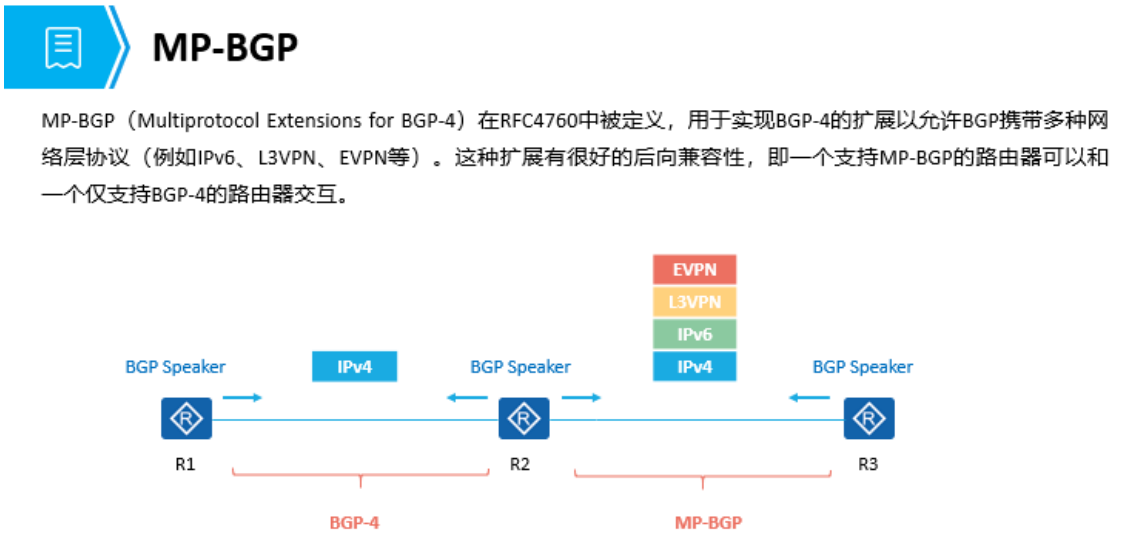
BGP EVPN 基础

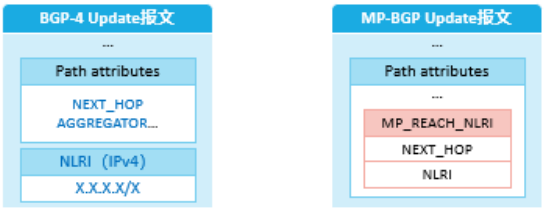
EVPN（Ethernet Virtual Private Network）是一种用于二层网络互联的 VPN 技术。EVPN 技术采用类似于 BGP/MPLS IP VPN 的机制，通过扩展 BGP 协议，使用扩展后的可达性信息，使不同站点的二层网络间的 MAC 地址学习和发布过程从数据平面转移到控制平面。

- 标准 BGP-4 仅支持 IPv4 单播地址，为了支持更多的网络层协议，MP-BGP（Multiprotocol Extensions for BGP-4）（RFC4760）被提出作为 BGP-4 的扩展允许不同类型的地址族在 BGP 中同时分发，例如 IPv4 组播、IPv6、L3VPN、EVPN 等。
- 随着 SDN 的发展和商用，EVPN（Ethernet VPN）在各解决方案中占据重要角色，应用覆盖全场景包括园区网络、数据中心网络、广域 IP 承载网络和 SD-WAN。
- 本章节将简单介绍 MP-BGP 概念、EVPN 的发展历史、常见的 EVPN 路由类型和应用场景。



BGP-4扩展

- BGP-4中IPv4特有的三个信息是NEXT_HOP属性、AGGREGATOR和IPv4 NLRI。因此为了支持多种网络层协议，BGP-4需要增加两种能力：
  - 关联其他网络层协议下一跳信息的能力。
  - 关联其他网络层协议NLRI的能力。
- 这种两种能力被互联网数字分配机构（IANA）统称为地址族（Address Family，AF）。
- 为了实现后向兼容性，协议规定MP-BGP增加两种新的属性，MP_REACH_NLRI和MP_UNREACH_NLRI，分别用于表示可达的目的信息和不可达的目的信息。这两种属性都属于可选非过渡（optional and non-transitive）。



- BGP-4 规定 IPv4 的 NEXT_HOP 和 AGGREGATOR 属于 Path attributes 字段，IPv4 的 NLRI 中携带 IPv4 的路由条目。
- MP-BGP 新增 Path attributes 的字段，将对应的网络层协议的 NEXT_HOP 字段和 NLRI 归属于 MP_REACH_NLRI。MP_REACH_NLRI 为 Path attributes 的新增字段。

MP_REACH_NLRI

- MP_REACH_NLRI被携带于BGP Update报文中，有以下作用：
  - 通告可达的路由给BGP邻居
  - 通告可达路的路由的下一跳给BGP邻居
- 其详细字段如下：

MP_REACH_NLRI格式	字段说明
Address Family Identifier (2 octets)	该字段标识了网络层协议，例如2表示IPv6
Subsequent Address Family Identifier (1 octet)	该字段和AFI一起使用，例如1表示unicast，结合AF为2表示IPv6单播
Length of Next Hop Network Address (1 octet)	该字段表示下一跳地址的长度
Network Address of Next Hop (variable)	该字段为此下一跳地址，格式由AF和SAFI决定
Reserved (1 octet)	全为0
Network Layer Reachability Information (variable)	该字段变长可包含可达的路由

- SAFI 字段中 1 表示单播，2 表示组播。值由 IANA 分配，其分配原则被定义于 RFC2434（Guidelines for Writing an IA

NA Considerations Section in RFCs）。

- 本章节后续学习 EVPN 的 AFI 为 25 (L2VPN) ，SAFI 为 70 (EVPN)。

MP_UNREACH_NLRI

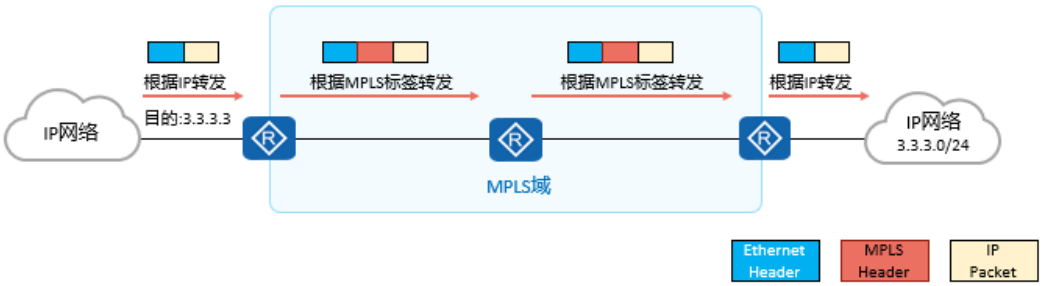
- MP_UNREACH_NLRI被携带于BGP Update报文中，用于撤销不可达的路由。
- 其详细字段如下：

MP_UNREACH_NLRI格式	字段说明
Address Family Identifier (2 octets)	该字段标识了网络层协议，例如2表示IPv6
Subsequent Address Family Identifier (1 octet)	该字段和AFI一起使用，例如1表示unicast，结合AF为2表示IPv6单播
Withdrawn Routes (variable)	该字段变长列举了需要被撤销的路由，其格式由AF和SAFI决定

- 例如 EVPN 的 AFI 为 25 (L2VPN) ，SAFI 为 70 (EVPN)。

MPLS简介

- MPLS （Multiprotocol Label Switching，多协议标记交换）位于TCP/IP协议栈中的数据链路层和网络层之间，在两层之间增加了额外的MPLS头部。报文转发直接基于MPLS头部。MPLS头部又被称为MPLS标签（Label）。
- MPLS以标签交换替代IP转发，实现了基于标签的快速转发。



- MPLS 起源于 IPv4（Internet Protocol version 4），其核心技术可扩展到多种网络协议，包括 IPv6（Internet Protocol version 6）、IPX（Internet Packet Exchange）、Appletalk、DECnet、CLNP（Connectionless Network Protocol）等。MPLS 中的“Multiprotocol”指的就是支持多种网络协议。
- MPLS 以标签交换替代 IP 转发。标签是一个短而定长的、

只具有本地意义的连接标识符，与 ATM 的 VPI/VCI 以及 Frame Relay 的 DLCI 类似。

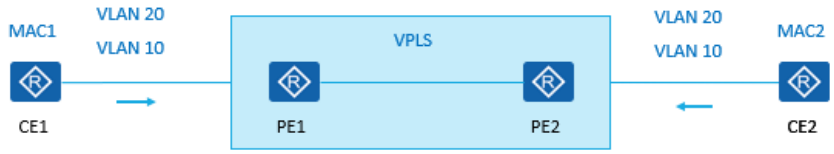
- MPLS 域（MPLS Domain）：一系列连续的运行 MPLS 的网络设备构成了一个 MPLS 域。

VPLS简介

VPLS（Virtual Private LAN Service）是一种基于以太网的二层VPN技术，它在MPLS网络上提供了类似LAN的业务，允许用户可以从多个地址位置接入网络、相互访问。

传统L2VPN

- 传统的L2VPN业务例如VPLS（Virtual Private LAN Service），提供用户远程站点之间二层连接服务。它组建二层交换网，像二层交换机一样透传以太报文。本例中PE1和PE2组建的VPLS网络透传CE1和CE2之间的VLAN流量。
- 因此在传统L2VPN中对于远端MAC地址的学习依靠ARP广播泛洪，PE设备将需要承载广播流量。广播占用较多的接口带宽，这是传统L2VPN的一个典型问题。

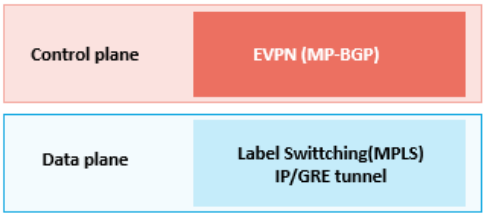


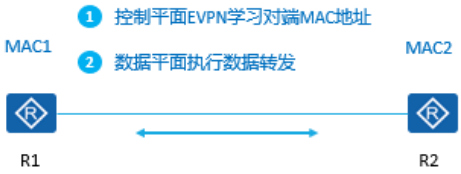
"二层交换机"

- VPLS 有更多的问题，例如不支持多活接入、故障收敛慢、不支持负载均衡等不在本课程介绍，请学习 HCIE-HCIE-Datacom《Etherent VPN》和 RFC 7209 - Requirements for Ethernet VPN (EVPN)。

EVPN的诞生

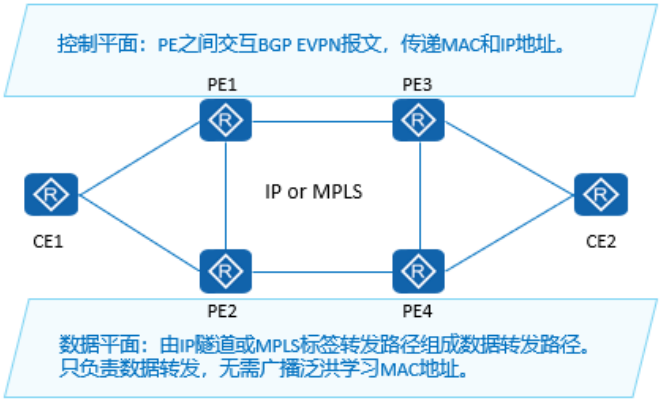
- 随着新技术和新场景对网络需求，VPLS被暴露出更多的问题无法满足二层VPN的需求。业界重新审视了对Ethernet VPN的需求（RFC 7209），提出新的解决方案EVPN（Ethernet VPN）。
- EVPN最初在RFC 7432中被定义，EVPN引入控制平面，用于更好的控制MAC地址学习过程。
- EVPN的控制平面采用MP-BGP，数据平面支持MPLS LSPs或者IP/GRE tunneling。





EVPN的优势

- EVPN颠覆了传统L2 VPN数据面学习的方式，引入控制面学习MAC和IP指导数据转发，实现了转控分离。
- EVPN解决传统L2 VPN的典型问题，带来双活，快速收敛，简化运维等更多的价值。



EVPN其他优势：

- 支持CE多活接入PE
- 支持PE成员自动发现
- 环路避免
- 广播流量优化
- 支持ECMP

- 更多详细 EVPN 原理，请参考 HCIE-Datacom《Ethernet VPN》。

EVPN NLRI

- EVPN定义了一种新的BGP NLRI（Network Layer Reachable Information）来承载所有的EVPN路由，被称为EVPN NLRI。
- EVPN NLRI是MP-BGP的新型扩展，被包含于MP_REACH_NLRI中，定义了新的NLRI。它规定了EVPN的AFI（Address Family Identifier）是25，SAFI（Subsequent Address Family Identifier）是70。

MP_REACH_NLRI格式

Address Family Identifier (2 octets)
Subsequent Address Family Identifier (1 octet)
Length of Next Hop Network Address (1 octet)
Network Address of Next Hop (variable)
Reserved (1 octet)
Network Layer Reachability Information (variable)

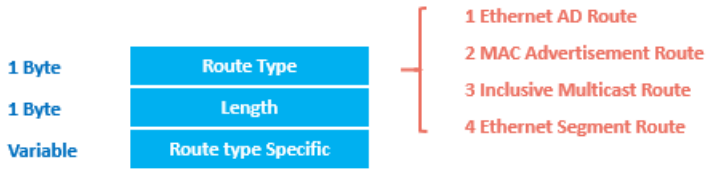
BGP EVPN的MP_REACH_NLRI

AFI: 25
SAFI: 70
该字段表示下一跳地址的长度。
该字段为EVPN路由的下一跳地址。
全为0
EVPN NLRI

EVPN路由

EVPN NLRI格式采用TLV（Type-Length-Value）三元组结构，使得报文具有很强的灵活性和扩展性：

- Route Type定义了不同的EVPN路由。RFC 7432中首先定义了四类路由。
- Length定义了字段的长度。
- Route Type Specifc则表示不同的路由类型有不同的字段填充。



EVPN NLRI格式　　Route Type字段定义的四种常见EVPN路由

- The NLRI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute contains the EVPN NLRI (encoded as specified above).
- The EVPN NLRI is carried in BGP [RFC4271] using BGP Multiprotocol Extensions [RFC4760] with an Address Family Identifier (AFI) of 25 (L2VPN) and a Subsequent Address Family Identifier (SAFI) of 70 (EVPN). The NLRI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute contains the EVPN NLRI (encoded as specified

above).

- In order for two BGP speakers to exchange labeled EVPN NLRI, they must use BGP Capabilities Advertisements to ensure that they both are capable of properly processing such NLRI. This is done as specified in [RFC4760], by using capability code 1 (multiprotocol BGP) with an AFI of 25 (L2VPN) and a SAFI of 70 (EVPN).

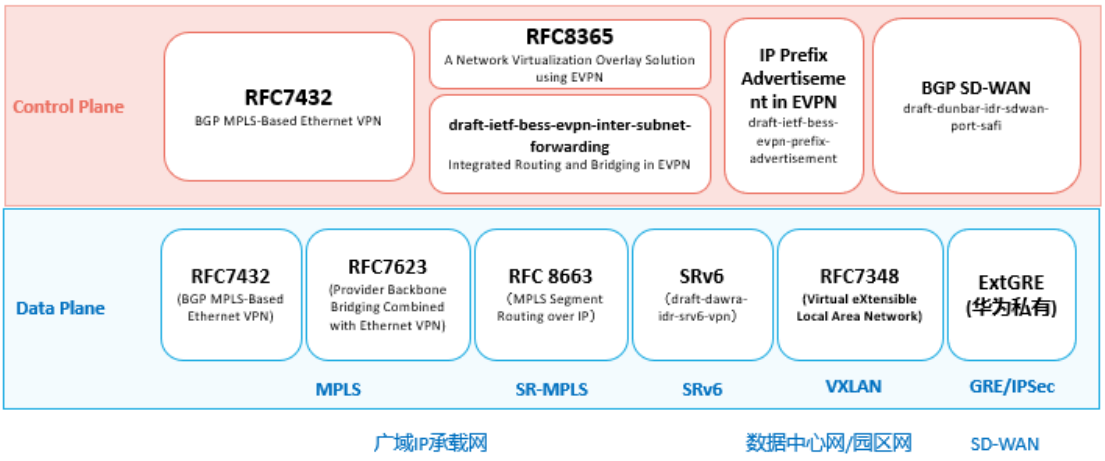
EVPN更多类型路由及作用

EVPN不仅限于二层VPN的应用，随着其EVPN路由类型的增加，支持更多的应用例如L3 VPN功能。

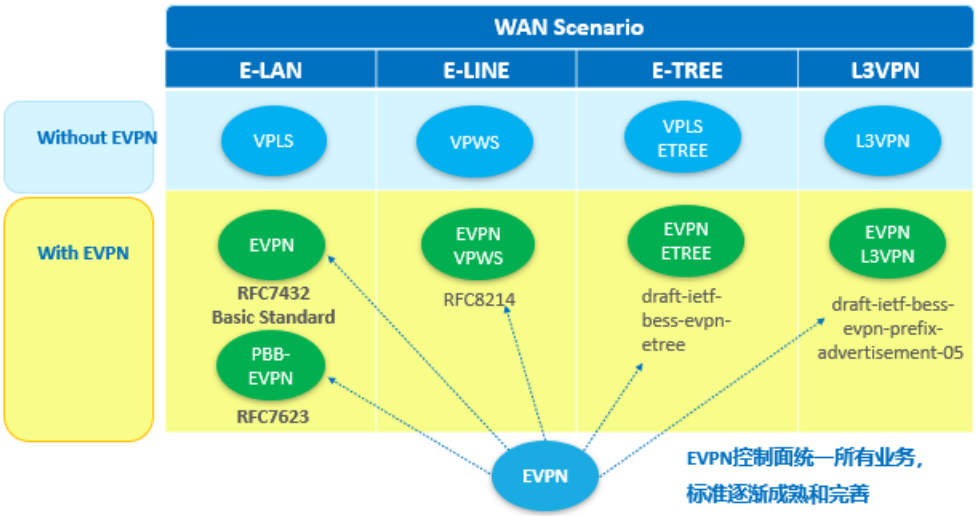
路由类型	作用	RFC
(Type 1) Ethernet A-D Route	• 别名 • MAC地址批量撤销 • 多活指示 • 通告ESI标签	RFC 7432
(Type 2) MAC/IP Advertisement Route	• MAC地址学习通告 • MAC/IP绑定 • MAC地址移动性	
(Type 3) Inclusive Multicast Route	组播隧道端点自动发现&组播类型自动发现	
(Type 4) Ethernet Segment Route	ES成员自动发现 DF选举	
(Type 5) IP Prefix Route	IP Prefix通告（支持L3 VPN）	draft-ietf-bess-evpn-prefix-advertisement

- Type5 类路由 IP Prefix Route 在标准化进程中目前处于草案阶段，draft-ietf-bess-evpn-prefix-advertisement。

EVPN协议标准



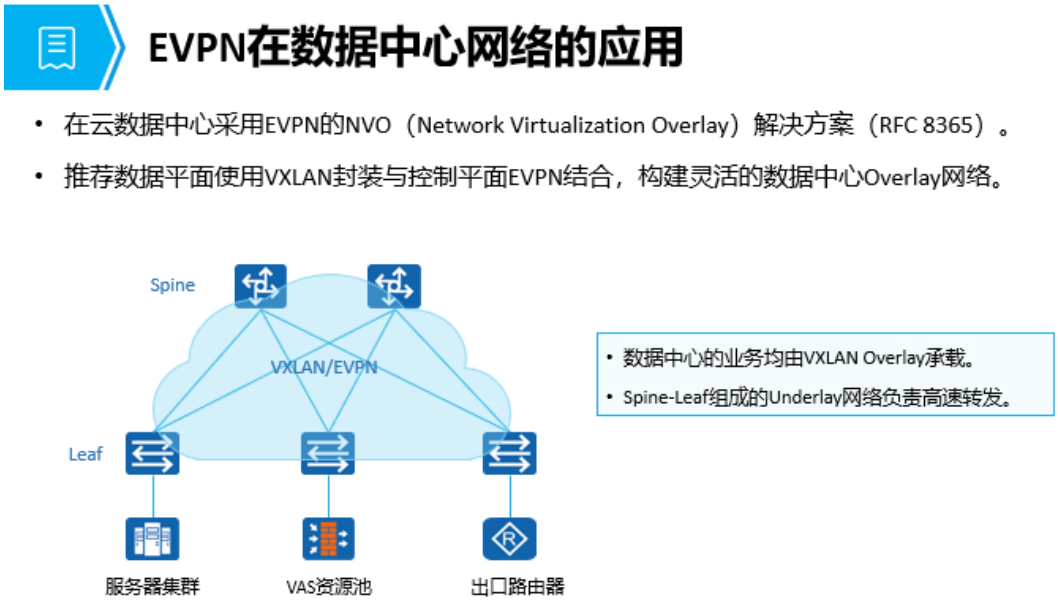
EVPN在广域IP承载网的应用



- E-LINE、E-TREE、E-LAN 是 EVC 定义的三种类型，具体请参考城域以太网标准。https://wiki.mef.net/display/CESG/E-Line
- MEF 中提到了三种 EVC 的服务种类，点到点 EVC（Point-to-Point EVC）、多点到多点 EVC（Mutlipoint-to-Multipoint EVC）和根到多点 EVC（Rooted-Multipoint EVC）。
- E-LINE：一条点到点的 EVC 将两个 UNI 严格地关联。
- E-LAN：一个多点到多点 EVC 可以将两个或者两个以上

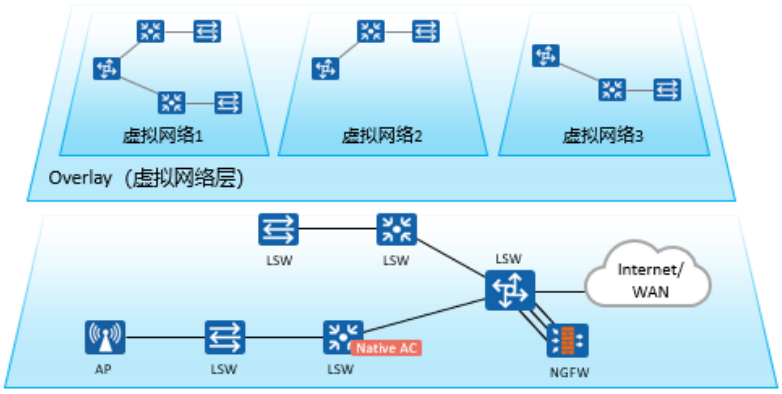
的 UNI 关联起来，而且用户/运营商可以在不影响其他 UNI 的前提下根据需要向这个 EVC 中添加任意个 UNI，或者将这个 EVC 中的某些 UNI 剔除。

- E-TREE：这种 EVC 类似于三层 VPN 中的 Hub-Spoke 模式，它包含一个或多个根 UNI（Root）和若干叶子 UNI（Leaf），其中根 UNI 可以和 EVC 中的所有 UNI 直接通信，而叶子 UNI 只能和 EVC 中的根 UNI 直接通信，两个叶子 UNI 之间不能直接通信。



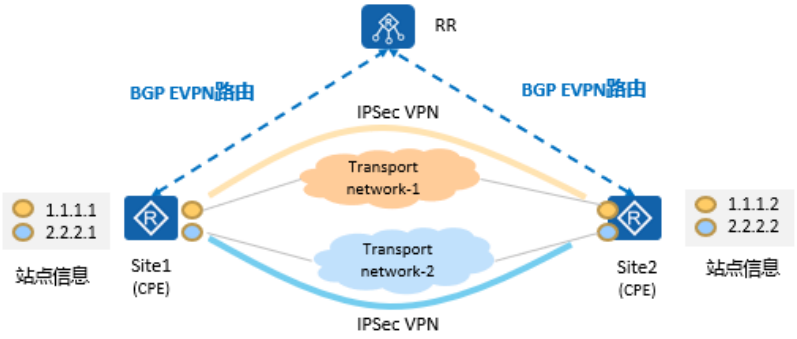
EVPN在园区网的应用

- 园区网虚拟化园区解决方案同在云数据中心相同，采用EVPN的NVO解决方案（RFC 8365）。
- 在不同的底层组网上使用VXLAN封装与控制平面EVPN结合，构建灵活的数据中心Overlay网络。



EVPN在SD-WAN的应用

- SD-WAN是新一代的企业分支互联解决方案，支持智能动态选路、ZTP和可视化等特性。
- SD-WAN解决方案中，在RR与CPE之间部署EVPN用于在控制平面传播SD-WAN的Overlay VPN路由，数据平面采用IPSec VPN构建安全的转发通道。



- Overlay VPN 路由包括站点 VPN 路由前缀、下一跳 TNP 路由信息以及用于 CPE 之间数据通道的数据加密所需要的 IP Sec 相关密钥等信息。详细内容请参考 SD-WAN 课程。
- CPE（Customer Premise Equipment，客户终端设备）。

思考题：

- （简答题）请简述 EVPN 的原理和常见的路由类型。
- （简答题）请简述 EVPN 的应用场景。

答案：

- EVPN 是 MP-BGP 的扩展，常见有五种路由类型，被用于作为 L2 或者 L3 隧道的控制平面。
- EVPN 可以被广泛用于企业全场景，例如 SD-WAN、园区网、数据中心和广域网。在数据中心和园区中，EVPN 与 VXLAN 结合构建业务 Overlay。在 SD-WAN 场景中 EVPN 与 IPSec 结合构建企业分支互联网络。在广域网中 EVPN 可以与各种底层隧道/标签技术结合，例如 MPLS/SR/VPLS/VPWS 等。