و از Routing Table چشم پوشی می کند. در واقع CEF با بهینه سازی ساختار FIB باعث می شود تا Delay مربوط به Forward پکت ها کاهش پیدا کرده و حجم پکت هایی که در ثانیه می توانند توسط روتر ارسال شوند, افزایش یابد. برای هر پکتی که روتر دریافت می کند, ابتدا جدول FIB اش را جستجو کرده تا یک ورودی منطبق با آن روت پیدا کند, سپس بر اساس آن چه از FIB به دست می آورد, adjacency table را جستجو خواهد کرد و entry منطبق با آن چه از FIB به دست آمده را پیدا کرده, این اطلاعات را تحت یک data-link هدر به ابتدای بسته اضافه می کند و بسته را ارسال می نماید. در تصویر زیر این مراحل به صورت گرافیکی نمایش داده شده است.



تصویر 3 IP Routing Table و CEF FIB(در این ساختار MPLS استفاده نشده است)

مروری بر MPLS Unicast IP Forwarding:

الگوی ارسال پکت ها در MPLS به این گونه است که: Host ها فقط پکت های بدون Label را generate می کنند. سپس تعدادی از روترها به این بسته های بدون Label, Label اضافه می کنند, تعداد دیگری از روترها وظیفه دارند این بسته های Label خورده را Forward کنند و بعضی از روترها نیز وظیفه حذف Label ها از بسته ها و ارسال بسته های بدون برچسب به سمت Host ها را بر عهده دارند. نهایتا کامپیوتر مقصد از وجود MPLS هیچ آگاهی نخواهد داشت. همانند آنچه در تصویر 4 مشاهده می نمایید:



تصویر4 MPLS Packet Forwarding

مراحل نشان داده شده در تصویر 4 به شرح ذیل می باشند:

1- ابتدا Host A یک پکت بدون Label را به مقصد 10.3.3.3 تولید و ارسال می کند.
2- روتر CE1 که فیچر MPLS بر روی آن پیکربندی نشده و هیچ آگاهی از MPLS ندارد, این پکت بدون برچسب را بر اساس آدرس IP مقصد Forward می کند(روی روتر CE1, فعال بودن یا نبودن CEF تفاوتی ندارد)
3- روتر PE1 که بخشی از شبکه MPLS ما می باشد, پکت بدون برچسب را دریافت کرده, یک Label با مقدار 22 به آن اضافه می کند و پکت را ارسال می کند.
4- روتر P1 پکت Label دار را دریافت کرده, Label آن را عوض کرده و یک Label جدید با مقدار 39 به آن اضافه کرده و بسته را ارسال می کند.
5- روتر PE2 پکت Label دار را دریافت کرده, Label آن را حذف می کند و پکت بدون Label را به سمت روتر CE2 ارسال می کند.
6- روتر CE2 که فیچر MPLS روی آن فعال نیست, پکت بدون Label را دریافت کرده و بر اساس آدرس IP مقصد, پکت را به مقصد مشخص ارسال می نماید(روی روتر CE2, فعال بودن یا نبودن CEF تفاوتی ندارد)

با توجه با آن چه در تصویر بالا مشاهده نمودید, می خواهیم با مفهومی به نام LSR آشنا شویم:

LSR: LSR یا Label Switching Router به هر روتری گفته می شود که Label, MPLS را بفهمد(تشخیص دهد.)

به عنوان مثال در تصویر 2 روترهای PE1, P1 و PE2 اصطلاحا LSR نامیده می شوند. در جدول زیر انواع LSR ها و عملی که هر نوع انجام می دهد را مشاهده می نمایید:

نوع LSR	عملی که این نوع LSR انجام می دهد
Label Switch Router(LSR)	هر روتری که برچسبی(Label) را به پکتی اضافه کند(عمل Push), برچسبی را از پکتی حذف نماید(عمل Pop), و یا پکتی Label دار را Forward نماید.
Edge LSR(E-LSR)	یک LSR که درلبه ی شبکه MPLS قرار دارد و فرآیندهای مربوط به Label زدن و حذف Label ها از پکت ها را بر عهده دارد.
Ingress E-LSR	برای پکت های مشخص, روتری که یک پکت بدون Label را دریافت کرده و یک Label به اول هدر IP آن پکت اضافه میکند.
Egress-LSR	برای پکت های مشخص, روتری که یک پکت Label دار را دریافت کرده و تمام MPLS Label های آن را حذف می کند و سپس یک پکت بدون Label را Forward می کند.

جدول 1 انواع LSR ها

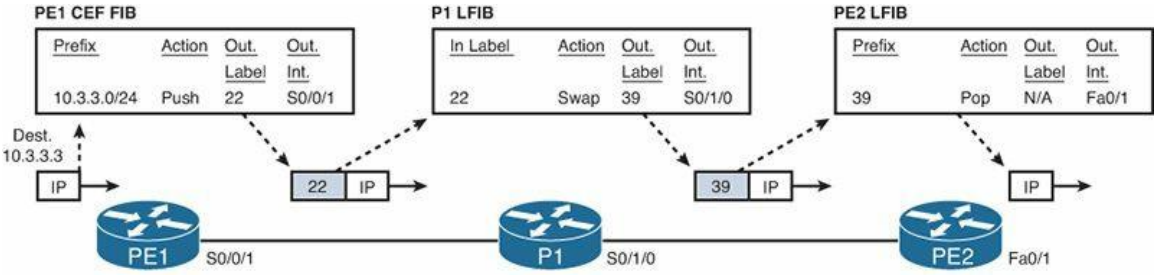
نحوه forwarding پکت ها بر اساس FIB و LFIB:

برای ارسال پکت ها LSR ها هم از CEF FIB و هم از MPLS LFIB استفاده می کنند. هم FIB و هم LFIB تمام اطلاعات ضروری Label ها را نگهداری می کنند مانند: اینترفیس خروجی و اطلاعات Next-hop.

در اصل FIB و LFIB با هم تفاوت دارند چرا که از یکی برای ارسال پکت های Label نخورده و از دیگری برای ارسال پکت های Label دار استفاده می شود به بیان بهتر:

FIB: اگر روتری پکتی بدون Label دریافت نماید, IOS سیسکو آدرس IP مقصد پکت را با Prefix های موجود در FIB مقایسه کرده و نهایتا بر اساس آنچه از FIB به دست می آورد, پکت را ارسال می کند.

LFIB: اگر روتر پکتی را دریافت کند که Label خورده باشد, IOS سیسکو به LFIB مراجعه کرده و Label پکت را با Label های موجود در LFIB مقایسه کرده و نهایتا بر اساس آن چه از LFIB به دست می آورد, پکت را Forward می کند. برای درک بهتر می توان تصویر زیر را در نظر گرفت:



تصویر5 استفاده از CEF FIB و MPLS LFIB برای ارسال پکت ها

مراحل تصویر بالا به شرح ذیل می باشند:

- **PE1**: روتر PE1 زمانی که پکت بدون Label را دریافت می کند از FIB استفاده می کند. در واقع در FIB به دنبال entry(ورودی) می گردد که با آدرس IP مقصد پکتی که دریافت کرده, مطابقت یابد که با توجه به شکل Prefix 10.3.3.0/24 را در FIB یافت می کند, اما این سطر یافت شده در جدول, حاوی اطلاعات دیگری نیز هست از جمله عملی که PE1 باید انجام دهد یعنی اضافه کردن Label(Push) و هم چنین مقدار این Label که 22 در نظر گرفته شده است.
- **P1**: از آن جایی که P1 یک پکت Label دار دریافت کرده از LFIB استفاده می کند, یعنی در LFIB به دنبال سطری که در ستون Label برایش مقدار 22 درج شده باشد می گردد, هنگامی که این entry را یافت به ستون Action نگاه کرده و با عمل swap مواجه می شود پس مقدار Label را به آن چه برایش به عنوان Out Label مشخص شده تغییر می دهد(یعنی مقدار 39) و بسته را ارسال می کند.
- **PE2**: روتر PE2 هم, چون یک پکت Label دار دریافت کرده از LFIB استفاده می کند, پس در LFIB به دنبال entry ای می گردد که Label 39 برایش ثبت شده باشد, هنگامی که این entry را یافت در ستون Action با عمل POP مواجه می شود, پس Label را از بسته حذف کرده و پکت بدون Label را به اینترفیس خروجی ای که برایش مشخص شده Forward می کند.

در این مثال روترهای P1 و PE2 هرگز از آدرس IP مقصد پکت در فرآیند ارسال پکت, استفاده نمی کنند.

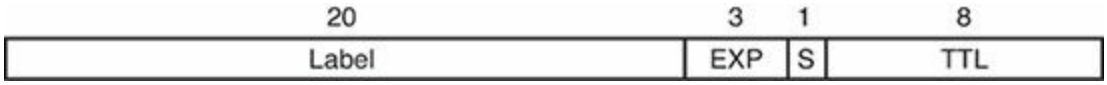
ساختار Header و Label, MPLS:

هدر MPLS, 4 بایت(32 بیت) می باشد که دقیقا قبل از هدر IP قرار می گیرد. همانند تصویر زیر:

Layer 2 Header	MPLS Header	Layer 3 Header(IP)	Payload

تصویر 6 قالب پکت در ساختار MPLS

در برخی موارد Label, MPLS را همان هدر MPLS معرفی می کنند که این امر اشتباه است چرا که Label, MPLSیک مقدار 20 بیتی درون هدر MPLS می باشد و در واقع بخشی از هدر می باشد.



تصویر 7 ساختار MPLS Header

در جدول زیر با وظیفه ی هر بخش هدر MPLS بیشتر آشنا می شویم:

Field	طول	وظیفه
Label	20	مشخص کننده مسیر برچسب خورده ای است که پکت باید به آن switch شود.
Experimental(EXP)	3	برای مباحث QoS مورد استفاده قرار می گیرد.
Bottom of Stack(S)	1	بیت Flag, اگر مقدار آن یک باشد یعنی برچسب بلافاصله قبل از هدر IP قرار گیرد.
Time to Live(TTL)	8	دقیقا همانند فیلد TTL در هدر IP عمل می کند.

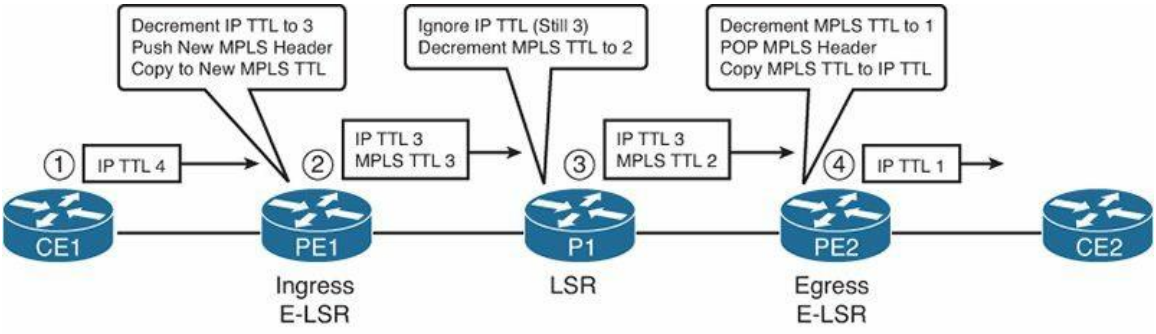
جدول 2 فیلدهای هدر MPLS

فیلد MPLS TTL و مفهوم TTL Propagation:

فیلد TTL موجود در هدر IP دو فیچر مهم را پشتیبانی می کند: 1- مکانیزمی برای شناسایی Loop 2- روشی برای دستور Traceroute برای پیدا کردن آدرس IP هر روتری که در یک مسیر مشخص end-to-end قرار دارد.

فیلد TTL موجود در هدر MPLS نیز دقیقا همان فیچرهای فیلد TTL هدر IP را فراهم می آورد. نیاز MPLS به فیلد TTL از آن جهت است که باید هدر IP زمانی که پکت وارد شبکه MPLS می شود, کلا نادیده گرفته شود. در واقع LSR ها مقدار فیلد TTL هدر MPLS را به جای فیلد TTL هدر IP کاهش می دهند. نحوه ی عملکرد ingress- ELSR و egress E-LSR ها برای کار با فیلد TTL به شرح ذیل می باشد:

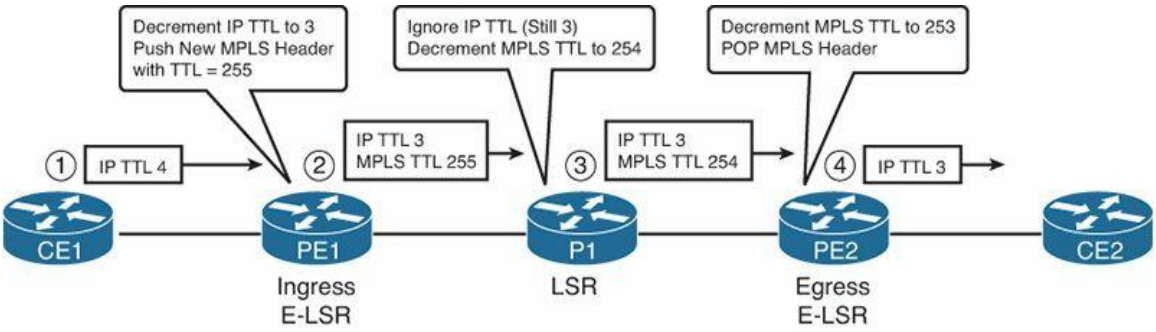
- **Ingress E-LSR**: ابتدا فیلد TTL هدر IP را یک واحد کاهش داده, سپس به پکت بدون برچسب, یک Label اضافه کرده و بعد مقدار فیلد TTL هدر IP را در فیلد TTL, هدر MPLS کپی می کند.
- **LSR**: زمانی که یک LSR, Label یک پکت را عوض می کند, فیلد TTL هدر MPLS را نیز کاهش داده و فیلد TTL هدر IP را نادیده می گیرد.
- **Egress E-LSR**: بعد از این که فیلد TTL هدر MPLS را یک واحد کاهش داد, هدر MPLS را از پکت کلا جدا کرده(حذف می کند) و مقدار فیلد TTL هدر MPLS را داخل فیلد TTL هدر IP کپی می کند.



تصویر 8 MPLS TTL Propagation

در ساختاری مانند شکل بالا روترهای MPLS در واقع همان مقداری را نهایتا برای TTL بیان می کنند که حتی اگر ساختار MPLS در این میان وجود هم نداشت, همین مقدار به دست می آمد, مشکل بزرگ این روش آن است که از سمت روترهای LSR هم, پیام ICMP Time Exceeded بر می گردد, در خیلی از موارد ما نمی خواهیم Customer ها از داخل ساختار SP مطلع شوند, اما در حالتی مانند تصویر 6 برای هر customer با دستور traceroute, شبکه MPLS قابل رویت خواهد بود. SP ها معمولا MPLS را پیاده سازی می کنند تا خدمات لایه 3, WAN را برای مشتریان خود فراهم آورند, حال اگر customer های SP بتوانند آدرس IP, LSR های شبکه MPLS را پیدا کنند, این موضوع هم برای customer ای که تنها می خواهد روتر customer دیگر در آن سمت کلود SP را ببیند چندان جالب نیست و هم بدتر از آن یک مشکل امنیتی برای SP محسوب می گردد.

روترهای سیسکو می توانند MPLS TTL Propagation را غیر فعال کنند. یعنی زمانی که یک روتر Ingress E-LSR یک پکت را دریافت می کند, اول TTL هدر IP آن پکت را یک واحد کاهش داده, سپس مقدار فیلد TTL هدر MPLS را برابر با 255 می گذارد و این هدر MPLS را به پکت اضافه کرده و آن را به داخل کلود MPLS ارسال می کند. روتر LSR که پکت را دریافت می کند, مقدار TTL هدر MPLS را یک واحد کم کرده و کلا هدر IP پکت را نادیده می گیرد و بسته را ارسال می کند. نهایتا روتر Egress E-LSR که پکت را دریافت می کند, فیلد TTL هدر MPLS را یک واحد کم کرده, هدر MPLS را از پکت جدا می کند(حذف می کند) و پکت را ارسال می کند(یعنی پکت با همان مقدار TTL هدر IP اش که در همان ابتدا فقط یک واحد کم شد, به خارج از کلود MPLS ارسال می گردد). در نتیجه از بیرون شبکه MPLS از دیدگاه TTL همانند آن است که پکت فقط یک روتر را رد کرده است و تمام روترهایی که در کلود MPLS پکت از آن ها عبور کرده, از دید customer پوشیده می ماند. این روند را می توان به خوبی در تصویر زیر مشاهده نمود:



تصویر 9 غیرفعال سازی MPLS TTL Propagation

می توان غیر فعال سازی انتشار TTL(TTL Propagation) را برای دو دسته از پکت ها در روترهای سیسکو انجام داد:

1- پکت هایی که از سمت customer ها به روترهای لبه ی شبکه ی MPLS می رسند
2- پکت هایی که در داخل شبکه MPLS جریان دارند

به عنوان مثال در تصویر 9 می توان کاری کرد که روتر PE1 برای پکت هایی که به صورت محلی ایجاد شده اند(یعنی پکت هایی که روتر PE1 برای ارتباط با روترهای داخل کلود MPLS از آن ها استفاده می کند) TTL Propagation فعال باشد, یعنی PE1 با کامند tracerouet بتواند به تمام روترهای داخل شبکه MPLS دسترسی داشته باشد. و هم چنین به طور همزمان اگر PE1 پکتی را از سمت customer ای دریافت کند که قرار باشد این پکت مربوط به customer را به داخل کلود MPLS فوروراد کند, برای این پکت ها TTL Propagation غیر فعال شود و به این ترتیب از فراگیری آدرس های IP داخل شبکه MPLS توسط customer ها جلوگیری می کنیم. برای استفاده از این فیچر نیز در محیط Global از کامند زیر استفاده می کنیم:

```
PE1(config)#no mpls ip propagate-ttl ?
  forwarded  Propagate IP TTL for forwarded traffic
  local      Propagate IP TTL for locally originated traffic
  <cr>
```

نکته اگر بخواهیم TTL Propagation را کلا در کلود MPLS غیر فعال کنیم, علاوه بر روتر PE1 باید دستور بالا در تمام روترهای داخل کلود MPLS نیز زده شود.

بررسی Control Plane در MPLS IP Forwarding:

در اصل عمل مسیر یابی IP بر اساس FIB صورت می گیرد, یعنی روترها با استفاده از Control Plane پروتکل های مسیریابیRouting Table را می سازند و براساس Routing Table, CEF FIB ساخته می شود.

عمل Forwarding در MPLS هم دقیقا همین گونه است, یعنی MPLS با استفاده از Control Plane پروتکل های مسیربابی و LDP, Label های MPLS که برای رسیدن به یک Prefix خاص نیاز دارد را, فرا گرفته, سپس FIB و

LFIB بر اساس این Label های فراگرفته شده, ساخته می شوند. پس ابتدا به بررسی نحوه عملکرد LDP(Label Distribution Protocol) می پردازیم.

نکته سیسکو برای Labelگذاری پکت ها قبل از معرفی LDP فیچری را معرفی نمود به نام TDP(Tag Distribution Protocol) که در TDP از اصطلاح Tag Switching به جای Label Switching استفاده می شد.

LDP چیست؟

LDP یا Label Distribution Protocol این امکان را فراهم می آورد تا برای هر Prefix ای که در Routing Table وجود دارد بتوانیم Label ای را advertise نماییم. در واقع LSR ها با استفاده از LDP به همسایه های خود پیام های LDP که حاوی Prefix ها به همراه Label اختصاص یافته به آن هاست را ارسال می کنند و از این طریق به همسایه ی خود می گویند که اگر می خواهی پکتی را به این IP Prefix ارسال کنی, به آن پکت MPLS Label ای که در آپدیت LDP برای این Prefix مشخص شده, اضافه کن و بعد پکت را برای من ارسال کن.

چگونه ارتباط LDP بین روترها برقرار می شود؟

LDP برای پیدا کردن سایر همسایه های LDP از Hello Protocol استفاده می کند. پیام های Hello از آدرس مالتی کست: 224.0.0.2 و هم چنین source و destination پورت 646, پروتکل UDP استفاده می کنند.

پیام های Hello شامل یک Transport Address هستند که از این آدرس برای برقراری TCP Session استفاده می شود, این Transport Address بر اساس LSR ID انتخاب می شود. قوانین انتخاب LSR-ID هم دقیقا همانند Router-ID است, یعنی اگر LSR-ID به صورت دستی پیکربندی نشده باشد, بزرگترین آدرس IP اینترفیس Loopback در وضعیت up/up, به عنوان LSR-ID انتخاب می شود در غیر این صورت بزرگترین آدرس IP اینترفیس فیزیکی در وضعیت up/up به عنوان LSR-ID انتخاب خواهد شد. LSR-ID هرچه که باشد همان مقدار برای Transport Address که بر مبنای آن ارتباط TCP شکل خواهد گرفت, در نظر گرفته می شود.

نکته برای تعیین LSR-ID به صورت دستی می توان از دستور زیر استفاده نمود:

```
(config)# mpls ldp router- id
```

بعد از این که همسایه ها با استفاده از پیام های Hello پیدا شدند, همسایگی از طریق پروتکل TCP و به صورت Unicast شکل خواهد گرفت. حال شرط لازم برای برقراری این ارتباط TCP آن است که Transport Address ای

که همسایه با پیام Hello برای روتر ارسال کرده, Reachable باشد.(به بیان دیگر حتما در Routing Table روتر باید مسیری برای رسیدن به Prefix مربوط به Transport Address وجود داشته باشد.)

بعد از این که TCP Connection برقرار شد هر روتر تمام Prefix های موجود در Routing Table اش را به همراه Label های مربوط به آن ها را به سایر همسایه های LDP خود advertise می کند. پس به صورت خلاصه می توان ویژگی های پروتکل LDP را در جدول زیر خلاصه نمود:

LDP Feature	LDP Implementation
Transport protocols	UDP (Hellos), TCP (updates)
Port numbers	646 (LDP), 711 (TDP)
Hello destination address	224.0.0.2
Who initiates TCP connection	Highest LDP ID
TCP connection uses this address	Transport IP address (if configured), or LDP ID if no transport address is configured
LDP ID determined by these rules, in order of precedence	Configuration Highest IP address of an up/up loopback when LDP comes up Highest IP address of an up/up nonloopback when LDP comes up

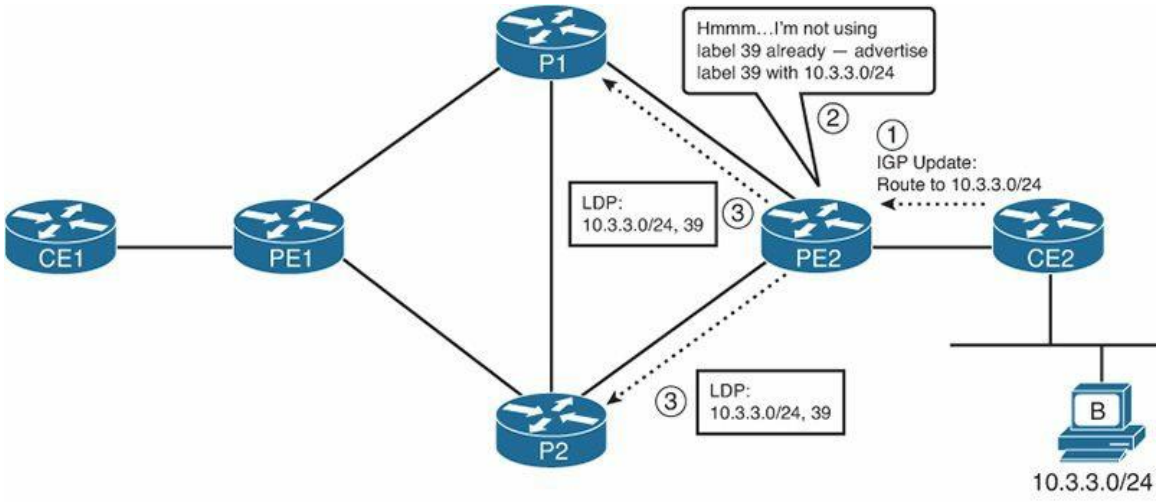
جدول 3 LDP References

نحوه ی عملکرد LDP:

به محض آن که روتر یک مسیر جدید را فرا بگیرد, LSR یک Label به آن مسیر اختصاص می دهد که به این Label, Local Label گویند پس:

Local Label: Label ای است که یک LSR از آن به عنوان نماینده یک Prefix موجود در routing Table, استفاده می کند.

زمانی که یک LSR یک مسیر جدید را فرا می گیرد, بعد از اضافه کردن آن مسیر به Routing Table, Label ای که تاکنون به هیچ مسیری اختصاص نیافته باشد را به آن Prefix اختصاص می دهد و آن Prefix را به همراه Label ای که به آن اختصاص داده در قالب یک پیام LDP برای همسایه های LDP اش ارسال می کند.



تصویر 10 فرآیند اجرای LDP برای یک مسیر جدید

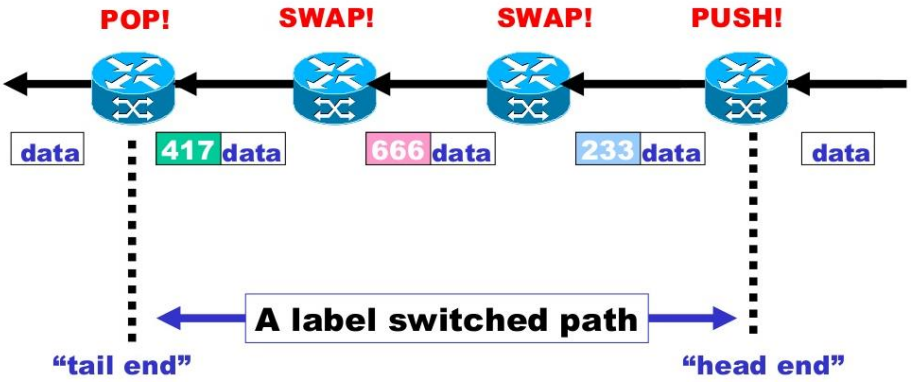
شرح مراحل نشان داده شده در تصویر بالا به صورت ذیل می باشد:

1- PE2 یک مسیر جدید را فرا می گیرد و این مسیر جدید را به Routing Table خود اضافه می کند.
2- PE2 به دنبال Label ای که تاکنون به هیچ Prefix ای اختصاص نیافته باشد می گردد و سپس این Local Label را به این مسیر جدید, اختصاص می دهد.
3- PE2 با استفاده از پیام LDP این Prefix جدید به همراه Label ای که به آن اختصاص یافته را برای همسایه های LDP اش advertise می کند.

در اینجا تعریفی مطرح می گردد با عنوان LSP:

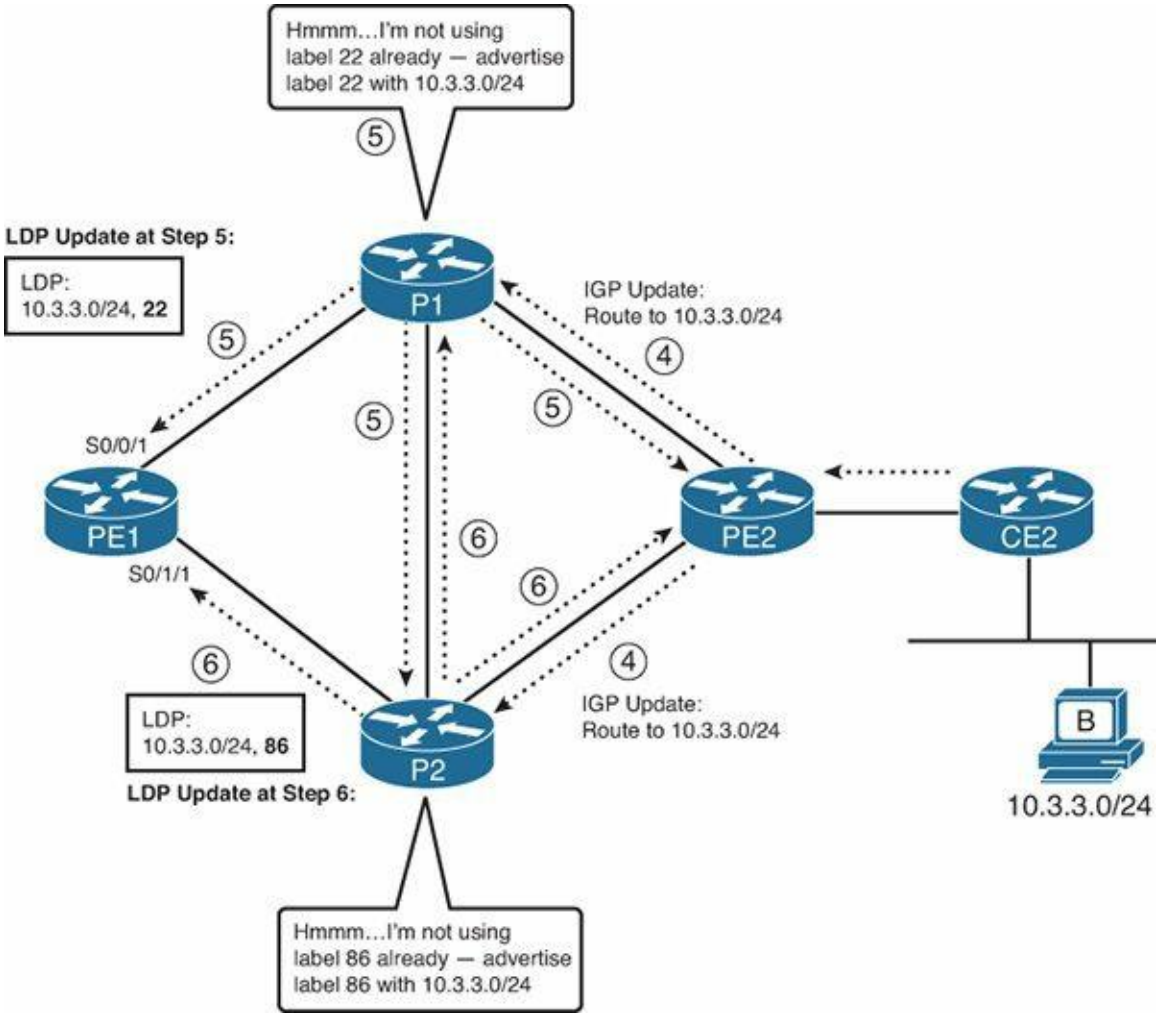
LSP(Label Switched Path): در واقع می شود مجموعه ای از مسیرهای برچسب خورده ای که برای Forward پکت ها از آن ها استفاده می شود. نکته آن است که LSP ها یک سویه(یک طرفه) هستند.

تصویر زیر این مفهوم را واضح تر نشان می دهد:



Label Switched Path(LSP) **تصویر 11**

حال برای ساختن یک LSP, ابتدا روترهای MPLS نیاز دارند تا توسط یک Routing Protocol, Route هایی که توسط LDP برایشان advertise می شود را فراگیرند, معمولا در شبکه MPLS از IGP ها برای این امر استفاده می گردد:



تصویر 12 فرآیند Advertise یک LSP درست

مراحل شکل بالا که به نوعی ادامه تصویر 10 می باشد, به شرح ذیل است:

4- PE2 از EIGRP برای تبلیغ مسیر 10.3.3.0/24 به P1 و P2 استفاده می کند.

5- P1 هنگام یادگیری این Route جدید یک Local Label با مقدار 22 را به این Prefix اختصاص داده و با استفاده از LDP این Prefix جدید و Label ای که به آن اختصاص داده را برای همسایه های LDP اش advertise می کند.

6- P2 هم هنگام یادگیری این Prefix جدید یک Local Label با مقدار 86 به این Prefix اختصاص داده و با استفاده از LDP این Prefix جدید به همراهLabel اختصاص یافته به آن را به تمام همسایگان LDP اش advertise می کند.

پس به ازای هر مسیری در Routing Table یک LSR, این فرآیند کلی رخ خواهد داد:

هر زمان LSR یک Route جدید فرا گیرد, یک Label Local به آن مسیر اختصاص داده و سپس Prefix جدید را به همراه Label ای که به آن map کرده به همسایگان LDP اش advertise می کند(حتی اگر این تبلیغ مفید نباشد).

بعد از آن که تمام LSR ها همه ی Prefix ها را با استفاده از یک پروتکل IGP فرا گرفتند و تمام map های مربوط به Prefix/Label توسط پیام های LDP تبلیغ شد, حالا تمام LSR ها اطلاعات کافی برای switch پکت های Label دار از Ingress E-LSR به Egress E-LSR را خواهند داشت. یعنی به بیان بهتر LSP برای Switch پکت ها در داخل کلود MPLS ساخته شده است.

حال که اندکی با LDP و نحوه ی عملکرد آن آشنا شدیم, بیایید به سراغ این موضوع برویم که چطور FIB و LFIB بر اساس اطلاعات LIB یا MPLS Label Information Base اطلاعات مورد نیاز خود را به دست می آورند.

MPLS Label Information Base(LIB):

LSR ها Label ها و اطلاعات مربوط به آن ها را در ساختار داده ای با نام LIB ذخیره می کنند. در واقع LIB شامل تمام Label ها و اطلاعات وابسته به آن ها است که ممکن است برای Forward مورد استفاده قرار گیرند, اما هرLSR باید بهترین Label و اینترفیس خروجی را انتخاب کرده و سپس اطلاعات مربوط به بهترین Label ها را در FIB و LFIB قرار دهد, در نتیجه FIB و LFIB تنها شامل Label های مربوط به بهترین LSR ها هستند, در حالی که LIB شامل تمام Label های شناخته شده برای LSR است, چه این Label برای Forward استفاده بشود, چه نشود.

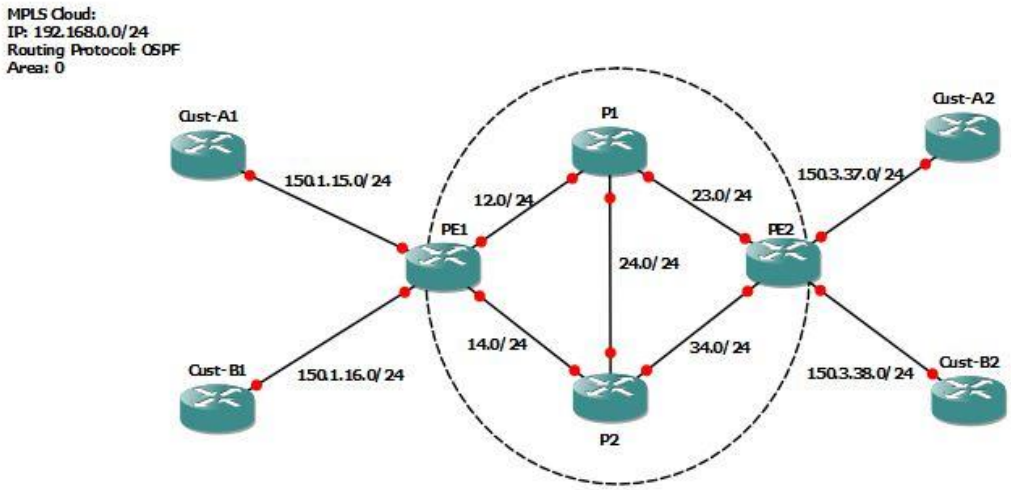
حال برای تصمیم گیری راجع به بهترین Label, LSR ها به تصمیمی که روتینگ پروتکل برای انتخاب بهترین مسیر گرفته است, اتکا می کنند(یعنی Label بهترین مسیری که Routing Protocol انتخاب کرده, به عنوان بهترین Label انتخاب می شود.) خوب این اتکا به تصمیم Routing Protocol دو مزیت را فراهم می کند: 1- LSR ها می توانند از مکانیزم جلوگیری از Loop روتینگ پروتکل بهره مند شوند(یعنی مسیری با بهترین Label و نهایتا LSP از هر گونه Loop ای مستثنی است.) 2- تمام این اتفاقات یعنی انتخاب بهترین Label و مسیری با بهترین Label همزمان با همگرایی در روتینگ پروتکل به وقوع می پیوندد, یعنی زمانی که روتینگ پروتکل به همگرایی رسید, بهترین Label نیز انتخاب شده است.

به طور خلاصه عملی که توسط LSR انجام می شود به شرح ذیل است:

برای هر مسیری در Routing Table بر اساس اینترفیس خروجی و آدرس next-hop ذکر شده برای آن Route, اطلاعات Label مربوط به آن مسیر از LIB استخراج شده و تمام این اطلاعات(یعنی Label, اینترفیس خروجی و next-hop) در FIB و LFIB ثبت می گردد.

پیاده سازی MPLS Unicast IP Forwarding:

تا این بخش کل مطالب تئوری مربوط به MPLS Unicast IP Forwarding تقریبا تشریح شد, برای درک بهتر نحوه عملکرد MPLS برای ارسال ساده پکت ها, سناریویی مانند آنچه در تصویر 13 مشاهده می کنید, پیاده سازی کرده و سعی می کنیم باقی نکات مربوط به نحوه عملکرد LDP را با سناریو توضیح دهیم:



تصویر 13 پیاده سازی شبکه MPLS

برای سناریوی بالا, در حال حاضر فقط بر ارتباطات روترهای داخل کلود MPLS تمرکز می کنیم. برای فعال سازی MPLS مراحل زیر را باید انجام دهیم:

1- فعال سازی CEF(این فیچر به صورت پیش فرض فعال هست اما اگر فعال نبود با کامند: **ip cef** آن را فعال می سازیم)
2- فعال سازی MPLS با استفاده از کامند: **mpls ip**
3- مشخص کردن نوع پروتکلی که برای Label گذاری می خواهیم از آن استفاده کنیم, با استفاده از دستور: **mpls label protocol [ldp|tdp]**

```
PE1(config)#ip cef
PE1(config)#mpls ip
PE1(config)#mpls label protocol ?
  ldp  Use LDP (default)
  tdp  Use TDP
```

دستورات بالا را در هر 4 روتر کلود MPLS می زنیم, سپس باید LDP را فعال سازیم برای انجام این کار دو راه داریم:

1- فعال سازی LDP در محیط پیکربندی روتینگ پروتکل با استفاده از دستور: **mpls ldp autoconfig**(از این دستور فقط برای پروتکل های OSPF و ISIS می توان استفاده کرد و برای EGRP کاربرد ندارد.)

2- فعال سازی LDP در محیط پیکربندی اینترفیس با استفاده از دستور: **mpls ip**

به عنوان مثال برای روتر PE1 در محیط پیکربندی اینترفیس های داخل کلود MPLS با دستور: mpls ip, LDP را بر روی این اینترفیس ها فعال می کنیم:

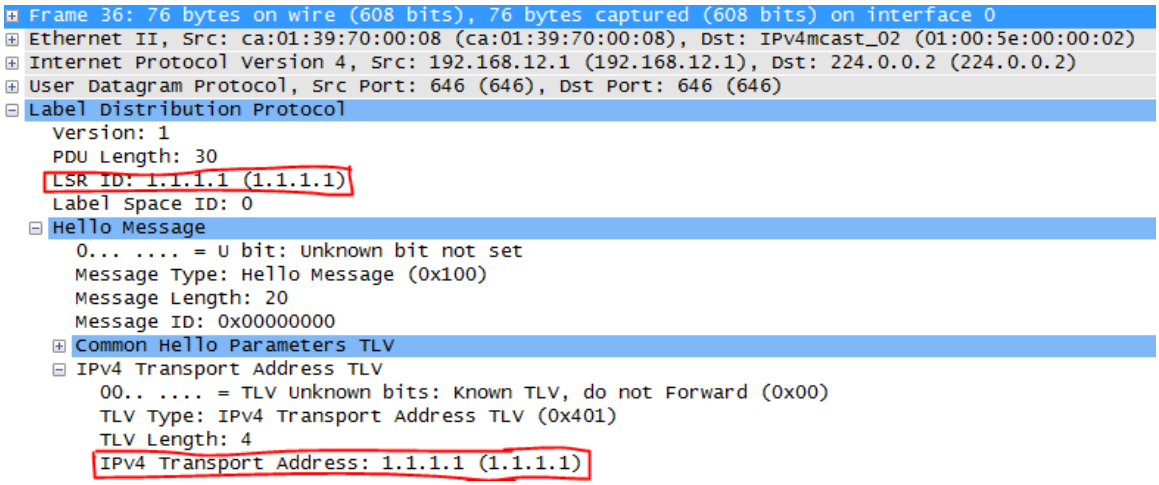
```
interface Ethernet0/0.12
 encapsulation dot1Q 12
 ip address 192.168.12.1 255.255.255.0
 mpls ip
!
interface Ethernet0/0.14
 encapsulation dot1Q 14
 ip address 192.168.14.1 255.255.255.0
 mpls ip
```

نکته در روترهای PE1 و PE2 در محیط پیکربندی اینترفیس های کانکت به Customer ها نباید دستور: mpls ip خورده شده باشد.

به محض این که دستور mpls ip زیر اینترفیس ها خورده شود, روتر شروع به ارسال پکت های LDP Hello برای یافتن همسایه های LDP اش می کند:

```
PE1#debug ip packet detail
IP packet debugging is on (detailed)
PE1#co
*Feb  5 17:05:43.259: IP: s=192.168.14.4 (Ethernet0/0.14), d=224.0.0.2, len 62, rcvd 0
*Feb  5 17:05:43.259:     UDP src=646, dst=646
*Feb  5 17:05:43.259: FIBipv4-packet-proc: route packet from Ethernet0/0.14 src 192.168.14.4 dst 224.0.0.2
*Feb  5 17:05:43.259: FIBfwd-proc: Default:224.0.0.0/24 receive entry
*Feb  5 17:05:43.259: FIBipv4-packet-proc: packet routing failed
*Feb  5 17:05:43.259: IP: s=192.168.14.4 (Ethernet0/0.14), d=224.0.0.2, len 62, input feature
*Feb  5 17:05:43.259:     UDP src=646, dst=646, packet consumed, MCI Check(101), rtype 0, forus FALSE, sendself FALSE, mtu 0, fwdchk FALSE
*Feb  5 17:05:43.894: IP: s=192.168.14.1 (local), d=224.0.0.2, len 62, local feature
*Feb  5 17:05:43.895:     UDP src=646, dst=646, Logical MN local(14), rtype 0, forus FALSE, sendself FALSE, mtu 0, fwdchk FALSE
*Feb  5 17:05:43.895: IP: s=192.168.14.1 (local), d=224.0.0.2 (Ethernet0/0.14), len 62, sending broad/multicast
*Feb  5 17:05:43.895:     UDP src=646, dst=646
*Feb  5 17:05:43.895: IP: s=192.168.14.1 (local), d=224.0.0.2 (Ethernet0/0.14), len 62, sending full packet
*Feb  5 17:05:43.895:     UDP src=646, dst=646
*Feb  5 17:05:44.415: IP: s=192.168.12.1 (local), d=224.0.0.2, len 62, local feature
*Feb  5 17:05:44.415:     UDP src=646, dst=646, Logical MN local(14), rtype 0, forus FALSE, sendself FALSE, mtu 0, fwdchk FALSE
PE1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
PE1(config)#
PE1(config)#
*Feb  5 17:05:44.415: IP: s=192.168.12.1 (local), d=224.0.0.2 (Ethernet0/0.12), len 62, sending broad/multicast
*Feb  5 17:05:44.415:     UDP src=646, dst=646
*Feb  5 17:05:44.415: IP: s=192.168.12.1 (local), d=224.0.0.2 (Ethernet0/0.12), len 62, sending full packet
*Feb  5 17:05:44.415:     UDP src=646. dst=646
```

این پیام Hello که ارسال می شود دارای یک Transport address است, همانگونه که در تصویر زیر مشاهده می کنید:



تا این Transport Address موجود در پیام Hello برای روترهای همسایه reachable نباشد, همسایگی LDP شکل نخواهد گرفت, پس برای آن که این همسایگی شکل بگیرد با استفاده از روتینگ پروتکل OSPF, Prefix مربوط به Transport Address که در واقع همان آدرس IP اینترفیس Loopback می باشد را برای همسایه های LDP, advertise می کنیم:

```
PE1(config)#do sho run | sec ospf
router ospf 1
 network 1.1.1.1 0.0.0.0 area 0
 network 192.168.0.0 0.0.255.255 area 0
```

نکته ما می توانیم با استفاده از فیچر MPLS LDP-IGP Sync کاری کنیم که قبل از این که IGP Path ها بخواهند برای Switch پکت ها مورد استفاده قرار گیرند, حتما اول LDP کامل Establish شود. پیش نیاز های استفاده از این فیچر عبارتند از:

- فقط روی اینترفیس هایی که OSPF یا IS-IS فعال باشد, این فیچر پشتیبانی می شود.
- زمانی این فیچر بر روی اینترفیس فعال می شود که روی آن اینترفیس LDP فعال شده باشد.
- TDP از این فیچر پشتیبانی نمی کند.
- این فیچر بر روی اینترفیس های تانل یا LC-ATM پشتیبانی نمی شود.

با استفاده از این فیچرمی توان LDP و IGP را به منظور کاهش Packet Loss همگام نمود. اگر همسایه LDP, Reachable باشد, IGP یک زمان مشخص برای رسیدن به همگام شدن با LDP صبر می کند که می توان مدت زمان انتظار IGP Session را با استفاده از دستور: mpls ldp igp sync holddown کاهش داد. اگر همسایه LDP در دسترس نباشد, IGP همسایگی خود را شکل می دهد و منتظر می ماند تا LDP Session نیز Establish گردد. اگر همسایگی IGP تشکیل شود اما هنوز LDP-IGP Sync رخ نداده باشد یا مثلا از بین رفته باشد, IGP یک max

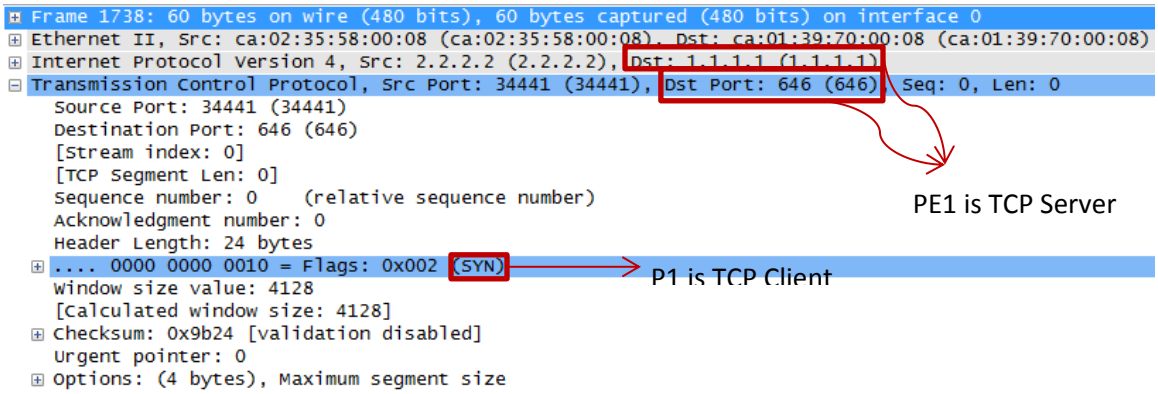
metric روی آن لینک advertise می کند. نکته حائز اهمیت آن است که اگر IGP nonstop forwarding فعال باشد, این فیچر پشتیبانی نخواهد شد. به دو طریق می توان این فیچر را فعال نمود:

1- در محیط پیکربندی روتینگ پروتکل با استفاده از دستور: mpls ldp sync
2- در محیط پیکربندی اینترفیس با استفاده از دستور: mpls ldp igp sync

OSPF را برروی هر چهار روتر کلود MPLS فعال می کنیم, به محض این که همسایگی OSPF شکل می گیرد, ارتباط TCP بین همسایه های LDP نیز شکل گرفته و همسایگی LDP برقرار می شود:

No.	Time	Source	Destination	Protocol	Length	Info
1738	2476.20078	2.2.2.2	1.1.1.1	TCP	60	34441→646 [SYN] Seq=0 Win=4128 Len=0 MSS=536
1739	2476.22493	1.1.1.1	2.2.2.2	TCP	60	646→34441 [SYN, ACK] Seq=0 Ack=1 Win=4128 Len=0 MSS=536
1740	2476.24484	2.2.2.2	1.1.1.1	TCP	60	34441→646 [ACK] Seq=1 Ack=1 Win=4128 Len=0
1741	2476.24490	2.2.2.2	1.1.1.1	LDP	108	Initialization Message
1742	2476.24692	1.1.1.1	2.2.2.2	TCP	60	646→34441 [ACK] Seq=1 Ack=55 Win=4074 Len=0
1743	2476.25794	1.1.1.1	2.2.2.2	LDP	116	Initialization Message Keep Alive Message
1744	2476.27691	2.2.2.2	1.1.1.1	LDP	72	Keep Alive Message
1745	2476.30202	1.1.1.1	2.2.2.2	LDP	169	Address Message Label Mapping Message Label Mapping Message Label Mapping Message
1746	2476.30985	2.2.2.2	1.1.1.1	LDP	169	Address Message Label Mapping Message Label Mapping Message Label Mapping Message
1747	2476.52217	1.1.1.1	2.2.2.2	TCP	60	646→34441 [ACK] Seq=178 Ack=188 Win=3941 Len=0
1793	2525.32239	1.1.1.1	2.2.2.2	LDP	72	Keep Alive Message
1794	2525.51236	2.2.2.2	1.1.1.1	TCP	60	34441→646 [ACK] Seq=188 Ack=196 Win=3933 Len=0
1801	2530.24271	2.2.2.2	1.1.1.1	LDP	72	Keep Alive Message
1802	2530.45948	1.1.1.1	2.2.2.2	TCP	60	646→34441 [ACK] Seq=196 Ack=206 Win=3923 Len=0

اگر به داخل اولین پیام TCP رد و بدل شده بین دو روتر نگاهی بیاندازیم, می توانیم از روی Source Port و Destination Port به راحتی TCP Server را مشخص نماییم:



هر روتری که TCP Source برای آن 646 درج شده باشد می شود TCP Server و روتر دیگر می شود TCP Clientبه عبارت دیگر هر روتری که پیام SYN را ارسال کند می شود TCP Client, اما چون در دیباگ: **debug ip packet detail** نوع پیام TCP مشخص نمی شود, برای تشخیص Server و Client بهتر است به همان شماره پورت مراجعه شود, در حالت عادی این که کدام روتر TCP Server و کدام یک Client است شاید چندان اهمیتی نداشته باشد اما در جایی که مباحث فیلترینگ مطرح می شود, تشخیص این دو از هم بسیار مهم می باشد.

حال که همسایگی LDP شکل گرفت نگاهی به LIB و LFIB می اندازیم, برای مشاهده ی LIB از دستور: show mpls ldp binding استفاده می کنیم:

```
PE1#sho mpls ldp bindings
  lib entry: 1.1.1.1/32, rev 14
        local binding:  label: imp-null
        remote binding: lsr: 4.4.4.4:0, label: 22
        remote binding: lsr: 2.2.2.2:0, label: 21
  lib entry: 2.2.2.2/32, rev 20
        local binding:  label: 22
        remote binding: lsr: 4.4.4.4:0, label: 21
        remote binding: lsr: 2.2.2.2:0, label: imp-null
  lib entry: 3.3.3.3/32, rev 18
        local binding:  label: 21
        remote binding: lsr: 4.4.4.4:0, label: 19
        remote binding: lsr: 2.2.2.2:0, label: 20
  lib entry: 4.4.4.4/32, rev 16
        local binding:  label: 20
        remote binding: lsr: 4.4.4.4:0, label: imp-null
        remote binding: lsr: 2.2.2.2:0, label: 19
  lib entry: 150.1.15.0/24, rev 9
        local binding:  label: imp-null
  lib entry: 150.1.16.0/24, rev 10
        local binding:  label: imp-null
  lib entry: 192.168.12.0/24, rev 11
        local binding:  label: imp-null
        remote binding: lsr: 4.4.4.4:0, label: 23
        remote binding: lsr: 2.2.2.2:0, label: imp-null
  lib entry: 192.168.14.0/24, rev 12
        local binding:  label: imp-null
        remote binding: lsr: 4.4.4.4:0, label: imp-null
        remote binding: lsr: 2.2.2.2:0, label: 22
  lib entry: 192.168.23.0/24, rev 24
        local binding:  label: 24
        remote binding: lsr: 4.4.4.4:0, label: 20
        remote binding: lsr: 2.2.2.2:0, label: imp-null
  lib entry: 192.168.24.0/24, rev 26
        local binding:  label: 25
        remote binding: lsr: 4.4.4.4:0, label: imp-null
        remote binding: lsr: 2.2.2.2:0, label: imp-null
  lib entry: 192.168.34.0/24, rev 22
        local binding:  label: 23
        remote binding: lsr: 4.4.4.4:0, label: imp-null
        remote binding: lsr: 2.2.2.2:0, label: 23
```

در خروجی دستور show mpls ldp binding ورودی های دیتابیس LIB را مشاهده خواهیم کرد, دو موضوع مهم از این خروجی قابل برداشت است:

1- از چه LSR هایی برای این Prefix, چه Label هایی دریافت کرده ایم.
2- چه Label ای را خود همین روتر برای همسایه های LDP اش ارسال نموده است.

Remote-binding می شود Label هایی که از همسایه LDP فرا گرفته شده و Local-binding هم می شود Label ای که همین روتر برای Prefix مورد نظر برای همسایه های LDP اش ارسال نموده, اما عبارت imp-null که در تصویر بالا مشخص گردیده به چه معنی است؟ هر LSR روت های directly-connected اش را با Label, imp-null برای همسایه های LDP اش advertise می کند و این بدان معنی است که به روتر ماقبل خود بیان می کند اگر

قرار است برای این Prefix پکتی را ارسال نمایی, MPLS Label را کلا از آن پکت حذف کن و پکت IP اصلی را برای من ارسال کن. در بخش بعد بیشتر با این مفهوم و کاربرد آن آشنا خواهید شد.

حال به سراغ بررسی LFIB برویم و ببینیم که چگونه LFIB اطلاعاتش را از LIB دریافت کرده, برای بررسی LFIB از دستور: **show mpls forwarding-table** استفاده می کنیم:

```
PE1#show mpls forwarding-table
Local      Outgoing    Prefix           Bytes Label   Outgoing    Next Hop
Label      Label       or Tunnel Id     Switched      interface
20         Pop Label   4.4.4.4/32       0             Et0/0.14    192.168.14.4
21         20          3.3.3.3/32       0             Et0/0.12    192.168.12.2
           19          3.3.3.3/32       0             Et0/0.14    192.168.14.4
22         Pop Label   2.2.2.2/32       0             Et0/0.12    192.168.12.2
23         Pop Label   192.168.34.0/24  0             Et0/0.14    192.168.14.4
24         Pop Label   192.168.23.0/24  0             Et0/0.12    192.168.12.2
25         Pop Label   192.168.24.0/24  0             Et0/0.12    192.168.12.2
           Pop Label   192.168.24.0/24  0             Et0/0.14    192.168.14.4
```

در خروجی دستور بالا همانطور که مشاهده می کنید برای برخی Prefix ها outgoing-label, POP درج شده, این Prefix ها در واقع همان Prefix هایی می باشند که در دیتابیس LIB برچسب imp-null دارند. با توجه به تصویر بالا ما برای رسیدن به Prefix 3.3.3.3/32 دو مقدار Label داریم, همانطور که قبلا نیز بیان کردیم, در LFIB تنها بهترین Label ها قرار می گیرند, بهترین Label ها هم مربوط به بهترین مسیرهایی هستند که توسط روتینگ پروتکل انتخاب می شوند, پس بیایید نگاهی به Routing Table بیاندازیم:

```
PE1#show ip rout
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       a - application route
       + - replicated route, % - next hop override

Gateway of last resort is not set

      1.0.0.0/32 is subnetted, 1 subnets
C        1.1.1.1 is directly connected, Loopback0
      2.0.0.0/32 is subnetted, 1 subnets
O        2.2.2.2 [110/11] via 192.168.12.2, 00:43:39, Ethernet0/0.12
      3.0.0.0/32 is subnetted, 1 subnets
O        3.3.3.3 [110/21] via 192.168.14.4, 00:43:39, Ethernet0/0.14
                 [110/21] via 192.168.12.2, 00:43:39, Ethernet0/0.12
      4.0.0.0/32 is subnetted, 1 subnets
O        4.4.4.4 [110/11] via 192.168.14.4, 00:43:39, Ethernet0/0.14
      150.1.0.0/16 is variably subnetted, 4 subnets, 2 masks
C        150.1.15.0/24 is directly connected, Ethernet0/0.15
L        150.1.15.1/32 is directly connected, Ethernet0/0.15
C        150.1.16.0/24 is directly connected, Ethernet0/0.16
L        150.1.16.1/32 is directly connected, Ethernet0/0.16
      192.168.12.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.12.0/24 is directly connected, Ethernet0/0.12
L        192.168.12.1/32 is directly connected, Ethernet0/0.12
      192.168.14.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.14.0/24 is directly connected, Ethernet0/0.14
L        192.168.14.1/32 is directly connected, Ethernet0/0.14
O     192.168.23.0/24 [110/20] via 192.168.12.2, 00:43:39, Ethernet0/0.12
O     192.168.24.0/24 [110/20] via 192.168.14.4, 00:43:39, Ethernet0/0.14
                      [110/20] via 192.168.12.2, 00:43:39, Ethernet0/0.12
O     192.168.34.0/24 [110/20] via 192.168.14.4, 00:43:39, Ethernet0/0.14
```

همانطور که در خروجی دستور **show ip route** نیز مشاهده می کنیم, برای رسیدن به Prefix 3.3.3.3/32 دو مسیر best وجود دارد یعنی دو مسیر با cost یکسان, پس انتظار می رود که برای رسیدن به Prefix موردنظر لود بالانسینگ داشته باشیم, اما آیا این لود بالانسینگ برای فورواد پکت های Label خورده نیز صدق می کند؟ بیایید این موضوع را با traceroute بررسی نماییم:

```
PE1#traceroute 3.3.3.3
Type escape sequence to abort.
Tracing the route to 3.3.3.3
VRF info: (vrf in name/id, vrf out name/id)
  1 192.168.12.2 [MPLS: Label 20 Exp 0] 2 msec
    192.168.14.4 [MPLS: Label 19 Exp 0] 0 msec
    192.168.12.2 [MPLS: Label 20 Exp 0] 1 msec
  2 192.168.34.3 1 msec
    192.168.23.3 1 msec
    192.168.34.3 1 msec
```

یک خروجی فوق العاده! به راحتی اثبات شد که برای پکت های Label دار نیز لود بالانسینگ را خواهیم داشت چرا که اگر روتر پکتی Label دار دریافت نماید, برای فوروارد آن پکت به LFIB مراجعه می کند, در LFIB نیز تنها بهترین Label ها قرار دارند که این بهترین Label ها مربوط به بهترین مسیرهایی هستند که توسط روتینگ پروتکل در Routing Table قرار گرفته اند, پس اگر روتینگ پروتکل برای روتی دو مسیر را به عنوان بهترین مسیرها انتخاب کرده باشد و لود بالانسینگ را ساپورت کند, قطعا Label مربوط به هر دوی این مسیرها در LFIB قرار گرفته و هر دوی این مسیرها برای فوروارد پکت استفاده خواهند شد.

با همین چند قاعده ساده به راحتی توانستیم MPLS Unicast IP Forwarding را پیاده سازی نماییم. در بخش بعد به یک کاربرد مهم MPLS در ساختار شبکه های SP می پردازیم.

خلاصه بخش دوم:MPLS Unicast IP Forwarding

مفاهیم:

Label Distribution Protocol(LDP): پروتکل استاندارد Label گذاری پکت ها, مورد استفاده MPLS, تعریف شده در RFC 3036

Tag Distribution Protocol (TDP): پروتکل اختصاصی سیسکو برای Label گذاری پکت ها که قبل از LDP معرفی و استفاده می گردید.

Label-Switched Path (LSP): یک مسیر یک سویه بین یک یا چند LSR که برای Forward پکت های Label خورده مورد استفاده قرار می گیرد.

Label Switching Router (LSR): هر روتری که بتواند MPLS Label را تشخیص دهد.
Label Information Base (LIB): دیتابیسی شامل تمام Label هایی که یک LSR توسط LDP یا TDP فرا می گیرد.
Forwarding Information Base (FIB): دیتابیسی که از آن برای فوروارد پکت های بدون Label استفاده می شود.
Label FIB (LFIB): دیتابیسی که از آن برای فوروارد پکت های Label خورده استفاده می شود.

پیاده سازی:

- ✓ Enable CEF, MPLS, Define Label Protocol

```
(config)#ip cef
(config)#mpls ip
(config)#mpls label protocol [ldp | tdp]
```

- ✓ Configure IGP Protocol
- ✓ Enable LDP

```
In IGP Process(Just OSPF & ISIS):
(config-router)# mpls ldp autoconfig

In Interface:
(config-if)#mpls ip
```

دستورات بررسی:

Verify LDP is enabled	-	show mpls interfaces
Verify LDP sessions	-	show mpls ldp neighbor
Verify LIB(or LRIB)	-	show mpls ldp binding
Verify LFIB	-	show mpls forwarding- table
Troubleshooting Forming LDP Adjacencies(TCP Server/Client)	-	debug ip packet detail
Troubleshooting LDP Adjacencies	-	debug mpls ldp transport events

بخش سوم: MPLS VPN

MPLS VPN این امکان را به SP ها یا حتی شرکت های بزرگ می دهد تا خدمات لایه 3, VPN فراهم آورند. در اکثر مواقع SPها, خدمات لایه 2 قدیمی مانند ATM و Frame Relay را با سرویس MPLS VPN جایگزین می کنند. با استفاده از MPLS VPN, SP ها می توانند خدمات متنوعی را برای Customer های خود فراهم آورند.

همانطور که در بخش قبل کامل توضیح داده شد, در داخل کلود SP از MPLS Unicast IP Forwarding استفاده می شود, اما برای پیاده سازی MPLS VPN علاوه بر MPLS Unicast IP Forwarding به دو جز اساسی دیگر نیز احتیاج داریم:

1- **Virtual Routing and Forwarding(VRF)**: برای جداسازی اطلاعات routing مربوط به Customer های مختلف
2- **MP-BGP**: برای برقراری ارتباط بین ingress LSR و egress LSR, در این صورت فقط روترهای edge نیاز دارند تا اطلاعات routing مربوط به Customer ها را داشته باشند و روترهای داخل کلود MPLS تنها باید مسیر رسیدن ingress E-LSR به egress E-LSR را بدانند و فقط بر اساس Label, پکت ها را Forward خواهند کرد.

اما بیایید ابتدا با مفهوم VRF و علت استفاده ی MPLS VPN از این فیچر بپردازیم.

VRF چیست؟

زمانی که یک SP خدمات لایه 2, WAN را برای Customer هایش فراهم می آورد, برایش هیچ اهمیتی ندارد که این Customer ها از چه آدرس IP هایی استفاده می کنند, اما زمانی که این Customer ها به خدمات لایه 3, WAN مهاجرت می کنند, SP باید رنج های آدرسی که هر Customer استفاده می کند را فرا گیرد و آن ها را به شبکه داخلی خود advertise نماید, حتی اگر SP بتواند تمام رنج آدرس IP ها را از تمام Customer ها فرا گیرد, از آن جایی که تمام سازمان ها برای شبکه داخلی سازمان خود از آدرس های IP Private استفاده می کنند, قطعا با مشکل Duplicate Address IP روبه رو خواهیم شد.

MPLS VPN برای حل این مشکل از VRF Table(Virtual Routing & Forwarding) استفاده می کند. در واقع VRF یعنی استفاده از چندین Routing Table که با استفاده از آن ها می توان روت های Customer های مختلف را از هم جدا کرد و مشکل Duplicate Address IP را برطرف نمود.

به عبارت دیگر VRF را می توان Virtual Router ها یا بهتر, یک Virtual copy از Routing Table داخل IOS معرفی نمود.

نکته ممکن است با مفهومی با عنوان VRF Lite مواجه شوید. VRF Lite در واقع VRF ای است که به پیکربندی MPLS و BGP VPN v4و Route distinguisher(RD) و Route Target(RT)(با این مفاهیم در ادامه آشنا خواهید شد) احتیاجی ندارد و تنها کاری که انجام می دهد آن است که یک کپی از Routing Table تهیه می کند. تفاوت اصلی VRF Lite با VRF مورد استفاده در MPLS VPN آن است که در VRF-Lite تمام دیوایس هایی که در مسیر Transit قرار دارند باید قادر باشند تا تمام روتها در تمام VRF ها را از خود عبور دهند, یعنی در این صورت حتی روترهای داخل کلود SP هم نیاز دارند تا تمام اطلاعات مسیریابی مربوط به Customer ها را داشته باشند, اما در MPLS VPN فقط روترهای PE باید از اطلاعات routing مربوط به Customer ها, اطلاع داشته باشند. توجه داشته باشید که نه MPLS به VRF وابسته است و نه VRF به MPLS, اما ترکیب این دو فیچر با هم MPLS Layer 3 VPN را برای ما فراهم می آورد. در ادامه این بخش منظور ما از VRF, VRF مورد استفاده ی MPLS VPN خواهد بود.

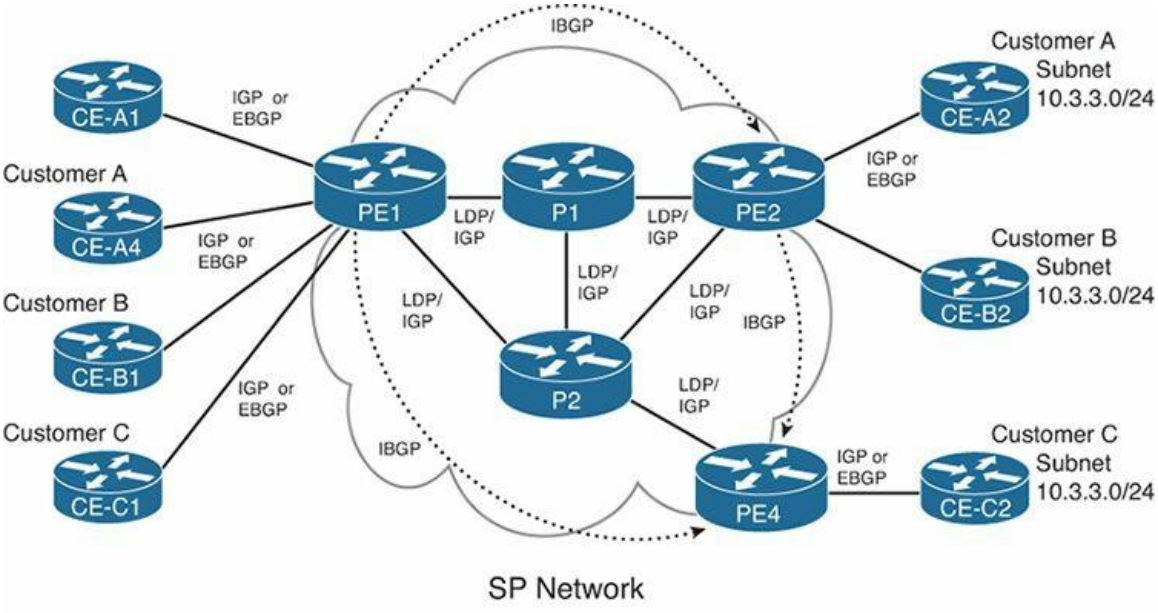
VRF سه جز اصلی دارد:

1- یک IP Routing Table(VRF RIB)
2- یک CEF FIB که بر اساس VRF RIB ساخته می شود
3- یک فرآیند مسیریابی برای مبادله Route ها با CE هایی که نیاز به پشتیبانی توسط VRF دارند

در ادامه با نحوه عملکرد VRF بیشتر آشنا خواهیم شد, اما قبل از آن بهتر است با سه اصطلاح که MPLS ازآن ها برای توصیف نقش روترهای شرکت کننده در ساختار MPLS VPN استفاده می کند, آشنا شویم:

- **Customer edge(CE)**: روتری که هیچ دانش و اطلاعی از پروتکل MPLS ندارد(یعنی روی آن MPLS پیاده سازی نشده است). این روتر هیچ پکت Label داری ارسال نمی کند, اما مستقیما به یک LSR یا همان PE در ساختار MPLS VPN متصل است.
- **Provider edge(PE)**: در قدیم از آن با عنوان LER هم یاد می شد, روتری که در لبه کلود MPLS قرار دارد و از یک سمت حداقل با یک CE و از سوی دیگر با P روترها در داخل کلود MPLS ارتباط دارد. PE باید قادر باشد تا هم IP Routing بر اساس FIB و هم MPLS Lookup بر اساس LFIB را انجام دهد. بر روی PE ها, iBGP و VRF Table را خواهیم داشت.
- **Provider(P)**: LSR ای که هیچ لینک ارتباطی با هیچ CE ای ندارد و در واقع تنها وظیفه ی آن Forward پکت های Label دار است و از روت های Customer ها هیچ آگاهی ندارد.

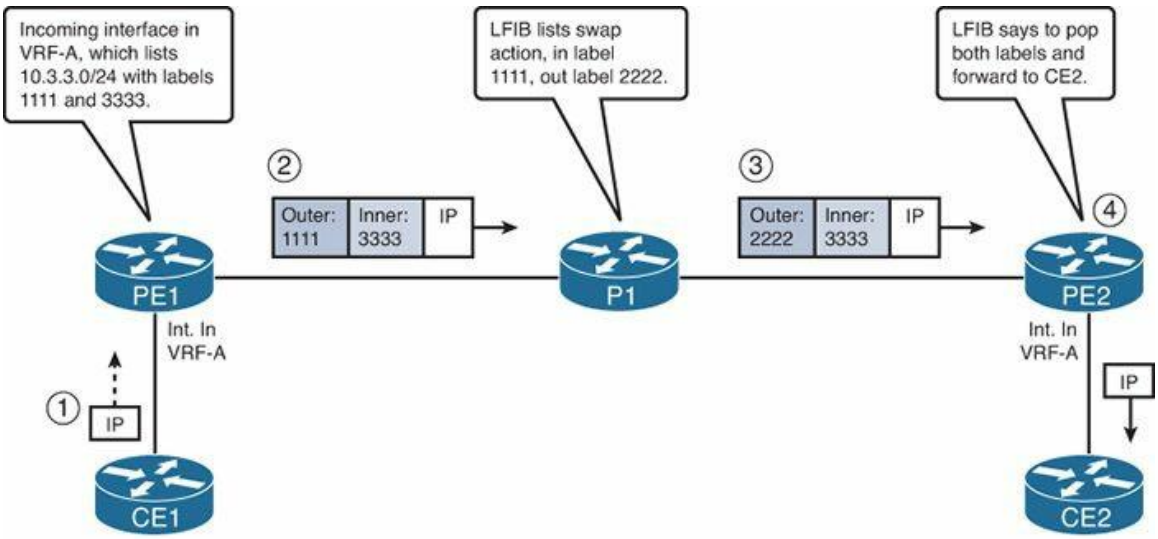
روترهای PE و CE برای آن که با هم ارتباط داشته باشند هم از IGP و هم از eBGP می توانند استفاده کنند. PE ها روت ها را از همسایه های CE خود فرا می گیرند اما برای آن که از Overlapping Prefix ها جلوگیری کنند, این روت ها را در داخل Routing Table معمول خود(که به آن Global Routing Table گوییم) قرار نمی دهند بلکه آن ها را در Routing Table های جداگانه ای که برای هر Customer در نظر گرفته اند که VRF نامیده می شوند, نگهداری می کنند. سپس روترهای PE از iBGP برای مبادله ی روت های Customer ها با سایر روترهای PE استفاده می کنند و هرگز این روت ها به P ها advertise نخواهد شد. این روندکلی اتفاقی است که در MPLS Layer3 VPN Control Plane رخ می دهد, تصویر زیر این روند را نمایش می دهد:



تصویر 14 بررسی اجمالی MPLS VPN Control Plane

اما MPLS VPN data plane باعث می شود که ingress PE ها دو Label به پکت اضافه کنند:

- هدر خارجی MPLS(با مقدار بیت S=0) همراه یک مقدار Label که بر اساس آن, پکت به سمت egress PE, ارسال خواهد شد.
- هدر داخلی MPLS(با مقدار بیت S=1) همراه با یک مقدار Label که VRF خروجی برای تصمیم گیری Forwarding را مشخص خواهد کرد(در واقع بر اساس این Label, egress PE متوجه می شود که باید به کدام VRF برای ارسال پکت, مراجعه کند.) تصویر زیر را در نظر بگیرید:



تصویر 15 بررسی اجمالی MPLS VPN Data Plane

مراحل شکل بالا به شرح ذیل می باشند:

1- CE1 یک پکت بدون Label را به PE1 ارسال می کند.

2- PE1 که یک پکت بر روی اینترفیسی که به VRF-A اختصاص پیدا کرده, دریافت می کند, مقصد پکت(10.3.3.3) را در VRF-A CEF FIB که اساس آن VRF-A Routing Table است, جستجو می کند و بر اساس entry ای که در VRF-A CEF FIB پیدا می کند, دو Label به پکت اضافه کرده و آن را به P1 ارسال می کند.

3- P1 دقیقا همانند زمانی که Unicast IP Forwarding داشتیم عمل می کند, یعنی پکت Label دار را دریافت کرده, در LFIB به دنبال entry مربوط به آن گشته و با عمل swap مواجه می شود, پس Label خارجی پکت را عوض کرده و آن را به سمت PE2 ارسال می کند.

4- PE2 بعد از دریافت پکت در LFIB اش به دنبال entry مربوط به Label 2222 می گردد, سپس در می یابد که برای این Label عمل POP قید شده پس Label خارجی را حذف می کند و با Label داخلی مواجه می شود, دوباره LFIB خود را برای entry مربوط به Label 3333 جستجو می کند, خوب PE2 دوباره با عمل POP مواجه می شود هم چنین اینترفیس خروجی را نیز به دست می آورد, در نتیجه PE2 یک پکت بدون Label برای CE2 ارسال می کند.

تا اینجا به صورت اجمالی نحوه عملکرد Control Plane و Data Plane را بررسی کردیم, حال بیایید کمی دقیق تر عملکرد آن ها را بررسی کنیم.

MPLS VPN Control Plane:

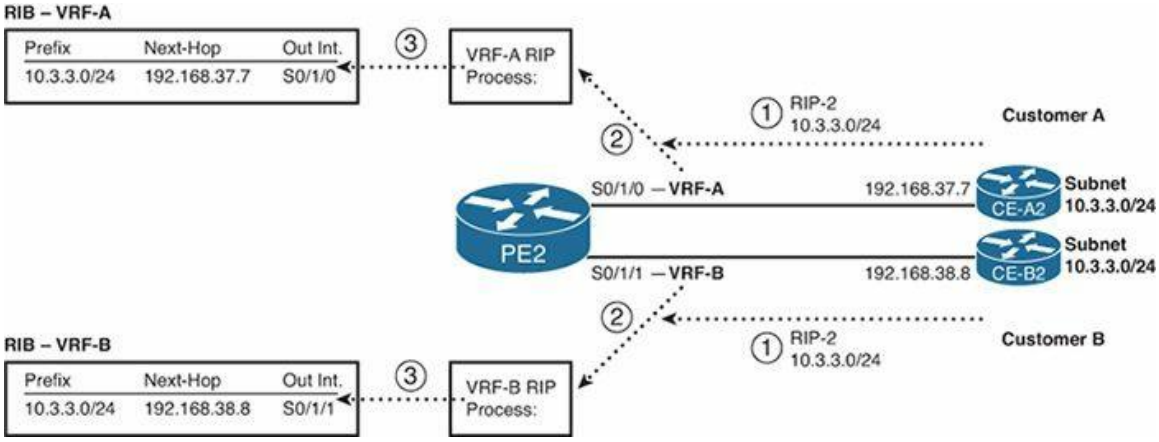
توسط MPLS VPN سه فیچر اصلی معرفی می گردد:

1- VRF
2- Route Distinguisher(RD)
3- Route Target(RT)

برای درک نحوه عملکرد Control Plan ابتدا این سه فیچر را بررسی می کنیم.

نحوه عملکرد VRF:

ما VRF را فقط بر روی PE ها در کلود MPLS خواهیم داشت. برای درک بهتر عملکرد VRF تصویر زیر را در نظر بگیرید:



تصویر 16 اضافه شدن روت فراگرفته از CE ها به VRF های مربوط به آن ها در روتر PE2

مراحل تصویر بالا به شرح ذیل می باشند:

1- روترهای CE که MPLS بر روی آن ها پیاده سازی نشده, با استفاده از پروتکل RIPv2 و به روش معمول, Prefix 10.3.3.0/24 را برای روتر PE2, advertise می کنند.
2- زمانی که PE2 یک RIP Update از روی اینترفیس se 0/1/0 دریافت می کند, که در واقع این اینترفیس به VRF-A اختصاص داده شده, فرآیند پردازش آپدیت RIP را برای روت دریافت شده انجام می دهد. روتر PE2 برای هر VRF یک فرآیند RIP جداگانه در نظر می گیرد. دقیقا همین عمل پردازش آپدیت RIP برای پیام آپدیت دریافت شده از روی اینترفیس se 0/1/1 که به VRF-B اختصاص یافته نیز, صورت می گیرد.
3- فرآیند مربوط به VRF-A ورودی 10.3.3.0/24 را به VRF-A RIB اضافه می کند. به طور مشابه همین عمل برای Prefix 10.3.3.0/24 مربوط به VRF-B نیز اتفاق می افتد.

توجه کنید که هر VRF یک FIB هم دارد که اطلاعاتش را از VRF RIB دریافت می کند, یعنی به ازای هر ورودی VRF RIB ما یک entry برای VRF FIB نیز خواهیم داشت که در تصویر بالا نمایش داده نشده است.

Route Distinguisher و نحوه ی عملکرد آن:

حال که روتر PE2 روت ها را از روترهای CE دریافت کرده, باید آن ها را به سایر PE ها, advertise کند. برای آن که مشکل Overlapping Prefix پیش نیاید, MPLS یک عدد به ابتدای BGP NLRI(یا همان Prefix) اضافه می کند. هر عدد به یک Customer اختصاص دارد و همین امر باعث تمایز NLRI ها می شود. حال برای این که MPLS بتواند این عدد را به اول NLRI اضافه کند از MP-BGP(**RFC 4760**) بهره می گیرد, که در واقع به ما امکان بازتعریف فیلد NLRI در BGP Update را می دهد. این عددی که به ابتدای NLRI اضافه می گردد براساس **RFC 4364** (**BGP/MPLS IP Virtual Private Networks (VPNs)**) نامیده می شود **Route Distinguisher** ,

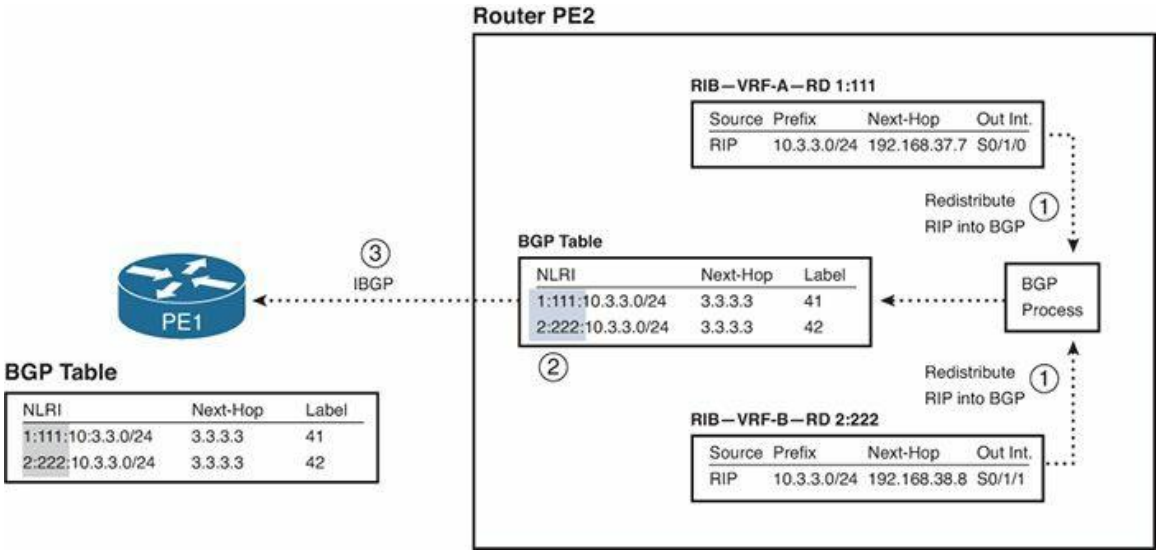
پس RD به ما این اجازه را خواهد داد که هم advertise و هم تمایز بین Duplicate IPV4 Prefix ها را داشته باشیم. درک این مفهوم ساده است:

هر NLRI همانند یک IPV4 Prefix معمولی advertise می شود با این تفاوت که یک عدد(RD) که باعث منحصر به فرد شدن این NLRI و شناسایی آن می گردد, به اول NLRI اضافه می شود, به این فرمت جدید NLRI, VPNv4 NLRI(or VPNv4 Route) گفته می شود که شامل دو بخش است:

- 64 بیت RD
- 32 بیت IPv4 Prefix

تصویر 17 VPNv4 NLRI

برای بررسی این موضوع که چطور RD به ابتدای یک NLRI اضافه می شود, تصویر زیر را در نظر بگیرید:



تصویر 18 یکتا شدن یک Prefix با استفاده از RD

در تصویر بالا PE2 از MP-BGP برای advertise, Prefix های 10.3.3.0/24 به PE1 استفاده می کند. همانطور که در شکل هم نشان داده شده, BGP Update یک فرم خاص و جدید را برای NLRI ها نمایش می دهد که در آن از RD 1:111 برای VPN-A و RD 2:222 برای VPN-B استفاده شده است.(3.3.3.3 که به عنوان Next-Hop در BGP Update وجود دارد, آدرس اینترفیس Loopback روتر PE2 می باشد.)

اگر در سناریوی نشان داده شده در تصویر بالا ازRD استفاده نشود, PE1 از دو (Prefix(10.3.3.0/24 که از همسایه ی iBGP اش دریافت کرده, تنها یکی را باید به عنوان بهترین روت انتخاب کرده و در BGP Table اش قرار دهد, اما با استفاده از RD, در واقع اکنون ما دو NLRI کاملا متفاوت داریم که به PE1 تبلیغ می شوند, در نتیجه PE1 هر دو NLRI را در BGP Table اش ذخیره خواهد نمود. مراحل شکل بالا به شرح ذیل می باشد:

1- PE2 پروتکل مسیریابی مربوط به هر VRF را داخل BGP, Redistribute می کند.
2- در طی این فرآیند Redistribution, RD مربوط به هر VRF به روت های داخل VRF RIB اضافه شده و سپس این روتها به همراه RD اضافه شده به آن ها, داخل BGP, Redistribute می شوند.
3- PE2 از iBGP برای advertise این روت ها به PE1 استفاده می کند. PE1 هر دو NLRI مربوط به Prefix 10.3.3.0/24 را در BGP Table اش قرار خواهد داد, چرا که آن ها توسط RD های در نظر گرفته شده برایشان, کاملا از هم متمایز می باشند.

RD, 8 بایت است که به فرم های زیر می توان آن را تعریف نمود:

- 2byte-integer:4-byte-integer
- 4byte-integer:2-byte-integer
- 4byte-dotted-decimal:2-byte-integer

مقدار قبل از colon یا باید شماره As باشد یا آدرس IP, اما مقدار بعد از colon می تواند هر عددی باشد.

Route Target و نحوه عملکرد آن:

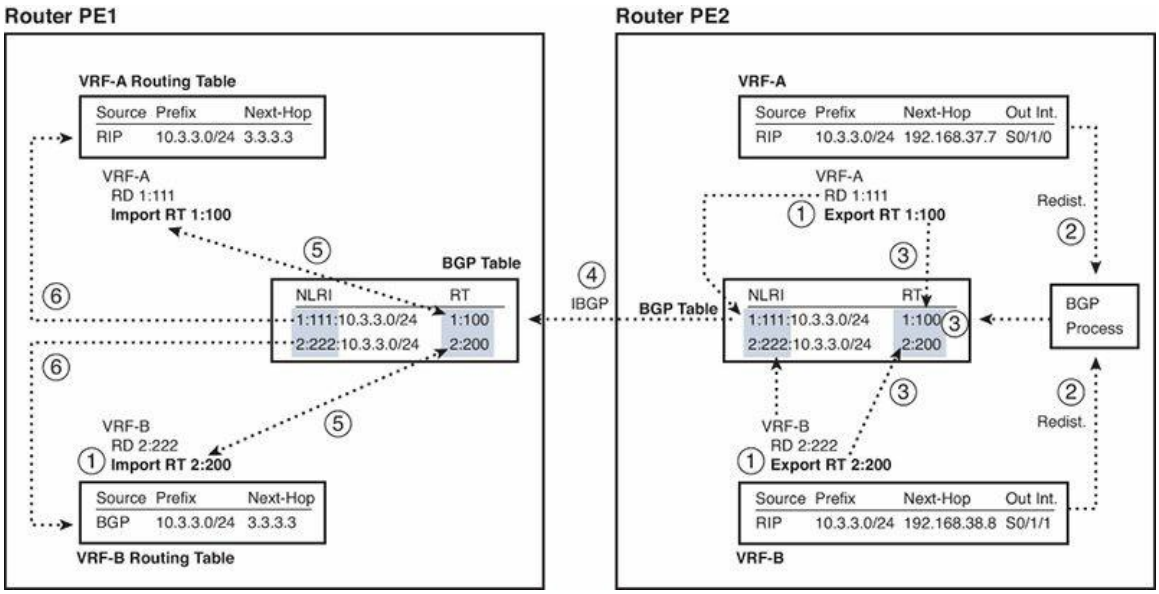
Route Target به MPLS این امکان را خواهد داد که توپولوژی های پیچیده ی VPN را پشتیبانی کند. PE ها, RT را در قالب یک BGP Extended Community PA, advertise می نمایند. BGP Extended Community ها دارای طول 8 بایت بوده و برای اهداف متنوعی مورد استفاده قرار می گیرند. MPLS دقیقا از همین BGP Extended Community برای encode کردن یک یا چند مقدار RT استفاده می کند. در واقع با استفاده از RT این جریان را کنترل می کنیم که چه روت های باید وارد/خارج VRF شوند. به بیان دیگر RT این موضوع را بیان می کند که چه روت هایی مربوط به چه سایت هایی می شوند.

برای این extended community دو پالیسی تعریف می شود:

1- **Export route-target**: که مشخص می کند چه روت هایی از VRF به داخل BGP, Redistribute شوند.
2- **Import route-target**: که مشخص می کند چه روت هایی از BGP به داخل VRF, Redistribute شوند.

فرمت مقداردهی به RT دقیقا همانند RD است, اما تفاوت RD و RT در این است که یک Prefix تنها می تواند یک مقدار برای RD داشته باشد, اما همین Prefix می تواند بیش از یک مقدار برای RT داشته باشد.

برای درک بهتر نحوه اضافه شدن RT به VPNv4 Route ها, تصویر زیر را در نظر بگیرید:



تصویر 19 MPLS Route Target(RT)

مراحل شکل بالا به شرح ذیل می باشد:

1- دو VRF روی روتر PE2 با یک Export Value, پیکربندی شده اند.
2- Redistribution روت های داخل VRF به درون BGP اتفاق می افتد.
3- طی فرآیند export, مقدار RT تنظیم شده برای هر VRF در BGP Table روتر PE2 قرار می گیرد.
4- PE2 روت ها را توسط iBGP برای PE1, advertise می کند.
5- PE1 ورودی های جدید BGP Table اش که در واقع همان VPNv4 Route هایی است که توسط PE2, advertise شده را بررسی می کند و مقدار RT این VPNv4 NLRI ها را با مقدار های RT import ای که برایش تعریف شده مقایسه می کند و متوجه می شود که این ورودی BGP Table به کدام VRF تعلق دارد.
6- PE1 بر اساس مطابقتی که بین RT های روت های دریافت شده و RT های پیکربندی شده بر روی خودش انجام می دهد, روت ها را به VRF های مربوطه Redistribute می کند.

توجه داشته باشید که پالیسی های import و export کاملا دو طرفه هستند, یعنی هر VRF هم به import و هم به export احتیاج دارد, اما در تصویر بالا فقط یک سمت این جریان نمایش داده شده است. جهت عکس تصویر بالا هم به این گونه خواهد بود که PE1 روت های مربوط به CE های کانکت به خود را با مقدار RT Export, داخل BGP, Redistribute کرده و آن ها را با iBGP برای PE2 ارسال می کند و دقیقا مراحل شرح داده شده در بالا برای PE2 رخ خواهد داد.

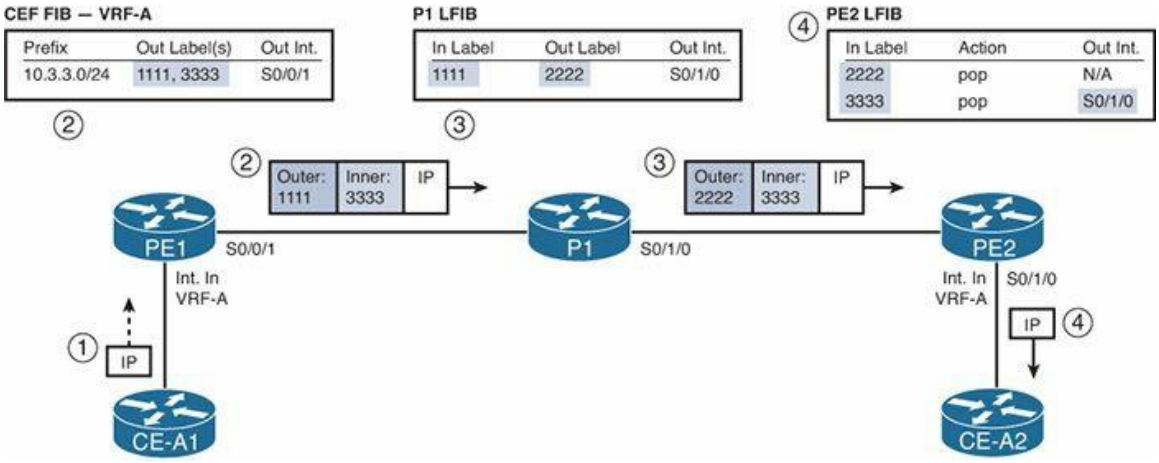
در اکثر پیاده سازی ها مقادیر RT مربوط به import و export یک VRF را یکسان در نظر می گیرند. و نکته بعد این که مقادیر VRF RD و VRF RT می توانند یکسان باشند.

با استفاده از این سه فیچر یعنی: VRF, RD و RT, نحوه عملکرد MPLS VPN Control Plane به خوبی روشن گردید, اما در ادامه به بررسی دقیقتر نحوه عملکرد MPLS VPN Data Plane خواهیم پرداخت.

MPLS VPN Data Plane:

VRF ها به PE ها اجازه می دهند که روت های مختلفی که از CE های متفاوت دریافت کرده اند را نگهداری نمایند, حتی اگر Prefix های CE ها, Overlap داشته باشند. RD به PE ها این امکان را می داد که Prefix را به صورت Unique بتوانند advertise کنند ونهایتا RT این امکان را برای PE ها فراهم می آورد که بتوانند تشخیص دهند کدام روت به کدام VRF تعلق دارد که به بیان بهتر این قابلیت را برای ما فراهم می آورد که بتوانیم به سایت ها از طریق چندین VPN دسترسی داشته باشیم.

حال برای Forward پکت ها, ingress PE ها نیاز دارد تا به FIB و LFIB خود مراجعه کند, در اینجا FIB و LFIB با آنچه در MPLS Unicast IP Forwarding مطرح شد کمی تفاوت دارد, تصویر زیر را در نظر بگیرید:



تصویر 20 نمایش FIB در ingress PE و LFIB در سایر روترها

مراحل شکل بالا به شرح ذیل می باشد:

1- یک پکت بدون Label روی یک اینترفیس که به VRF-A اختصاص پیدا کرده, دریافت می شود که باعث می شود ingress PE1 با استفاده از VRF-A FIB برای ارسال پکت تصمیم گیری نماید.

2- Ingress PE1 برای ورودی جدول VRF-A FIB یعنی Prefix 10.3.3.0/24 یک اینترفیس خروجی(out int در جدول) مشخص می کند(که می شود se 0/0/1) و یک Label برای آن تعیین می کند که این Label خود از دو بخش تشکیل می شود: یک Label خارجی که Outer Label نامیده می شود و در اینجا مقدار 1111 برای آن در نظر گرفته شده است و یک Label داخلی یا inner Label که می شود 3333. بنابراین PE1 این پکت را با دو Label ای که به اول هدر IP آن اضافه کرده, ارسال می کند.

3- P1 در LFIB خود به دنبال entry مربوط به Label ای با مقدار 1111 می گردد, با یافتن این entry با عمل swap مواجه می شود, پس مقدار Label خارجی را به 2222 تغییر داده و پکت را ارسال می کند.(توجه داشته باشید که P1 به محض مواجه شده با عمل swap خارجی ترین Label را تغییر داده و باافاصله پکت را ارسال می کند و از وجود Label داخلی خبر هم ندارد).

4- PE2 با جستجو در LFIB خود, entry مربوط به Label با مقدار 2222 را یافته و با عمل POP مواجه می شود, پس Label خارجی را حذف کرده, حال با یک Label دیگر مواجه می شود که همان inner Label است, پس دوباره در LFIB خود به دنبال entry مربوط به Label ای با مقدار 3333 می گردد, با پیدا کردن Entry مورد نظر با عمل POP مواجه می شود اما در این Entry اینترفیس خروجی هم برای ارسال پکت, مشخص شده است, پس PE2, Label را از پکت حذف کرده و پکت بدون Label را از طریق اینترفیس خروجی مشخص شده, ارسال می کند.

بر اساس آنچه شرح داده شد پس هدف کلی دو Label, Outer و Inner را به صورت زیر می توان بیان نمود:

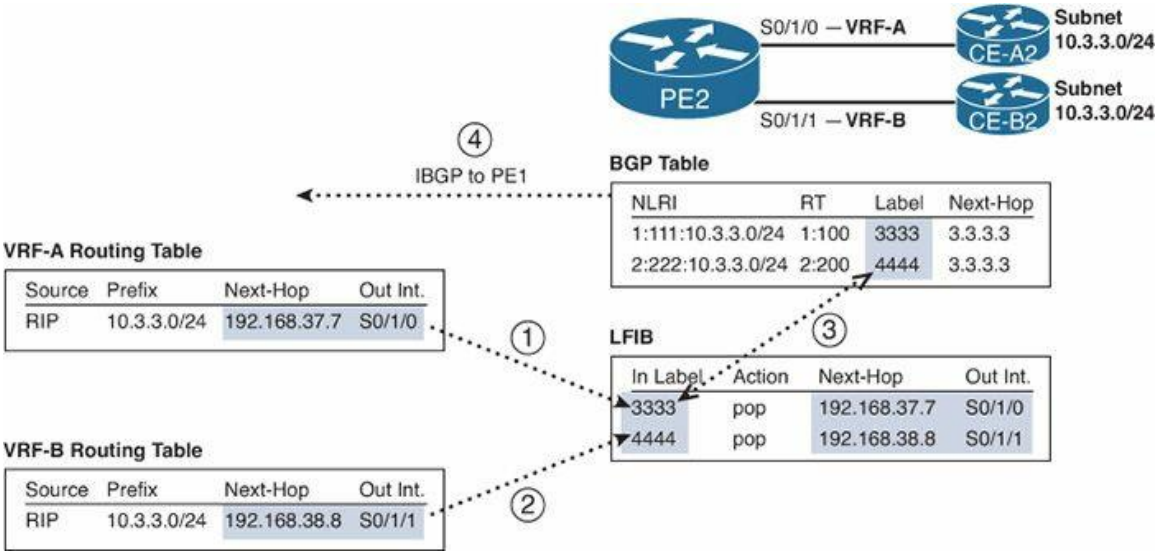
- **Outer Label**: این Label در واقع فقط مسیر بین ingress PE و egress PE را مشخص می کند(LSP) اما مشخص نخواهد کرد که egress PE چطور باید پکت را Forward کند.

- **Inner Label**: در واقع این Inner Label است که برای egress PE جزییات ارسال پکت را مشخص می کند(که مشخصا منظور تعیین اینترفیسی است که پکت بدون Label باید از طریق آن ارسال شود.)

این نحوه عملکرد کلی MPLS VPN Data Plane بود اما بیایید بررسی کنیم که چطور FIB و LFIB اطلاعات خود را به دست می آورند.

بررسی نحوه ایجاد LFIB در PE2:

تصویر زیر را در نظر بگیرید:



تصویر 21 نحوه ایجاد LFIB در Egress PE

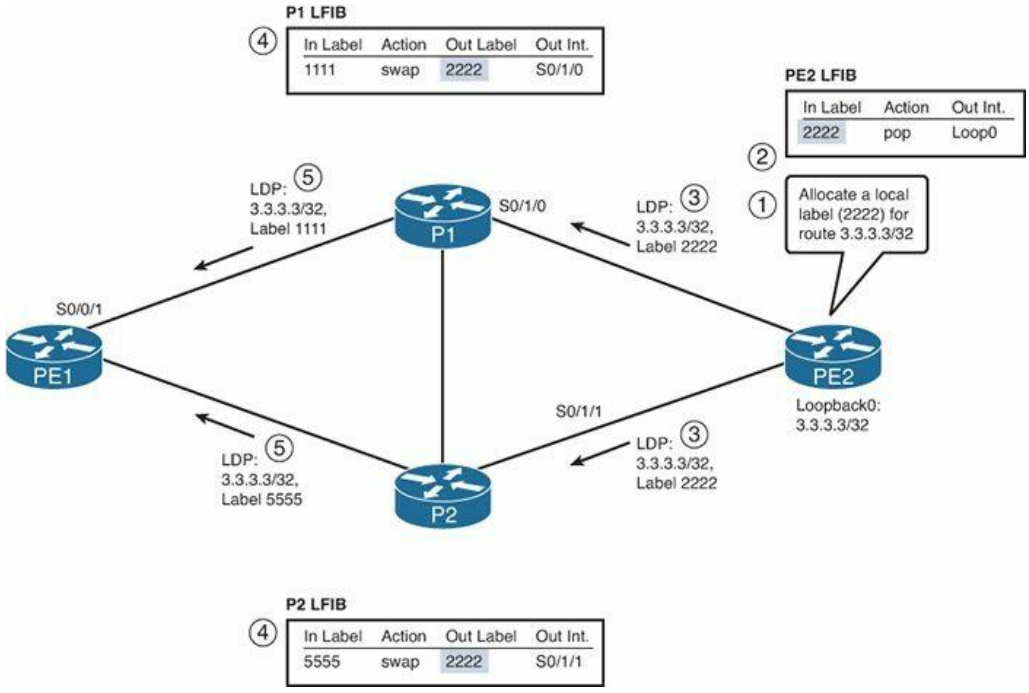
مراحل شکل بالا به شرح ذیل می باشد:

1- بعد از این که یک روت برای 10.3.3.0/24 در VRF-A ثبت می شود, PE2 یک Local Label(با مقدار 3333) به این Prefix اختصاص داده و این Local Label به همراه آدرس IP Next Hop و اینترفیس خروجی مربوط به آن روت را در LIB و LFIB قرار می دهد.
2- دقیقا آنچه در مرحله اول شرح داده شد برای 10.3.3.0/24 که از CE-B2 دریافت می شود نیز, اتفاق می افتد.(به روت مربوط به VRF-B, مقدار Local Label 4444 اختصاص یافته است.)
3- زمانی که این روت ها داخل BGP, Redistribute می شوند, روت ها به همراه Local Label های اختصاص یافته به آن ها در LFIB, در BGP Table ثبت می شوند.
4- PE2 از iBGP برای advertise روت ها به PE1 استفاده می کند و برای این کار نیز از یک BGP Update که شامل VPN Label های هر Prefix است, استفاده می کند.

خوب پس تا اینجا مشخص گردید که inner Label توسط PE ها مشخص شده و از طریق VPN v4 MP-BGP, advertise می شوند. اما Outer Label چگونه مشخص می شود؟ Outer Label در واقع مسیر بین ingress PE و egress PE را مشخص می کند, پس می توان این گونه بیان کرده که:

1- یک egress PE روتهایی را از تعدادی از Customer ها دریافت می کند.
2- Egress PE از iBGP برای advertise, VPNv4 Route ها(که شامل: NLRI/RD/RT/VPN Label هستند) به ingress PE استفاده می کند.
3- حال این روت های iBGP که فرا گرفته شده اند, یک آدرس IP Next-hop را مشخص می کنند.
4- برای این که MPLS VPN بتواند کار خود را انجام دهد, باید PE و P ها نحوه رسیدن به BGP Next Hop بعدی را بدانند
5- هم چنین برای این که MPLS VPN کار بکند باید برای هر روتی که به Next Hop منتهی می شود باید یک Label توسط LDP, advertise شود.

خوب با توجه به مثال بالا می بینیم که برای VPNv4 Route ها آدرس Next-Hop, 3.3.3.3 تعریف شده که در واقع همان آدرس IP اینترفیس Loopback روتر PE2 می باشد. کمی به عقبتر برگردید, زمانی که MPLS Unicast IP Forwarding را شرح می دادیم, داخل کلود SP ما برای فعال کردن MPLS نیاز به LDP داشتیم, شرط همسایگی LDP آن بود که همسایه های LDP حتما باید Reachable باشند, برای برقراری این شرط از یک IGP استفاده نمودیم, به محض این که همسایه ها قابل دسترس می شدند همسایگی LDP شکل می گرفت و Label ها برای Prefix های موجود در Routing Table توسط LDP بین همسایه ها advertise می گردید. حالا در همین ساختار که LDP فعال شده, ما iBGP را پیاده سازی می کنیم, در همسایگی iBGP شرط آن است که همسایه ها Reachable باشند خوب برای برقراری این شرط از IGP ها استفاده می کردیم, خوب پس Prefix مربوط به next-hop توسط IGP, advertise می گردد, از طرف دیگر LDP نیز همزمان این Prefix و Label اختصاص داده شده به آن را برای تمام همسایه های LDP روتر advertise می کند, پس می توان گفت outer Label بر اساس همان منطق MPLS Unicast IP Forwarding تعیین می شود.برای واضح تر شدن این مطلب شکل زیر را در نظر بگیرید:



تصویر 22 نحوه ی ایجاد LFIB Entry برای رسیدن به Egress PE's BGP Next Hop

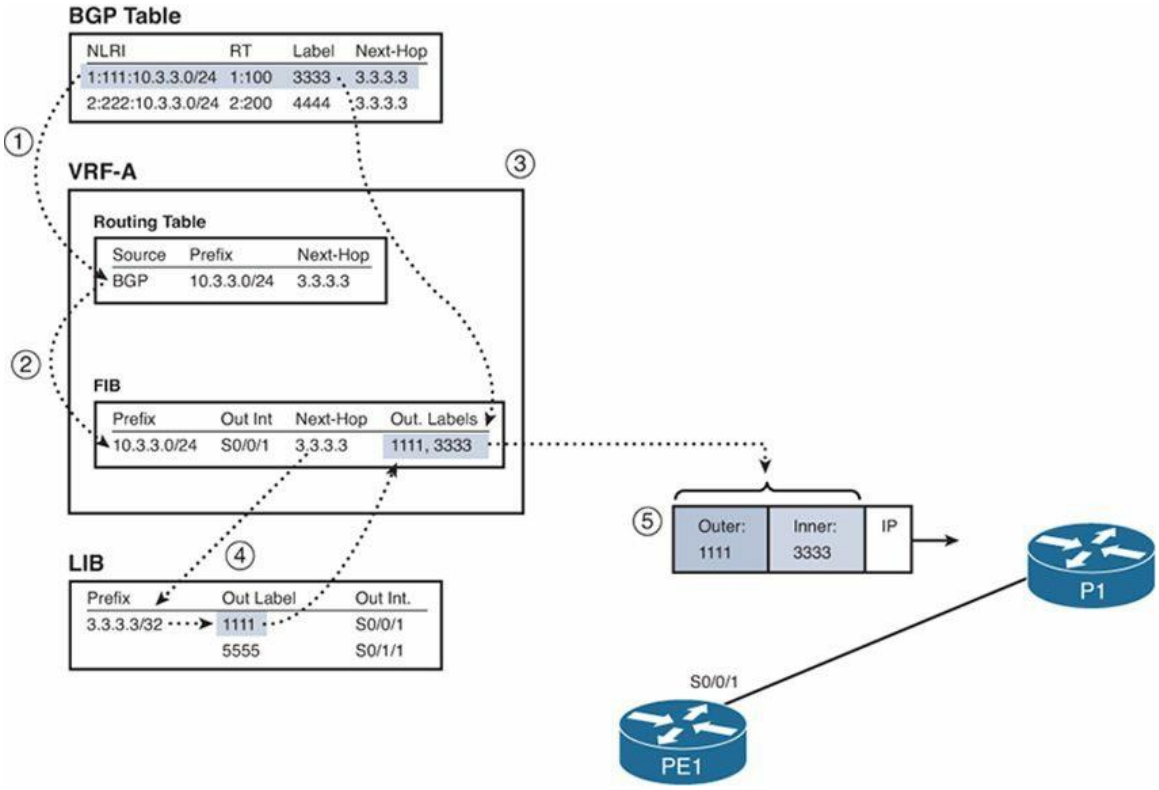
خوب تا اینجا نحوه ی ایجاد inner Label(VPN Label) و outer Label(Transit Label) را بررسی نمودیم , حال بیایید بررسی کنیم که VRF FIB در ingress PE چگونه ساخته می شود:

مراحل ایجاد VRF FIB در PE1:

منطق ارسال پکت های بدون Label در ingress PE به صورت زیر است:

1- پردازش پکتها بر اساس VRF اختصاص داده شده به اینترفیس ورودی
2- ارسال پکت ها بر اساس VRF FIB

ورودی جدول FIB برای آن که بتواند از MPLS VPN پشتیبانی کند به دو Label احتیاج دارد: یک Outer Label که بر اساس LDP آن را فراگرفته و در دیتابیسLIB اش قرار دارد, یک inner Label که از طریق iBGP Update از egress PE دریافت کرده و آن را در BGP Table خود دارد. برای درک بهتر تصویر زیر را در نظر بگیرید:



تصویر 23 ایجاد یک FIB Entry برای VRF-A در ingress PE1

تمام مراحل بالا زمانی اتفاق می افتد که روتر PE1 تمام پیش نیازهای BGP و اطلاعات LDP را فرا گرفته باشد, حالا می تواند VRF Routing Table و VRF FIB اش را بسازد. مراحل تصویر بالا به شرح ذیل می باشد:

1- PE1 براساس import RT متوجه می شود که هر Prefix به کدام VRF اختصاص دارد, پس Prefix مربوط به VRF-A را به داخل VRF-A RIB, redistribute می کند.
2- PE1 ورودی مربوط به VRF-A FIB را برای روتی که به VRF-A RIB اضافه شده, می سازد.
3- حال این ورودی FIB به یک VPN Label احتیاج دارد که آن را از BGP Table به دست می آورد.
4- هم چنین این ورودی FIB به یک outer Label احتیاج دارد که آن را از طریق LIB به دست می آورد(بهترین Label انتخاب خواهد شد.)
5- نهایتا روتر PE1, هدر MPLS را که شامل دو Label است, به اول پکت اضافه می کند.

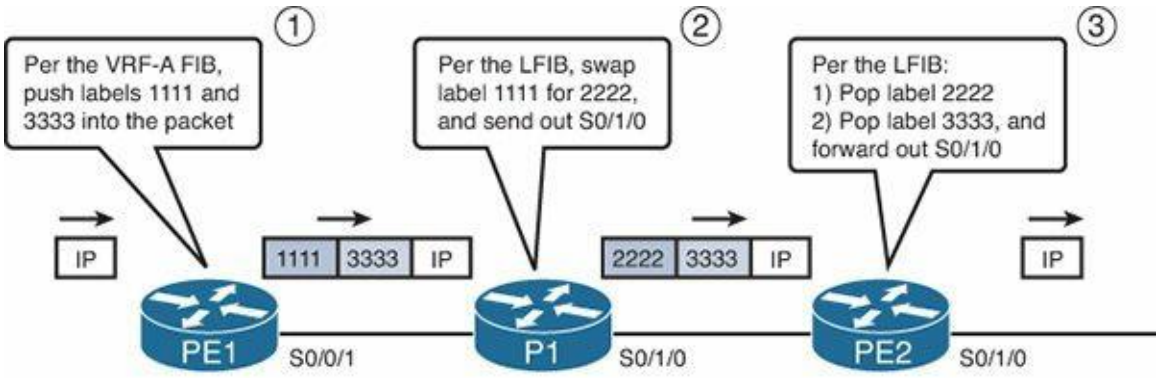
در نتیجه مراحل بالا, زمانی که PE1 یک پکت را از اینترفیس اختصاص یافته به VRF-A دریافت کند, ابتدا به VRF-A FIB اش نگاهی می اندازد, اگر این پکت برای مقصد 10.3.3.0/24 بود, PE1 آن را با ورودی FIB که در تصویر بالا نمایش داده شده, مطابقت داده و پکت را از طریق اینترفیس Se 0/0/1 و با دو Label, 1111 و 3333 ارسال می کند.

قبل از این که به سراغ پیکربندی عملی برویم و نکات ریز باقی مانده را مطرح کنیم, آخرین مفهوم باقی مانده از این بخش را شرح داده و بعد به سراغ سناریو می رویم.

Penultimate Hop Popping:

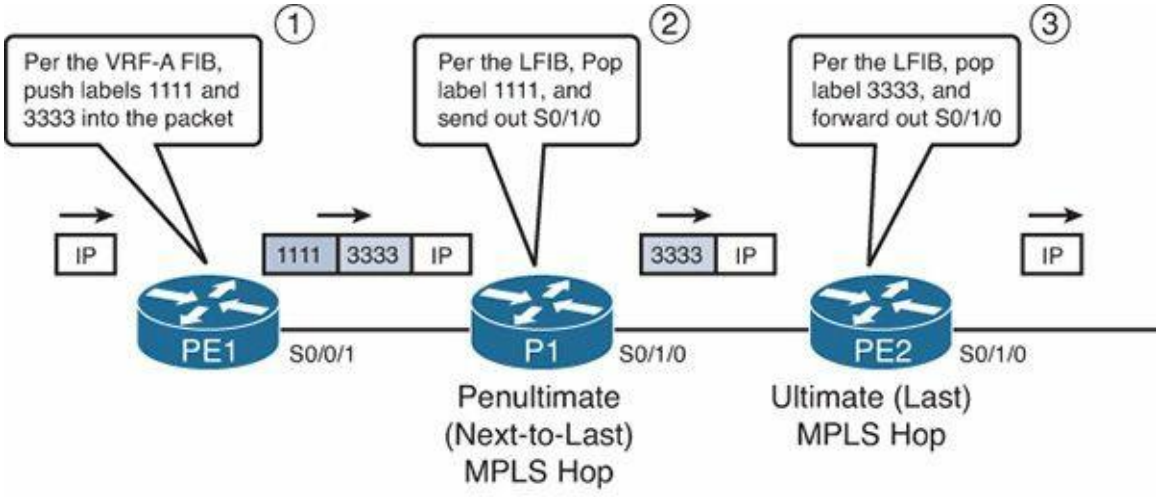
قبلا در بخش MPLS Unicast IP Forwarding اشاره مختصری در رابطه با این موضوع داشتیم که روترها روت های connected خود را با مقدار Label, imp-null برای همسایه های LDP خود advertise می کنند که این مقدار یعنی انجام عمل POP, یعنی روتر ماقبل این روتر اگر بخواهد روتی را برای این Prefix که مقدار Label آن imp-null ذکر شده ارسال کند, اول Label را از اول پکت IP حذف کند بعد پکت IP را ارسال نماید.

خوب حال در MPLS VPN با شرحی که داده شد, درست است که عملکرد Data Plane درست و کامل انجام می شود, اما فرآیندی که روی egress PE صورت می گیرد, کمی ناکارآمد است. این ناکار آمدی هم مربوط به این موضوع است که egress PE بعد از این پکتی با دو Label دریافت می کند, باید دو جستجو در LFIB اش انجام دهد, تصویر زیر را در نظر بگیرید:



تصویر 24 Two LFIB Lookups Required on the Egress PE

همانطور که در تصویر بالا هم مشاهده می شود, روتر egress PE مجبور است دو جستجو در LFIB اش داشته باشد, برای جلوگیری از انجام این عمل اضافی, از فیچری به نام Penultimate Hop Popping یا PHP استفاده می شود, عملی که این فیچر انجام می دهد آن است که به LSR یکی مانده به آخرین LSR(یعنی egress PE) بیان می کند که خارجی ترین Label را از اول Label Stack حذف نماید, بنابراین آخرین LSR(یعنی egress PE) پکتی را دریافت خواهد کرد که تنها یک Label دارد و آن هم VPN Label است, در نتیجه حالا egress PE فقط به یک جستجو در LFIB اش احتیاج دارد, همانطور که در تصویر زیر مشاهده می کنید:

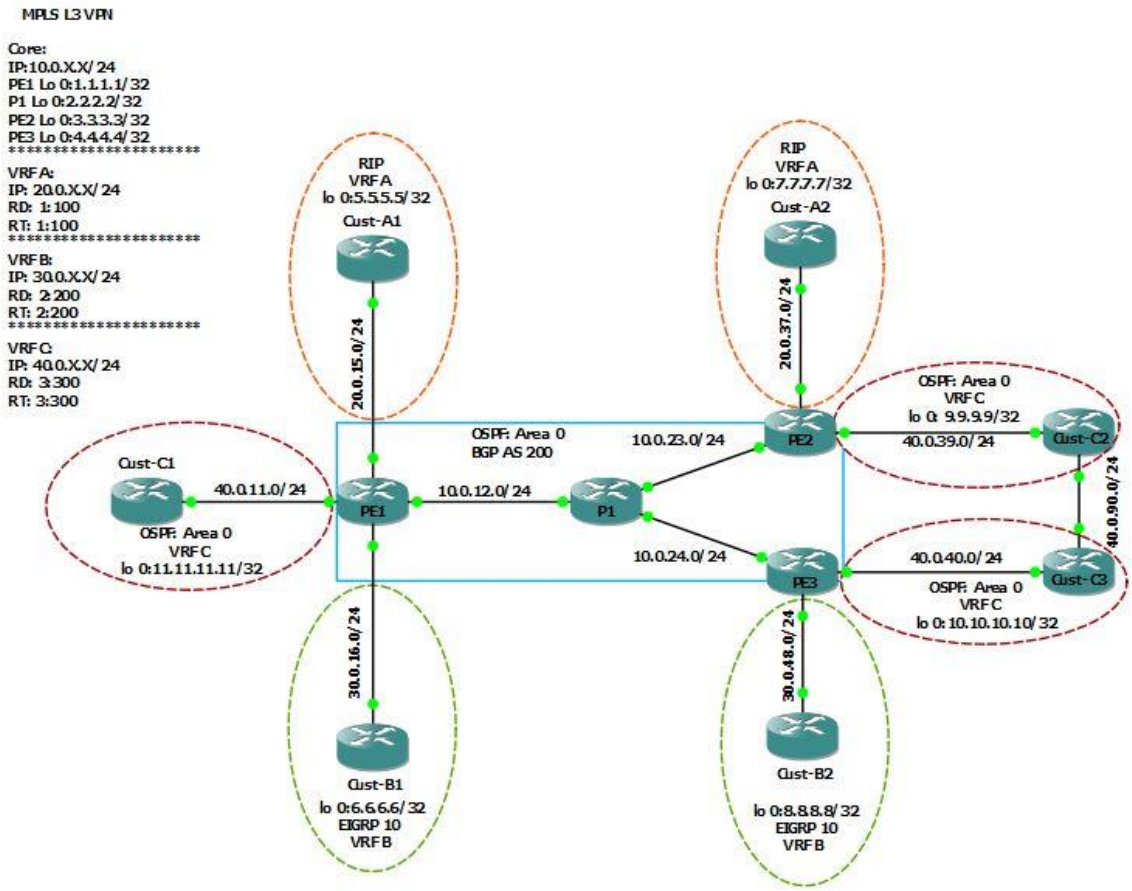


تصویر 25 Single LFIB Lookup on Egress PE Because of PHP

خوب حال برای درک بهتر آنچه تا کنون در این بخش مطرح شد به سراغ سناریوی خود خواهیم رفت.

پیاده سازی MPLS L3 VPN:

در این بخش برای آن که بتوانیم انواع حالت های مختلف پیاده سازی IGP ها بین PE و CE ها را نمایش دهیم, سناریویی همانند تصویر زیر را بررسی و پیاده سازی خواهیم نمود:



مراحل پیکربندی روترهای داخل کلود SP دقیقا مشابه آنچه در بخش MPLS Unicast IP Forwarding شرح داده شد می باشد. پس به سراغ پیکربندی فیچرهای MPLS L3 VPN خواهیم رفت.

مرحله اول, پیکربندی VRF ها و تعیین RD و RT هاست. به دو طریق می توان VRF را ایجاد کرد:

1- با استفاده از دستور: **ip vrf [name]**
2- با استفاده از دستور: **vrf definition [name]**

Vrf definition یک فرمت جدید در IOS های جدید می باشد که این امکان را می دهد که VRF همزمان هم IPv4 و هم IPv6 را ساپورت کند.

دستورات پیکربندی VRF ها بر روی روترهای PE سناریو به شرح زیر می باشد:

```
PE1#sho run | sec vrf
ip vrf Cust-A1
 rd 1:100
 route-target export 1:100
 route-target import 1:100
ip vrf Cust-B1
 rd 2:200
 route-target export 2:200
 route-target import 2:200
ip vrf Cust-C1
 rd 3:300
 route-target export 3:300
 route-target import 3:300
```

```
PE2#sho run | sec vrf
ip vrf Cust-A2
 rd 1:100
 route-target export 1:100
 route-target import 1:100
ip vrf Cust-C2
 rd 3:300
 route-target export 3:300
 route-target import 3:300
```

```
PE3#sho run | sec vrf
ip vrf Cust-B2
 rd 2:200
 route-target export 2:200
 route-target import 2:200
ip vrf Cust-C3
 rd 3:300
 route-target export 3:300
 route-target import 3:300
```

اگر پالیسی های import و export مربوط به RT یکسان در نظر گرفته شوند, به جای زدن دو دستور یکی برای import و دیگری برای export می توان از کامند: route-target both استفاده نمود, سپس خود IOS, این دو پالیسی را از هم متمایز خواهد کرد.

حال باید این VRF را به اینترفیس های متصل به Customer ها اعمال نماییم, نکته حائز اهمیت آن است که به محض اختصاص پیدا کردن یک VRF به یک اینترفیس, آدرس IPv4 برای آن پورت disable می گردد, علت هم آن است که آن اینترفیس از Global Routing Table به VRF Routing Table منتقل می شود و باید دوباره پورت را آدرس دهی کنیم:

```
PE2(config-vrf)#int fa 0/1
PE2(config-if)#ip vrf forwarding Cust-A2
% Interface FastEthernet0/1 IPv4 disabled and address(es) removed due to enabling VRF Cust-A2
PE2(config-if)#ip add 20.0.37.3 255.255.255.0
PE2(config-if)#
PE2(config-if)#int fa 1/0
PE2(config-if)#ip vrf forwarding Cust-C2
% Interface FastEthernet1/0 IPv4 disabled and address(es) removed due to enabling VRF Cust-C2
PE2(config-if)#ip add 40.0.39.3 255.255.255.0
```

دقیقا همین پیکربندی ها در PE1 و PE3 نیز انجام می شود. دقت داشته باشید اگر از دستور ip vrf forwarding زیر اینترفیس استفاده شود, در صورتی که آن لینک آدرس IPv6 هم داشته باشد, تنها آدرس IPv4 آن به VRF موردنظر انتقال می یابد و آدرس IPv6 آن هم چنان در Global Routing Table باقی می ماند و این امر به نوعی خطا در پیکربندی محسوب می شود چون آدرس IPv4 و IPv6 دیگر هیچ ارتباطی با هم ندارند, در چنین سناریوهایی که لینک هم آدرس IPv4 و هم آدرس IPv6 دارد یا باید از کامند: vrf definition استفاده شود یا CLI مربوط به VRF را با کامند زیر Upgrade کنیم:

```
(config)#vrf upgrade-cli multi-af-mode common-policies
```

حال برای بررسی آن که آیا اینترفیس های مورد نظر به VRF مربوطه اختصاص پیدا کرده اند یا نه از دستور: **show ip route vrf [name]** استفاده می کنیم:

```
Routing Table: Cust-A1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      20.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        20.0.15.0/24 is directly connected, FastEthernet0/1
L        20.0.15.1/32 is directly connected, FastEthernet0/1
PE1(config)#do sho ip route vrf Cust-B1

Routing Table: Cust-B1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      30.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        30.1.16.0/24 is directly connected, FastEthernet1/0
L        30.1.16.1/32 is directly connected, FastEthernet1/0
PE1(config)#do sho ip route vrf Cust-C1

Routing Table: Cust-C1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      40.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        40.0.11.0/24 is directly connected, FastEthernet1/1
L        40.0.11.1/32 is directly connected, FastEthernet1/1
```

```
PE1(config-if)#do sho ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      1.0.0.0/32 is subnetted, 1 subnets
C        1.1.1.1 is directly connected, Loopback0
      2.0.0.0/32 is subnetted, 1 subnets
O        2.2.2.2 [110/2] via 10.0.12.2, 01:18:13, FastEthernet0/0
      3.0.0.0/32 is subnetted, 1 subnets
O        3.3.3.3 [110/3] via 10.0.12.2, 01:17:32, FastEthernet0/0
      4.0.0.0/32 is subnetted, 1 subnets
O        4.4.4.4 [110/3] via 10.0.12.2, 01:13:55, FastEthernet0/0
      10.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
C        10.0.12.0/24 is directly connected, FastEthernet0/0
L        10.0.12.1/32 is directly connected, FastEthernet0/0
O        10.0.23.0/24 [110/2] via 10.0.12.2, 01:17:42, FastEthernet0/0
O        10.0.24.0/24 [110/2] via 10.0.12.2, 01:18:13, FastEthernet0/0
```

همانطور که در دو تصویر اخیر نیز مشاهده می کنید Prefix های مربوط به اینترفیس های اختصاص یافته به VRFها از Global Routing Table حذف شده و در VRF RIB قرار گرفته اند.

در این حالت اگر تست ping را به صورت معمول با Cust-A1داشته باشیم, ping نخواهیم داشت برای حل این مشکل باید ping را براساس VRF اختصاص یافته به آن لینک انجام دهید:

```
PE1#ping 20.0.15.5
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.15.5, timeout is 2 seconds:
.....
Success rate is 0 percent (0/5)
PE1#ping vrf Cust-A1 20.0.15.5
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 20.0.15.5, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 8/14/36 ms
```

PE-CE Routing with RIPv2:

در سناریوی بالا, سایت A برای برقرای ارتباط با PE ها باید از پروتکل RIP استفاده کند. اولین گام قبل از پیکربندی Routing Protocol تست ارتباط بین روتر PE و CE است و ارتباط egress PE و ingress PE می باشد, پس با Ping برقراری این ارتباط ها را چک کرده, در صورت عدم مشکل, RIP را پیکربندی می کنیم. در سمت CE ها پیکربندی RIP به صورت رایج و معمول است, اما در سمت PE ها از فیچر address family برای تمایز سایت های مختلف بهره می بریم(شاید همزمان به یک PE دو CE متصل باشند که هر کدام VRF خاص خود را دارند ولی هر دو برای ارتباط با PE از RIP استفاده می کنند, پس برای تمایز آن ها از هم از address family استفاده می شود.)

```
Cust-A1#sho run | sec rip
router rip
 version 2
 network 5.0.0.0
 network 20.0.0.0
 no auto-summary
```

```
PE1(config)#router rip
PE1(config-router)#
PE1(config-router)#address-family ipv4 vrf Cust-A1
PE1(config-router-af)#ver 2
PE1(config-router-af)#network 20.0.0.0
PE1(config-router-af)#no auto-summary
```

همین پیکربندی ها در سمت PE2 و Cust-A2 نیز انجام می شود:

```
Cust-A2#sho run | sec rip
router rip
 version 2
 network 7.0.0.0
 network 20.0.0.0
 no auto-summary
```

```
PE2(config)#router rip
PE2(config-router)#
PE2(config-router)#address-family ipv4 vrf Cust-A2
PE2(config-router-af)#ver 2
PE2(config-router-af)#network 20.0.0.0
PE2(config-router-af)#no auto-summary
```

برای بررسی آن که آیا روت ها از طریق RIP فراگرفته شده اند یا نه, از دو دستور زیر می توان استفاده نمود:

1- **Show ip rip database vrf [name]**
2- **Show ip route vrf [name]**

```
PE1#sho ip rip database vrf Cust-A1
5.0.0.0/8    auto-summary
5.5.5.5/32
    [1] via 20.0.15.5, 00:00:24, FastEthernet0/1
20.0.0.0/8    auto-summary
20.0.15.0/24    directly connected, FastEthernet0/1
```

```
PE2#sho ip route vrf Cust-A2

Routing Table: Cust-A2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      7.0.0.0/32 is subnetted, 1 subnets
R        7.7.7.7 [120/1] via 20.0.37.7, 00:00:02, FastEthernet0/1
      20.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        20.0.37.0/24 is directly connected, FastEthernet0/1
L        20.0.37.3/32 is directly connected, FastEthernet0/1
```

حال به سراغ پیکربندی MP-BGP VPNv4 و برقراری ارتباط همسایگی iBGP بین روترهای ingress PE و egress PE خواهیم رفت, در این حالت ما می خواهیم در آپدیت های BGP اطلاعات بیشتری شامل MPLS Label, Extended community ها و ... را داشته باشیم, پس در این حالت ما فقط IP-Unicast Forwarding ساده نخواهیم داشت, بلکه نیاز به session ها مختلف داریم, خوب MP-BGP این امکان را فراهم می آورد, پس چون در اینجا تمام همسایه های iBGP به یک نشست خاص(address-family) مربوط می شوند می توان خاصیت پیش فرض BGP که همسایگی iBGP را بر اساس IPv4 Unicast ساده در نظر می گرفت و روت هایی که از این طریق می آموخت را در Global Routing Table ذخیره می کرد, با دستور: **no bgp default ipv4-unicast** غیر فعال نماییم. پس پیکربندی BGP بر روی روترهای PE به صورت زیر خواهد بود:

```
PE1(config)#router bgp 200
PE1(config-router)#no bgp default ipv4-unicast
PE1(config-router)#neighbor 3.3.3.3 remote-as 200
PE1(config-router)#neighbor 3.3.3.3 update-source lo0
PE1(config-router)#neighbor 4.4.4.4 remote-as 200
PE1(config-router)#neighbor 4.4.4.4 update-source lo0
PE1(config-router)#address-family vpnv4
PE1(config-router-af)#neighbor 3.3.3.3 activate
PE1(config-router-af)#neighbor 4.4.4.4 activate
```

```
PE2(config)#router bgp 200
PE2(config-router)#no bgp default ipv4-unicast
PE2(config-router)#neighbor 1.1.1.1 remote-as 200
PE2(config-router)#neighbor 1.1.1.1 update-source lo0
PE2(config-router)#neighbor 4.4.4.4 remote-as 200
PE2(config-router)#neighbor 4.4.4.4 update-source lo0
PE2(config-router)#address-family vpnv4
PE2(config-router-af)#neighbor 1.1.1.1 activate
PE2(config-router-af)#neighbor 4.4.4.4 activate
```

```
PE3(config)#router bgp 200
PE3(config-router)#no bgp default ipv4-unicast
PE3(config-router)#neighbor 3.3.3.3 remote-as 200
PE3(config-router)#neighbor 3.3.3.3 update-source lo0
PE3(config-router)#neighbor 1.1.1.1 remote-as 200
PE3(config-router)#neighbor 1.1.1.1 update-source lo0
PE3(config-router)#address-family vpnv4
PE3(config-router-af)#neighbor 3.3.3.3 activate
PE3(config-router-af)#neighbor 1.1.1.1 activate
```

همانطور که در تصاویر بالا هم می بینید ما کامند: **neighbor x.x.x.x send-community extended** را استفاده نکردیم, اما به طور خودکار این کامند توسط IOS اضافه خواهد شد, نکته حائز اهمیت اینجاست که اگر بنا به هر دلیلی این کامند خورده نشده باشد یا حذف شده باشد, حتی اگر روتینگ پروتکل بین PE و CE را در BGP, redistribute هم بکنیم, باز هم هیچ روتی در VRF RIB ما ثبت نخوهد شد, چرا که مقدار RT که تعیین کننده آن است که هر روت به چه VRF ای بستگی دارد, توسط extended community موجود در پیام آپدیت BGP برای روتر مشخص می گردد.

برای بررسی آن که همسایگی BGP درست برقرار شده یا نه از دستور: **show bgp vpnv4 unicast all summary** استفاده می کنیم:

```
PE1#show bgp vpnv4 unicast all summary
BGP router identifier 1.1.1.1, local AS number 200
BGP table version is 1, main routing table version 1

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
3.3.3.3         4          200      18      18        1    0    0 00:13:53        0
4.4.4.4         4          200      18      18        1    0    0 00:13:33        0
```

```
PE3#show bgp vpnv4 unicast all summary
BGP router identifier 4.4.4.4, local AS number 200
BGP table version is 1, main routing table version 1

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
1.1.1.1         4          200      19      19        1    0    0 00:14:31        0
3.3.3.3         4          200      17      18        1    0    0 00:14:35        0
```

خوب حال نوبت به آن می رسد که در روترهای متصل به سایت های A, روت های فراگرفته شده از طریق RIP را داخل BGP, redistribute گردانیم. برای انجام این کار نیز, هم در سمت BGP هم در سمت RIP از address-family استفاده می گردد(برای جدا کردن session های مربوط به هر VRF):

```
PE1:
router bgp 200
address-family ipv4 vrf Cust-A1
redistribute rip
!
router rip
address-family ipv4 vrf Cust-A1
redistribute bgp 200 metric transparent
```

```
PE2:
router bgp 200
address-family ipv4 vrf Cust-A2
redistribute rip
!
router rip
address-family ipv4 vrf Cust-A2
redistribute bgp 200 metric transparent
```

زمانی که قرار است یک روتینگ پروتکل دیگر داخل RIP, Redistribute شود, حتما باید برای آن متریک تعریف نمود, حال این تعریف متریک یا می تواند به صورت Globally باشد با کامند default metric یا می تواند برای همان روتینگ پروتکل به سه صورت زیر تعریف گردد:

1- با ساب کامند: metric [0-15]
2- با استفاده از ساب کامند: metric transparent
3- با استفاده از ساب کامند: metric route-map

زمانی که از ساب کامند **metric transparent** استفاده می کنیم, یعنی روتینگ پروتکلی که قرار است روت را داخل RIP, Redistribute گرداند, هر متریکی که خود برای آن روت حساب کرده, همان را RIP به عنوان متریک قبول می کند. در اینجا قرار است روت های BGP, داخل RIP, Redistribute شوند, پس برای آن روت, BGP هر MED ای مشخص کرده باشد, همان مقدار برای متریک RIP در نظر گرفته می شود, برای درک بهتر, روتر PE1, Prefix 5.5.5.5/32 را با متریک 1 در VRF Cust-A RIB قرار داده است:

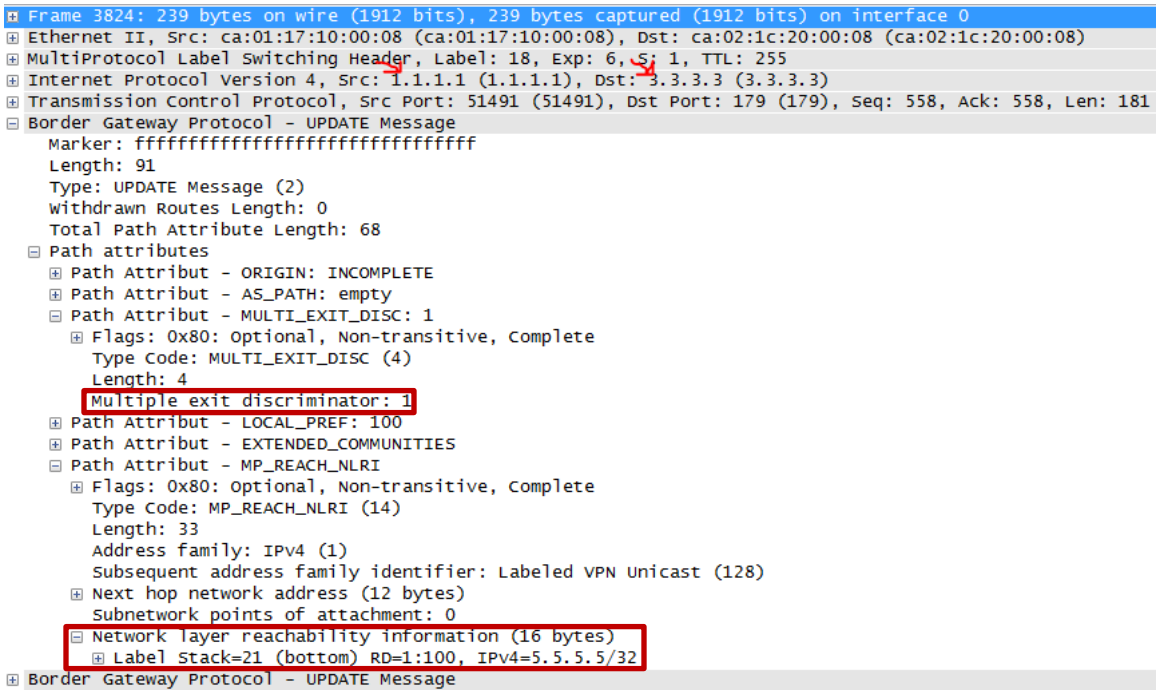
```
PE1(config-router)#do sho ip route vrf Cust-A1

Routing Table: Cust-A1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      5.0.0.0/32 is subnetted, 1 subnets
R        5.5.5.5 [120/1] via 20.0.15.5, 00:00:24, FastEthernet0/1
      20.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        20.0.15.0/24 is directly connected, FastEthernet0/1
L        20.0.15.1/32 is directly connected, FastEthernet0/1
```

هنگامی که روت های RIP داخل BGP, Redistribute می شوند, این متریک RIP, در داخل MED کپی می شود و روتر PE1 یک BGP Update برای PE2 ارسال می کند که در آن برای Prefix 5.5.5.5/32 مقدار MED, یک می باشد.



روتر PE2 این آپدیت را دریافت کرده و روت ها را در BGP Table خود قرار میدهد, زمانی که قرار است این روت ها را داخل RIP, Redistribute گرداند, همین مقدار MED به عنوان متریک RIP برای این روت در نظر گرفته می شود:

```
PE2#sho ip route vrf Cust-A2

Routing Table: Cust-A2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      5.0.0.0/32 is subnetted, 1 subnets
B        5.5.5.5 [200/1] via 1.1.1.1, 01:06:12
      7.0.0.0/32 is subnetted, 1 subnets
R        7.7.7.7 [120/1] via 20.0.37.7, 00:00:21, FastEthernet0/1
      20.0.0.0/8 is variably subnetted, 3 subnets, 2 masks
B        20.0.15.0/24 [200/0] via 1.1.1.1, 01:06:12
C        20.0.37.0/24 is directly connected, FastEthernet0/1
L        20.0.37.3/32 is directly connected, FastEthernet0/1
```

حال اگر نگاهی به جدول مسیریابی روتر Cust-A2 بیندازیم می بینیم که این Prefix را با متریک 2 در یافت نموده, و این یعنی هیچ کدام از روترهایی که در طی مسیر بین دو CE قرار گرفته اند در محاسبه متریک لحاظ نشده اند و مانند آن است که به نوعی ما ارتباط end-to-end داشته باشیم:

```
Cust-A2#sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      5.0.0.0/32 is subnetted, 1 subnets
R        5.5.5.5 [120/2] via 20.0.37.3, 00:00:00, FastEthernet0/0
      7.0.0.0/32 is subnetted, 1 subnets
C        7.7.7.7 is directly connected, Loopback0
      20.0.0.0/8 is variably subnetted, 3 subnets, 2 masks
R        20.0.15.0/24 [120/1] via 20.0.37.3, 00:00:00, FastEthernet0/0
C        20.0.37.0/24 is directly connected, FastEthernet0/0
L        20.0.37.7/32 is directly connected, FastEthernet0/0
```

مرحله آخر برای تست این که تمام تنظیمات درست بوده آن است که روترهای Cus-A1 و Cust-A2 بتوانند, Ping ساب نت های یکدیگر را داشته باشند پس:

```
Cust-A1#ping 7.7.7.7
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 7.7.7.7, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 40/46/60 ms
Cust-A1#traceroute 7.7.7.7
Type escape sequence to abort.
Tracing the route to 7.7.7.7
VRF info: (vrf in name/id, vrf out name/id)
  1 20.0.15.1 32 msec 16 msec 12 msec
  2 10.0.12.2 [MPLS: Labels 18/21 Exp 0] 44 msec 28 msec 36 msec
  3 20.0.37.3 [MPLS: Label 21 Exp 0] 32 msec 32 msec 32 msec
  4 20.0.37.7 48 msec 28 msec 36 msec
```

همانطور که در تصویر هم مشخص است, تست Ping موفقیت آمیز بود ☺

PE-CE Routing with EIGRP:

به سراغ پیکربندی ارتباطات روترهای PE با روترهای CE مربوط به سایت B می رویم.همانند RIP, پیکربندی EIGRP در سمت روترهای CE نیز به صورت معمول خواهد بود, اما در سمت PE ها باز هم از Address-family بهره خواهیم گرفت, در اینجا ارتباط PE1 با Cust-B1 را از طریق EIGRP Classic و ارتباط PE3 با Cust-B2 را از طریق Named EIGRP پیکربندی می کنیم:

```
Cust-B1#sho run | sec eig
router eigrp 10
 network 0.0.0.0

PE1#sho run | sec eigrp
router eigrp 1
 !
 address-family ipv4 vrf Cust-B1 autonomous-system 10
  network 0.0.0.0
 exit-address-family
```

```
Cust-B2(config)#router eigrp B2
Cust-B2(config-router)#address-family ipv4 unicast autonomous-system 10
Cust-B2(config-router-af)#network 0.0.0.0 255.255.255.255
```

```
PE3(config)#router eigrp PE3
PE3(config-router)#address-family ipv4 vrf Cust-B2 autonomous-system 10
PE3(config-router-af)#network 0.0.0.0 255.255.255.255
```

برای اطمینان از برقراری همسایگی از دستور: **show ip eigrp vrf [name] neighbor** استفاده می کنیم:

```
PE1#sho ip eig vrf Cust-B1 neig
EIGRP-IPv4 Neighbors for AS(10) VRF(Cust-B1)
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
0   30.0.16.6               Fa1/0             14 03:35:33   32    200  0  2
```

هم چنین با نگاهی به VRF Cust-B RIB از فراگیری روت های Cust-B مطمئن می شویم:

```
PE1#sho ip route vrf Cust-B1

Routing Table: Cust-B1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      6.0.0.0/32 is subnetted, 1 subnets
D        6.6.6.6 [90/156160] via 30.0.16.6, 02:47:10, FastEthernet1/0
      30.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        30.0.16.0/24 is directly connected, FastEthernet1/0
L        30.0.16.1/32 is directly connected, FastEthernet1/0
```

```
PE3#sho ip route vrf Cust-B2

Routing Table: Cust-B2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      8.0.0.0/32 is subnetted, 1 subnets
D        8.8.8.8 [90/156160] via 30.0.48.8, 00:01:40, FastEthernet0/1
      30.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        30.0.48.0/24 is directly connected, FastEthernet0/1
L        30.0.48.4/32 is directly connected, FastEthernet0/1
```

در حالت Named Mode برای بررسی EIGRP Topology Table از دستور زیر استفاده می کنیم:

```
PE3#show eigrp address-family ipv4 vrf Cust-B2 topology
EIGRP-IPv4 VR(PE2) Topology Table for AS(10)/ID(30.0.48.4)
           Topology(base) TID(0) VRF(Cust-B2)
Codes: P - Passive, A - Active, U - Update, Q - Query, R - Reply,
       r - reply Status, s - sia Status

P 30.0.48.0/24, 1 successors, FD is 28160
        via Connected, FastEthernet0/1
P 8.8.8.8/32, 1 successors, FD is 156160
        via 30.0.48.8 (156160/128256), FastEthernet0/1
```

حال باید redistribution را انجام دهیم:

```
PE1(config)#router bgp 200
PE1(config-router)#address-family ipv4 vrf Cust-B1
PE1(config-router-af)#redistribute eigrp 10
PE1(config-router-af)#
PE1(config-router-af)#router ei 1
PE1(config-router)#address-family ipv4 vrf Cust-B1
PE1(config-router-af)#redistribute bgp 200
```

```
PE3(config)#router bgp 200
PE3(config-router)#address-family ipv4 vrf Cust-B2
PE3(config-router-af)#redistribute eigrp 10
PE3(config-router-af)#
PE3(config-router-af)#router ei PE3
PE3(config-router)#address-family ipv4 vrf Cust-B2 autonomous-system 10
PE3(config-router-af)#topology base
PE3(config-router-af-topology)#redistribute bgp 200
```

همانطور که در دستورات بالا مشاهده می کنید, برای redistribute bgp داخل eigrp از هیچ متریکی بهره نگرفته ایم در صورتی که اگر بخواهیم هر روتینگ پروتکی را داخل eigrp, redistribute گردانیم باید حتما یا Default metric تعریف کنیم یا حداقل برای آن روتینگ پروتکل, متریک پارامترهای محاسبه متریک را تعیین کنیم, در غیر این صورت روتهایی که از آن روتینگ پروتکل داخل eigrp, redistribute شده بودند داخل Routing Table قرار نمی گیرند اما با نگاهی به VRF RIB در روتر PE1 و PE3 می بینیم که روت های redistribute شده از bgp به داخل EIGRP درون VRF RIB ثبت شده اند!

```
PE3(config)#do sho ip route vrf Cust-B2

Routing Table: Cust-B2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      6.0.0.0/32 is subnetted, 1 subnets
B        6.6.6.6 [200/156160] via 1.1.1.1, 00:03:28
      8.0.0.0/32 is subnetted, 1 subnets
D        8.8.8.8 [90/156160] via 30.0.48.8, 00:02:55, FastEthernet0/1
      30.0.0.0/8 is variably subnetted, 3 subnets, 2 masks
B        30.0.16.0/24 [200/0] via 1.1.1.1, 00:03:28
C        30.0.48.0/24 is directly connected, FastEthernet0/1
L        30.0.48.4/32 is directly connected, FastEthernet0/1
```

```
PE1(config)#do sho ip route vrf Cust-B1

Routing Table: Cust-B1
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      6.0.0.0/32 is subnetted, 1 subnets
D        6.6.6.6 [90/156160] via 30.0.16.6, 03:38:13, FastEthernet1/0
      8.0.0.0/32 is subnetted, 1 subnets
B        8.8.8.8 [200/156160] via 4.4.4.4, 00:07:29
      30.0.0.0/8 is variably subnetted, 3 subnets, 2 masks
C        30.0.16.0/24 is directly connected, FastEthernet1/0
L        30.0.16.1/32 is directly connected, FastEthernet1/0
B        30.0.48.0/24 [200/0] via 4.4.4.4, 00:07:29
```

اما چرا این اتفاق می افتد؟ روترهای EIGRP برای محاسبه متریک از فرمولی استفاده می کند که Composit metric یا متریک ترکیبی نامیده می شود, در این فرمول پارامترهای مختلفی شرکت دارند که از میان آن ها در حالت پیش فرض تنها از Bandwidth و Delay استفاده می شود(این پارامترها در ضرایبی که K-Value نامیده می شوند ضرب شده و چون ضرایب پارامترهای غیر از Bandwidth و Delay صفر می باشد, مقدار این پارامترها در محاسبه متریک تاثیری نخواهد داشت) برای مشاهده مقادیر این پارامترها برای هر روتی که در توپولوژی تیبل EIGRP قرار گرفته, می توان از دستور زیر استفاده نمود(مثلا Prefix 6.6.6.6/32 روی روتر PE1):

```
PE1#show ip eigrp vrf Cust-B1 topology 6.6.6.6/32
EIGRP-IPv4 Topology Entry for AS(10)/ID(30.0.16.1) VRF(Cust-B1)
EIGRP-IPv4(10): Topology base(0) entry for 6.6.6.6/32
  State is Passive, Query origin flag is 1, 1 Successor(s), FD is 156160
  Descriptor Blocks:
  30.0.16.6 (FastEthernet1/0), from 30.0.16.6, Send flag is 0x0
      Composite metric is (156160/128256), route is Internal
      Vector metric:
        Minimum bandwidth is 100000 Kbit
        Total delay is 5100 microseconds
        Reliability is 255/255
        Load is 1/255
        Minimum MTU is 1500
        Hop count is 1
        Originating router is 6.6.6.6
```

حال این پارامترها زمانی که روت داخل BGP, redistribute می گردد, در قالب BGP Extended commiunity, encode شده و همراه BGP Update برای همسایه ی iBGP, ارسال می شوند, بیایید به روتر PE3 رفته و نگاهی به Prefix 6.6.6.6/32 که توسط BGP Update دریافت کرده بیاندازیم:

```
PE3#show bgp vpnv4 unicast vrf Cust-B2 6.6.6.6/32
BGP routing table entry for 2:200:6.6.6.6/32, version 4
Paths: (1 available, best #1, table Cust-B2)
  Not advertised to any peer
  Local
    1.1.1.1 (metric 3) from 1.1.1.1 (1.1.1.1)
      Origin incomplete, metric 156160, localpref 100, valid, internal, best
      Extended Community: RT:2:200 Cost:pre-bestpath:128:156160
        0x8800:32768:0 0x8801:10:130560 0x8802:65281:25600 0x8803:65281:1500
        0x8806:0:101058054
      mpls labels in/out nolabel/23
```

مقادیری که در تصویر بالا مشخص شده اند در واقع همان پارامترهای Vector metric می باشند, برای اطمینان بیایید کمی حساب و کتاب انجام دهیم ☺:

0X8800: مقدار Flag را مشخص می کند که در اینجا 32768 عنوان شده, این بدان معناست که روت باید internal در نظر گرفته شود, اگر همین مقدار 0 باشد یعنی روت external است.

0x8801: مقدار 10, نشان دهنده شماره AS می باشد, و مقدار 130560 نشان دهنده ی Delay است, اما در Vector metric این مقدار 5100 در نظر گرفته شده, خوب یک ضرب و تقسیم ساده ما را به نتیجه خواهد رساند, Delay ای که در بخش Vector metric مشخص شده در واقع مجموع Delay هاست که برای قرار گرفتن در فرمول متریک ترکیبی باید در رابطه زیر قرار گیرد, تا نهایتا بتوان Dealy نهایی را به دست آورد, حال 130560 همان delay نهایی است و ما باید به نوعی مجموع delay ها را به دست آوریم:

Delay=(256 * ها delay مجموع)/10

(130560*10)/256=1500

0x8802: مشخص کننده Relaiability می باشد, هگزا دسیمال عدد 65281=ff01 که در اینجا ff یعنی Relaiability, 100٪ است یعنی 255/255 و 01 یعنی Prefix مورد نظر یک گام دورتر است.

0x8803: مشخص کننده Load و MTU است, عدد 65281 می شود ff01 که ff یعنی 1/255 و 1500 هم نشان دهنده ی MTU است.

جدول زیر برای یادآوری مقادیر شرح داده شد می تواند مفید باشد:

EIGRP Attribute	Type	Usage	Value
General	0x8800	EIGRP General Route Information	Route Flag and Tag
Metric	0x8801	EIGRP Route Metric Information and AS	AS and Delay
	0x8802	EIGRP Route Metric Information	Reliability, Next Hop, and Bandwidth
	0x8803	EIGRP Route Metric Information	Reserve, Load, and Maximum Transmission Unit (MTU)
	0x8804	EIGRP External Route Information	Remote AS and Remote ID
External	0x8805	EIGRP External Route Information	Remote Protocol and Remote Metric

جدول 4 BGP extended communities for EIGRP PE-CE routing

آخرین مرحله برای اطمینان از برقراری ارتباط بین Customer های سایت B, بررسی روتینگ تیبل و تست Ping می باشد:

```
Cust-B1#sho ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      6.0.0.0/32 is subnetted, 1 subnets
C        6.6.6.6 is directly connected, Loopback0
      8.0.0.0/32 is subnetted, 1 subnets
D        8.8.8.8 [90/158720] via 30.0.16.1, 01:59:09, FastEthernet0/0
      30.0.0.0/8 is variably subnetted, 3 subnets, 2 masks
C        30.0.16.0/24 is directly connected, FastEthernet0/0
L        30.0.16.6/32 is directly connected, FastEthernet0/0
D        30.0.48.0/24 [90/30720] via 30.0.16.1, 01:59:09, FastEthernet0/0
```

همانطور که در تصویر بالا هم مشاهده می کنید, AD برای روت های فرا گرفته شده از Cust-B2 برابر 90 می باشد یعنی این روت به عنوان یک internal route در نظر گرفته شده است.

```
Cust-B2#ping 6.6.6.6
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 6.6.6.6, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 44/49/72 ms
Cust-B2#
Cust-B2#traceroute 6.6.6.6
Type escape sequence to abort.
Tracing the route to 6.6.6.6
VRF info: (vrf in name/id, vrf out name/id)
  1 30.0.48.4 8 msec 24 msec 20 msec
  2 10.0.24.2 [MPLS: Labels 16/23 Exp 0] 52 msec 40 msec 44 msec
  3 30.0.16.1 [MPLS: Label 23 Exp 0] 12 msec 32 msec 28 msec
  4 30.0.16.6 60 msec 56 msec 52 msec
```

حال اگر بین روترهای سایت B غیر از MPLS یک لینک ارتباطی دیگر هم باشد مثلا point to point connected باشند(به آن Backdoor Route گویند), EIGRP برای ارسال ترافیک از Backdoor Route استفاده خواهد کرد, برای این که بخواهیم لینک مربوط به MPLS, Primary و لینک Backdoor, به عنوان لینک Backup در نظر گرفته شود, می توان با تغییر Delay در محیط پیکربندی لینک مسیر Backup, این طرح را اجرا کرد.

نکته بعد از پیکربندی EIGRP و Redistribution حتما EIGRP Topology Table را بررسی کنید, اگر غیر از روت های connected, روتی در توپولوژی تیبل وجود نداشت و از پیکربندی درست تمام تنظیمات هم مطمئن هستید, یک دور VRF تیبل مربوطه را پاک کنید با استفاده از کامند: **clear ip route vrf [name] ***

PE-CE Routing with OSPF:

حال به سراغ برقراری ارتباط بین روترهای سایت C می رویم. پیکربندی OSPF بر روی روترهای CE به صورت معمول و روی روترهای PE با استفاده از Address-family می باشد:

```
Cust-C1(config)#router ospf 10
Cust-C1(config-router)#network 0.0.0.0 255.255.255.255 area 0
```

```
PE1(config)#router ospf 10 vrf Cust-C1
PE1(config-router)#
PE1(config-router)#int fa 1/1
PE1(config-if)#ip ospf 10 area 0
```

به همین ترتیب Cust-C2 و Cust-C3 و هم چنین روترهای PE2 و PE3 را نیز پیکربندی می کنیم.

همانطور که در پیکربندی های بالا نیز مشاهده می کنید, Process ID جدیدی بر روی روتر PE در نظر گرفته شده, علت هم آن است که ما برای ارتباطات داخل کلود MPLS, از OSPF به عنوان IGP استفاده کردیم, Process-ID یک

مقدار Local ای است و اهمیتی ندارد که یکسان باشد یا نه, اما برای آن که پردازش های مربوط به Customer ها را از پردازش های مربوط به Global Table متمایز کنیم, OSPF ای با P-ID جدید در نظر می گیریم.

برای بررسی آن که همسایگی برقرار شده و روت ها از جانب Customer در VRF مربوطه قرار گرفته یا نه, Routing Table را چک می کنیم:

```
PE2#sho ip route vrf Cust-C2

Routing Table: Cust-C2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
O        9.9.9.9 [110/2] via 40.0.39.9, 00:03:23, FastEthernet1/0
      40.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        40.0.39.0/24 is directly connected, FastEthernet1/0
L        40.0.39.3/32 is directly connected, FastEthernet1/0
```

گام بعدی redistribute, OSPF داخل BGP و بالعکس می باشد:

```
PE1(config)#router bgp 200
PE1(config-router)#address-family ipv4 vrf Cust-C1
PE1(config-router-af)#redistribute ospf 10
PE1(config-router-af)#
PE1(config-router-af)#router ospf 10
PE1(config-router)#redistribute bgp 200 subnet
```

دقیقا دستورات بالا در روترهای PE2 و PE3 نیز پیاده سازی می شود. حال بیایید نگاهی به Routing Table, Cust-C3 بیاندازیم و ببینیم آیا را فرا گرفته است یا نه:

```
Cust-C3#sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
O IA     9.9.9.9 [110/3] via 40.0.40.4, 00:02:27, FastEthernet0/0
      10.0.0.0/32 is subnetted, 1 subnets
C        10.10.10.10 is directly connected, Loopback0
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.40.4, 00:02:27, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.40.4, 00:02:27, FastEthernet0/0
O IA     40.0.39.0/24 [110/2] via 40.0.40.4, 00:02:27, FastEthernet0/0
C        40.0.40.0/24 is directly connected, FastEthernet0/0
L        40.0.40.10/32 is directly connected, FastEthernet0/0
```

خوب همه چیز درست است اما همانطور که مشاهده می کنید روت های سایر Customer ها با مارک IA یعنی OSPF Inter Area دریافت شده اند, علت چیست؟

زمانی که از OSPF به عنوان روتینگ پروتکل بین PE ها و CE ها استفاده می کنیم, کلود MPLS به عنوان یک Super area 0 عمل خواهد کرد که به آن OSPF Super Backbone گویند. در این حالت دیگر نیازی نیست که حتما تمام سایت ها در area 0 باشند.

نکته در طراحی ها یک نکته را باید مد نظر قرار داد آن هم این است که در این حالت ما برای طراحی ساختار area ها باید یک طراحی سلسله مراتبی را رعایت کنیم, یعنی یا باید تمام CE های متصل به PE ها در Area 0 باشند یا همه باید در یک Area غیر 0 قرار داشته باشند. اگر در ساختاری روتر CE به یک روتر دیگر متصل باشد که در یک area دیگر قرار دارد, اگر قرار باشد روت های روتر پشت CE هم بتوان advertise نمود, باید روتر CE متصل به کلود MPLS حتما در area 0 باشد, اگر عکس این اتفاق بیافتد, یعنی روتر پشت CE در area 0 و روتر CE متصل به کلود MPLS در area ای غیر از area 0 باشد, در این صورت دیگر روت های روتر پشت CE به سایر سایت ها advertise نخواهد شد. پس اگر سناریویی این چنینی داشتیم دیگر استفاده از روش دوم پیشنهاد نمی گردد و بنابراین می توان گفت باید ساختار سلسله مراتبی در ارتباط با کلود MPLS به صورت زیر رعایت شود(روش اول):

ار دید روترهای CE, روترهای PE به عنوان ABR در نظر گرفته می شوند و تمام روتهای OSPF که داخل MP-BGP , redistribute می شوند به عنوان Inter-Area Route(IA) در نظر گرفته خواهند شد. MP-BGP برای عبور Prefix های OSPF در طول کلود MPLS از سه Extended Community بهره خواهد برد:

- **OSPF Domain-ID**: که در واقع همان Process-ID روی روتر PE, مربوط به پردازش OSPF بین PE و CE می باشد. خوب حال اگر تمام P-ID های OSPF یک VPN یکسان باشند, روت های OSPF با LSA Type 3, advertise شده و به عنوان Inter-Route در نظر گرفته خواهند شد, اما اگر این P-ID ها متفاوت باشد, در این صورت روت های OSPF با LSA Type 5, advertise خواهند شد و به عنوان External Route در نظر گرفته می شوند.

نکته Domain-ID را به صورت دستی نیز می توان تعیین کرد از طریق دستور زیر:

```
PE1(config-router)#domain-id ?
  A.B.C.D  OSPF domain ID in IP address format
  null     Null Domain-ID
  type     OSPF domain ID type in Hex format
```

- **OSPF Route Type**: سه فیلد اصلی دارد:
  1- Source Area
  2- Route Type
  3- Option

به فرم X:Y:Z نمایش داده می شود که X همان شماره Area, Y همان Rout type یا به بیان بهتر همان LSA Type می باشد. اما Z مشخص کننده Metric Type مربوط به LSA Type های 5 و 7 می باشد(E1, E2, N1, N2)

- **OSPF Router-ID**: که در واقع همان Router-ID روتر PE, در پردازش OSPF مربوط به ارتباط PE-CE را مشخص می کند.

خوب پس با استفاده از صفت MP-BGP Extended Community ما اطلاعات Prefix ها و LSA Type ها را در طول کلود MPLS منتقل خواهیم کرد, هم چنین از صفت MP-BGP MED برای حمل متریک OSPF بهره خواهیم گرفت.

زمانی که OSP با MP-BGP استفاده شود, از مکانیزم های جلوگیری از Loop مختلفی بهره می برد:

- تمام LSA Type 3 های Redistribute شده از MP-BGP داخل OSPF در هدر LSA خود فیلدی دارند به نام down bit که اگر این بیت set شود, یعنی این روت قبلا دریافت شده و روتر این روت را Drop خواهد کرد. این مکانیزم از آن جهت استفاده می شود که فرض کنیم PE یک روت redistribute شده داخل OSPF را به یک CE ارسال می کند, و آن CE نیز آن روت را از طریق یک Backdoor Path برای یک CE دیگر Advertise می کند و این CE هم دوباره این روت را به یک PE دیگر در کلود MPLS, advertise کرده, در نتیجه این روتر PE به این نتیجه میرسد که روت OSPF ای که از CE دریافت کرده با AD=110 خیلی بهتر از روت iBGP است که AD=200 است پس در نتیجه دوباره همین روت را به PE ای که اول روت را generate کرده بود ارسال می کند و این یعنی Loop. پس down bit برای یک PE مشخص می کند که روت های که Down bit آن ها set شده هرگز نباید دوباره از OSPF به داخل MP-BGP, Redistribute گردند. دقیقا این رفتار مانند آن است که ما دستی به روت ها Tag بزنیم. یک فیچر دیگر که دقیقا مرتبط با همین down bit می باشد, routing bit است که باعث می شود دامین OSPF یک Customer به عنوان Transit part برای شبکه MPLS در نظر گرفته نشود. این فیچر بیان می کند که تمام روت های که down bit برای آن ها set شده, باید routine-bit شانClear شود , در این صورت حتی اگر این روت توسط الگوریتم SPF به عنوان بهترین روت نیز انتخاب شود هرگز در Routing Table قرار نخواهد گرفت.
- مکانیزم بعدی جلوگیری از Loop مربوط به LSA Type های 5 می شود.زمانی که LSA Type 5 داریم دیگر Down bit برای ما کارایی نخواهد داشت چرا که در این حالت ما دامین های OSPF مختلف داریم و زمانی

که قرار است یک روت از یک دامین به دامین دیگر وارد شود, down bit آن Clear خواهد شد, پس دوباره احتمال این که یک PE روتی را از طریق MP-BGP دوباره برای یک PE داخل کلود MPLS ارسال کند پیش خواهد آمد. در اینجا می گوییم برای روتهایی با LSA Type 5 از فیلدی به نام tag Field استفاده می شود که مقدار آن برابر خواهد شد با شماره AS مر بوط به BGP, در نتیجه اگر یک روتر PE یک روت را دریافت کند که tag-Field آن مقداری برابر با BGP ASN فعال بر رویش داشته باشد, هرگز آن روت را به داخل MP-BGP, advertise نخواهد کرد.

برای کسب اطلاعات بیشتر **RFC 4577** را می توانید مطالعه نمایید.

خوب ابتدا نگاهی به رفتار پروتکل OSPF فعال شده روی PE1 در ارتباط با Cust-C1 بیاندازیم:

```
PE1#sho ip ospf 10
 Routing Process "ospf 10" with ID 40.0.11.1
   Domain ID type 0x0005, value 0.0.0.10
 Start time: 00:00:15.664, Time elapsed: 02:12:28.196
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
 Supports Link-local Signaling (LLS)
 Supports area transit capability
 Supports NSSA (compatible with RFC 1587)
 Connected to MPLS VPN Superbackbone, VRF Cust-C1
 Event-log disabled
 It is an area border and autonomous system boundary router
 Redistributing External Routes from,
    bgp 200, includes subnets in redistribution
 Router is not originating router-LSAs with maximum metric
 Initial SPF schedule delay 5000 msecs
 Minimum hold time between two consecutive SPFs 10000 msecs
 Maximum wait time between two consecutive SPFs 10000 msecs
 Incremental-SPF disabled
 Minimum LSA interval 5 secs
 Minimum LSA arrival 1000 msecs
 LSA group pacing timer 240 secs
 Interface flood pacing timer 33 msecs
 Retransmission pacing timer 66 msecs
 Number of external LSA 0. Checksum Sum 0x000000
 Number of opaque AS LSA 0. Checksum Sum 0x000000
 Number of DCbitless external and opaque AS LSA 0
 Number of DoNotAge external and opaque AS LSA 0
 Number of areas in this router is 1. 1 normal 0 stub 0 nssa
 Number of areas transit capable is 0
 External flood list length 0
 IETF NSF helper support enabled
 Cisco NSF helper support enabled
 Reference bandwidth unit is 100 mbps
    Area BACKBONE(0)
        Number of interfaces in this area is 1
        Area has no authentication
        SPF algorithm last executed 01:40:48.196 ago
        SPF algorithm executed 2 times
```

همانطور که در تصویر بالا هم مشخص شده, روتر PE1 به backbone Super متصل است و به عنوان ABR در نظر گرفته شده است.

خوب بیایید نگاهی به دیتابیس Cust-C1 بیاندازیم:

```
Cust-C1#sho ip ospf data

            OSPF Router with ID (11.11.11.11) (Process ID 10)

                Router Link States (Area 0)

Link ID         ADV Router      Age         Seq#       Checksum Link count
11.11.11.11     11.11.11.11     988         0x80000002 0x0063C6 2
40.0.11.1       40.0.11.1       989         0x80000002 0x0086D7 1

                Net Link States (Area 0)

Link ID         ADV Router      Age         Seq#       Checksum
40.0.11.1       40.0.11.1       989         0x80000001 0x00046D

                Summary Net Link States (Area 0)

Link ID         ADV Router      Age         Seq#       Checksum
9.9.9.9         40.0.11.1       949         0x80000001 0x00B9A8
10.10.10.10     40.0.11.1       949         0x80000001 0x008BD2
40.0.39.0       40.0.11.1       949         0x80000001 0x0096A1
40.0.40.0       40.0.11.1       949         0x80000001 0x008BAB
```

همانطور که در تصویر بالا هم مشخص گردیده, روتر Cust-C1 روتر PE1 را به عنوان ABR برای خود در نظر گرفته است(علت آن که Router-ID برای PE1, 40.0.11.1 که آدرس IP لینک متصل به Cust-C1 می باشد, آن است که روی روتر PE1 یک پردازش OSPF دیگر برای کلود MPLS فعال می باشد و دو پردازش OPSF همزمان نمی توانند یک Router-ID یکسان داشته باشند.)

حال نگاهی به آپدیت MP-BGP بر روی یک روتر PE می اندازیم:

```
PE3#show bgp vpnv4 unicast all 11.11.11.11/32
BGP routing table entry for 3:300:11.11.11.11/32, version 32
Paths: (1 available, best #1, table Cust-C3)
  Not advertised to any peer
  Local
    1.1.1.1 (metric 3) from 1.1.1.1 (1.1.1.1)
      Origin incomplete, metric 2, localpref 100, valid, internal, best
      Extended Community: RT:3:300 OSPF DOMAIN ID:0x0005:0x0000000A0200
        OSPF RT:0.0.0.0:2:0 OSPF ROUTER ID:40.0.11.1:0
      mpls labels in/out nolabel/25
```

- **Domain-ID**: x0000000**0A**02000 مشخص کننده OSPF P-ID می باشد که ما این P-ID را 10 در نظر گرفته بودیم(0A=10)
- **OSPF RT**: **0.0.0.0:2:0** مشخص کننده ی Area 0, LSA Type 2 و Metric Type 0 می باشد.
- **OSPF Router-ID**: مشخص کننده R-ID می باشد که برابر است با 40.0.11.1

OSPF Sham-Link:

دوباره نگاهی به سناریو بیاندازید, بین روترهای Cust-C2 و Cust-C3 یک ارتباط Point-to-Point داریم(که اصطلاحا Backdoor Path نامیده می شود) که فعلا آن را پیکربندی نکردیم. همانطور که در تصویر زیر مشاهده می کنید ارتباط این دو سایت از طریق MPLS VPN می باشد:

```
Cust-C2#sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O IA     10.10.10.10 [110/3] via 40.0.39.3, 03:27:38, FastEthernet0/0
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 02:57:57, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 02:57:57, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O IA     40.0.40.0/24 [110/2] via 40.0.39.3, 03:27:38, FastEthernet0/0
```

اگر ارتباط OSPF بین Cust-C2 و Cust-C3 از طریق Backdoor Path برقرار کنیم, این دو روتر برای ارتباط با هم از این لینک مستقیم استفاده می کنند به جای آن که از طریق MPLS VPN این ارتباط برقرار شود, علت هم واضح است, روت هایی که از طریق Backdoor Path دریافت می شوند چون هر دو روتر در یک area هستند, به عنوان Intra area route در نظر گرفته می شوند, مراحل انتخاب بهترین مسیر در OSPF را به یاد بیاورید, قبل از این که به Cost یا هر چیز دیگری OSPF توجه کند, ابتدا Route Type را مد نظر قرار می دهد و روتهای Intra Area(O) الویت بالاتری نسبت به روت های Inter Area دارند. هدف از این Backdoor Path تنها داشتن یک مسیر Backup بوده که اگر ارتباط MPLS VPN بنا به هر دلیلی قطع شد, این دو سایت بتوانند با هم ارتباط داشته باشند در صورتی که با این شرایط, Backdoor Path به عنوان مسیر اصلی برای ارتباط این دو سایت در نظر گرفته شده است:

```
Cust-C2(config-if)#do sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/2] via 40.0.90.10, 00:00:52, FastEthernet0/1
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 03:01:42, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 6 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 03:01:42, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O        40.0.40.0/24 [110/2] via 40.0.90.10, 00:00:52, FastEthernet0/1
C        40.0.90.0/24 is directly connected, FastEthernet0/1
L        40.0.90.9/32 is directly connected, FastEthernet0/1
```

برای آن که بتوانیم این قانون تصمیم گیری OSPF برای انتخاب بهترین مسیر را عوض کنیم, بین Egress PE و Ingress PE(که در سناریوی ما برای این دوسایت, PE2 و PE3 مد نظر می باشد) یک Virtual Link به نام Sham

Link زده می شود. توجه داشته باشید فقط در شرایطی که ما دوسایت VPN داشته باشیم که از طریق یک Backdoor Link نیز با هم ارتباط داشته باشند, نیاز به استفاده از Sham Link خواهیم داشت, در غیر این صورت به این فیچر نیازی نیست.

برای ساخت Sham Link, روی روترهای PE یک اینترفیس Loopback تعریف کرده , VRF Cust مورد نظر را به این اینترفیس loopback جدید اختصاص می دهیم و برای آن یک آدرس 32/ تعریف می کنیم. آدرس این Loopback از طریق OSPF, advertise نخواهد شد؛ این آدرس فقط از طریق BGP آن هم در محیط VRF مربوطه, advertise خواهد شد.

```
PE2(config)#int lo 3
PE2(config-if)#ip vrf forwarding Cust-C2
PE2(config-if)#ip add 3.3.3.3 255.255.255.255
PE2(config-if)#
PE2(config-if)#router bgp 200
PE2(config-router)#address-family ipv4 vrf Cust-C2
PE2(config-router-af)#network 3.3.3.3 mask 255.255.255.255
```

```
PE3(config)#int lo 4
PE3(config-if)#ip vrf forwarding Cust-C3
PE3(config-if)#ip add 4.4.4.4 255.255.255.255
PE3(config-if)#
PE3(config-if)#router bgp 200
PE3(config-router)#address-family ipv4 vrf Cust-C3
PE3(config-router-af)#network 4.4.4.4 mask 255.255.255.255
```

برای تست درستی تنظیمات نگاهی به VRF RIB Cust-C2 در روتر PE2 می اندازیم:

```
PE2#sho ip route vrf Cust-C2

Routing Table: Cust-C2
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      3.0.0.0/32 is subnetted, 1 subnets
C        3.3.3.3 is directly connected, Loopback3
      4.0.0.0/32 is subnetted, 1 subnets
B        4.4.4.4 [200/0] via 4.4.4.4, 00:02:36
      9.0.0.0/32 is subnetted, 1 subnets
O        9.9.9.9 [110/2] via 40.0.39.9, 04:15:44, FastEthernet1/0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/3] via 40.0.39.9, 00:44:39, FastEthernet1/0
      11.0.0.0/32 is subnetted, 1 subnets
B        11.11.11.11 [200/2] via 1.1.1.1, 03:45:39
      40.0.0.0/8 is variably subnetted, 5 subnets, 2 masks
B        40.0.11.0/24 [200/0] via 1.1.1.1, 03:45:39
C        40.0.39.0/24 is directly connected, FastEthernet1/0
L        40.0.39.3/32 is directly connected, FastEthernet1/0
O        40.0.40.0/24 [110/3] via 40.0.39.9, 00:44:39, FastEthernet1/0
O        40.0.90.0/24 [110/2] via 40.0.39.9, 00:44:49, FastEthernet1/0
```

آدرس loopback تازه پیکربندی شده, از طریق BGP دریافت شده, اما اگر نگاهی به Routing Table روترهای CE نیز بیاندازیم می بینیم آن ها هم این Prefixها را دریافت کرده اند:

```
Cust-C2#sho ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      3.0.0.0/32 is subnetted, 1 subnets
O E2     3.3.3.3 [110/1] via 40.0.39.3, 00:14:30, FastEthernet0/0
      4.0.0.0/32 is subnetted, 1 subnets
O E2     4.4.4.4 [110/1] via 40.0.39.3, 00:13:49, FastEthernet0/0
      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/2] via 40.0.90.10, 00:56:02, FastEthernet0/1
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 03:56:52, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 6 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 03:56:52, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O        40.0.40.0/24 [110/2] via 40.0.90.10, 00:56:02, FastEthernet0/1
C        40.0.90.0/24 is directly connected, FastEthernet0/1
L        40.0.90.9/32 is directly connected, FastEthernet0/1
```

روترهای CE نباید از این Prefix اگاه شوند, پس با استفاده از Route-map از advertise, این Prefix ها به روترهای CE جلوگیری می کنیم:

```
PE3(config)#ip prefix Sham-link seq 5 permit 3.3.3.3/32
PE3(config)#ip prefix Sham-link seq 10 permit 4.4.4.4/32
PE3(config)#
PE3(config)#route-map bgp-vpnv4 deny 10
PE3(config-route-map)#match ip add prefix-list Sham-link
PE3(config-route-map)#route-map bgp-vpnv4 permit 20
PE3(config-route-map)#
PE3(config-route-map)#router ospf 10
PE3(config-router)#redistribut bgp 200 route-map bgp-vpnv4
```

دقیقا دستورات بالا در PE2 نیز پیاده سازی می شود, حال دوباره نگاهی به جدول روتینگ Cust-C2 می اندازیم:

```
Cust-C2#sho ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/2] via 40.0.90.10, 01:14:57, FastEthernet0/1
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 04:15:47, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 6 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 04:15:47, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O        40.0.40.0/24 [110/2] via 40.0.90.10, 01:14:57, FastEthernet0/1
C        40.0.90.0/24 is directly connected, FastEthernet0/1
L        40.0.90.9/32 is directly connected, FastEthernet0/1
```

خوب فیلترینگ موفقیت آمیز بود, حال نوبت به تعریف Sham Link میرسد, دستور پیاده سازی Sham-Link مشابه Virtual Link می باشد: **area [area-num] sham-link source-address destination-address**

```
PE2(config)#router ospf 10
PE2(config-router)#area 0 sham-link 3.3.3.3 4.4.4.4



PE3(config)#router ospf 10
PE3(config-router)#area 0 sham-link 4.4.4.4 3.3.3.3
```

خوب حال برای این که مطمئن شویم Sham-Link درست پیکربندی شده و روت های سایت C2 و C3 به جای Backdoor Link از طریق MPLS VPN فرا گرفته می شوند, نگاهی به جدول روتینگ CE2 می اندازیم:

```
Cust-C2#sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/2] via 40.0.90.10, 00:09:45, FastEthernet0/1
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 00:14:58, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 6 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 00:14:58, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O        40.0.40.0/24 [110/2] via 40.0.90.10, 00:09:45, FastEthernet0/1
C        40.0.90.0/24 is directly connected, FastEthernet0/1
L        40.0.90.9/32 is directly connected, FastEthernet0/1
```

همانطور که مشاهده می کنید با وجود پیکربندی Sham-Link روتر Cust-C2 هم چنان از مسیر Backdoor برای رسیدن به Prefix های روتر Cust-C3 استفاده می کند, اما دلیل چیست؟ OSPF بعد از بررسی Route Type در صورتی که Route Type ها یکسان باشند به سراغ Cost خواهد رفت, بیایید بررسی کنیم که Cost برای Prefix 10.10.10.10/32 از رطیق Backdoor Link چه مقداری است:

```
Cust-C2#sho ip route 10.10.10.10
Routing entry for 10.10.10.10/32
  Known via "ospf 10", distance 110, metric 2, type intra area
  Last update from 40.0.90.10 on FastEthernet0/1, 00:18:32 ago
  Routing Descriptor Blocks:
  * 40.0.90.10, from 10.10.10.10, 00:18:32 ago, via FastEthernet0/1
      Route metric is 2, traffic share count is 1
```

خوب متریک 2 می باشد(1 گام می شود Backdoor Link, یک گام هم خود Loopback, در نتیجه: 2=1+1) حال اگر بخواهیم از طریق Sham-Link به این Prefix برسیم, Cost چقد خواهد شد؟ یک گام تا PE3, یک گام Sham-Link, یک گام از PE3 تا Cust-C3 و یک گام هم Loopback, در نتیجه Cost=4, پس انتظار می رود اگر cost مسیر Backdoor بیش از 4 شود, Sham-Link به عنوان مسیر اصلی در نظر گرفته شود, بیایید Cost لینک Backdoor را در روترهای Cust-C2 و Cust-C3 افزایش دهیم:

```
Cust-C2(config)#int fa 0/1
Cust-C2(config-if)#ip ospf cost 10
```

```
Cust-C3(config)#int fa 0/1
Cust-C3(config-if)#ip ospf cost 10
```

حال نگاهی به جدل روتینگ Cust-C2 می اندازیم تا ببینیم آیا حدسمان در رابطه با افزایش cost درست بوده است یا نه؟☺

```
Cust-C2#sho ip rou
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 1 subnets
C        9.9.9.9 is directly connected, Loopback0
      10.0.0.0/32 is subnetted, 1 subnets
O        10.10.10.10 [110/4] via 40.0.39.3, 00:05:31, FastEthernet0/0
      11.0.0.0/32 is subnetted, 1 subnets
O IA     11.11.11.11 [110/3] via 40.0.39.3, 00:38:27, FastEthernet0/0
      40.0.0.0/8 is variably subnetted, 6 subnets, 2 masks
O IA     40.0.11.0/24 [110/2] via 40.0.39.3, 00:38:27, FastEthernet0/0
C        40.0.39.0/24 is directly connected, FastEthernet0/0
L        40.0.39.9/32 is directly connected, FastEthernet0/0
O        40.0.40.0/24 [110/3] via 40.0.39.3, 00:05:31, FastEthernet0/0
C        40.0.90.0/24 is directly connected, FastEthernet0/1
L        40.0.90.9/32 is directly connected, FastEthernet0/1
```

```
Cust-C2#sho ip route 10.10.10.10
Routing entry for 10.10.10.10/32
  Known via "ospf 10", distance 110, metric 4, type intra area
  Last update from 40.0.39.3 on FastEthernet0/0, 00:07:06 ago
  Routing Descriptor Blocks:
  * 40.0.39.3, from 10.10.10.10, 00:07:06 ago, via FastEthernet0/0
      Route metric is 4, traffic share count is 1
```

خوب یک نتیجه عالی! پیکربندی Sham-Link و حدسمان در رابطه با cost کاملا درست بود! به این ترتیب تنها در صورتی از مسیر Backdoor استفاده خواهد شد که مسیر MPLS VPN با مشکل روبه رو شود.

تا اینجا تمام مباحث مهم مربوط به پیکربندی, بررسی و Trobleshooting, MPLS L3 VPN تقریبا بیان شد, اما برای تمرین و ورود به این وادی زیبا, از شما دعوت می کنم تا سناریوی بالا را در حالتی که ارتباط بین روترهای PE-CE, بر اساس BGP می باشد, پیاده سازی نمایید ;)

خلاصه بخش سوم:MPLS L3 VPN

مفاهیم:

PE(Provider Edge): روتری که در لبه کلود MPLS قرار دارد و از یک سمت با روترهای core و از طرف دیگر با CE ها در ارتباط است

P(Provider): روتری که هیچ لینک ارتباطی با هیچ CE ندارد

CE(Customer Edge): روتری که هیچ آگاهی از MPLS ندارد و به یک PE متصل می باشد.

Global Routing Table: Routing Table اصلی روتر

VRF(Virtual Routing & Forwarding): یک کپی از Routing Table داخل IOS که سبب تمایز Prefix های Customer ها می شود.

RD(Route Distinguisher): یک VPNv4 Attribute که به اول NLRI ها اضافه شده و سبب تمایز NLRI ها می شود.

RT(Route Target): یک BGP Extended Community که مشخص می کند هر روت به کدام VRF تعلق دارد. دارای دو پالیسی است: import و export

RT import Policy: چه روت هایی از BGP داخل VRF ریخته شود.

RT export Policy: چه روت هایی از VRF به داخل BGP ریخته شود.

VPN Label(Local Label or Inner Label): Label ای که از طریق BGP به دست می آید.

Transit Label(Outer Label): Label ای که از طریق LDP به دست می آید.

VPNv4 Route: NLRI ای که به اول آن RD اضافه شده است.

MPLS VPN Control Plane: بر اساس سه فاکتور عمل می کند:

1- VRF
2- RD
3- RT

} از طریق VPNv4 Update به دست می آیند.

MPLS VPN Data Plane: بر اساس دو فاکتور عمل می کند:

1- Transport Lable: از طریق IGP+LDP به دست می آید.
2- VPN Label: از طریق BGP Update به دست می آید.

پیاده سازی:

Configure MPLS Core(P & PE): LDP+IGP

Configure VRF in PE Router:

ip vrf [name]
rd x:xxx
route-target both x:xxx

int [type][number]
ip vrf forwarding [vrf-name]

Configure PE-CE Connection: IGP or Static or BGP:

CE: Normal Configuration

PE:

- ✓ Use: **address-family ipv4 vrf [name]** per routing protocol configuration
- ✓ Assign VRF to Interface: **ip vrf forwarding [vrf-name]**

Configure MP-BGP VPNv4:

Router bgp [ASN]
no bgp default ipv4-unicast
Neighbor x.x.x.x remote-as [ASN]
Address-family vpnv4
neighbor 3.3.3.3 activate

Configuring MP-BGP VPNv4 Prefixes exchange:

PE(both bgp and PE-CE Routing Protocol):
router [routing-protocol]
address-family ipv4 vrf [vrf-name]
redistribute [routing-protocol]

دستورات بررسی:

Verify the VRF was properly allocated	- Show [ip] vrf - Show [ip] vrf detail ---> View RD & RT - Show run vrf
Verify interfaces were properly allocated to the VRF	- Show ip route vrf - Show ipv6 route vrf
PE to CE Routing Verification	- CE: show ip route - PE: show ip route vrf [name]
Verify VPNv4 BGP establish	- show bgp vpnv4 unicast all summary
Verify extended communities are being sent	- debug bgp vpnv4 unicast updates
Verify RT import/export policy is correct	- show [ip] vrf detail - show run vrf
Verify IGP to BGP redistribution occur	- show bgp vpnv4 unicast all - show ip ospf database - show ip eigrp database - show ip rip database
Verify VPNv4 routes are being sent/received	- show bgp vpnv4 unicast all - show bgp vpnv4 unicast all neighbor advertised-- routes - debug bgp vpnv4 unicast updates - clear p ip bgp vpnv4 unicast * [in|out]