学习推荐

- 华为培训与认证官方网站
  - http://learning.huawei.com/cn/
- 华为在线学习
  - https://ilearningx.huawei.com/portal/#/portal/ebg/26
- 华为职业认证
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=zh
- 查找培训入口
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=zh

更多信息

- 华为培训APP

华为认证系列教程

HCIA-Routing & Switching 进阶

华为网络技术与设备

HUAWEI

华为技术有限公司

版权声明

华为认证系列教程

HCIA-Routing & Switching 华为网络技术与设备

第 2.5 版本

前言

简介

本书为 HCIA 认证培训教程，专门适用于准备参加 HCIA 考试的学员。对于希望通过在华为 VRP 平台上进行实际操作和演练，从而加强认识和理解数据通信原理的在校学生和专业人员，本身也极具参考价值。

内容描述

本书共包含六个 Module，全面介绍了增强企业网络各种重要特性所涉及的相关技术，以及这些技术是如何在 VRP 上配置和实现的。

Module 1 介绍了提升企业网络效率与可靠性的相关技术，内容包含链路聚合、VLAN、VLAN 间路由。

Module 2 介绍了企业网络远程互连时常用的广域网技术，包括 HDLC、PPP、PPPoE。

Module 3 介绍了基于华为路由交换产品的企业网安全技术，包含了 ACL、AAA、IPSec VPN 和 GRE。

Module 4 介绍了企业网管系统的基本原理以及配置与实现，包含 SNMP 技术。

Module 5 介绍了 IPv6 基础、IPv6 路由协议 RIPng 和 OSPFv3、以及 DHCPv6。

Module 6 介绍了现网中常见的 MPLS 和 SR 技术。

目录

链路聚合

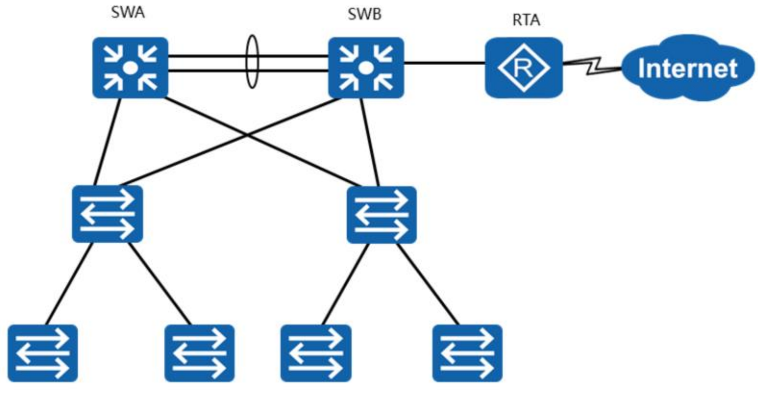
前言

- 随着网络规模不断扩大，用户对骨干链路的带宽和可靠性提出了越来越高的要求。在传统技术中，常用更换高速率的接口板或更换支持高速率接口板的设备的方式来增加带宽，但这种方案需要付出高额的费用，而且不够灵活。
- 采用链路聚合技术可以在不进行硬件升级的条件下，通过将多个物理接口捆绑为一个逻辑接口，来达到增加链路带宽的目的。在实现增大带宽目的的同时，链路聚合采用备份链路的机制，可以有效的提高设备之间链路的可靠性。

目标

- 学完本课程后，您将能够：
  - 掌握链路聚合的原理
  - 掌握链路聚合的配置

链路聚合的应用场景



- 链路聚合一般部署在核心结点，以便提升整个网络的数据吞吐量。

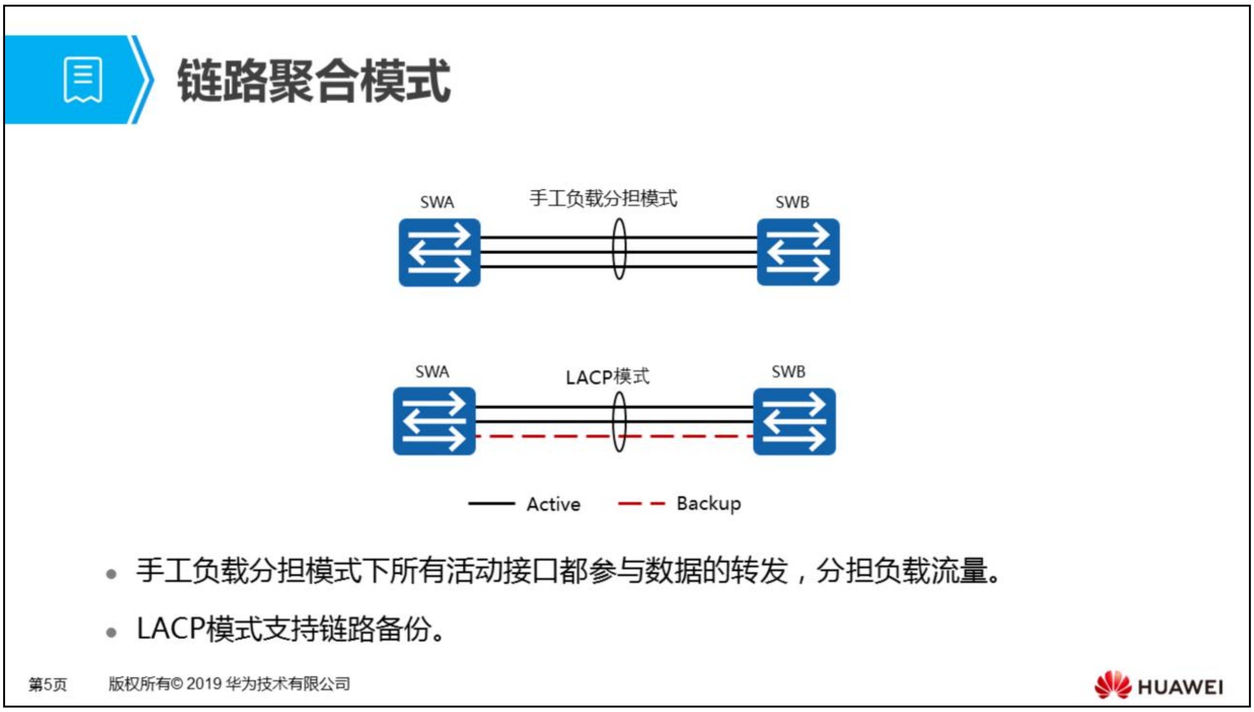
HUAWEI

- 在企业网络中，所有设备的流量在转发到其他网络前都会汇聚到核心层，再由核心区设备转发到其他网络，或者转发到外网。因此，在核心层设备负责数据的高速交换时，容易发生拥塞。在核心层部署链路聚合，可以提升整个网络的数据吞吐量，解决拥塞问题。本示例中，两台核心交换机SWA和SWB之间通过两条成员链路互相连接，通过部署链路聚合，可以确保SWA和SWB之间的链路不会产生拥塞。

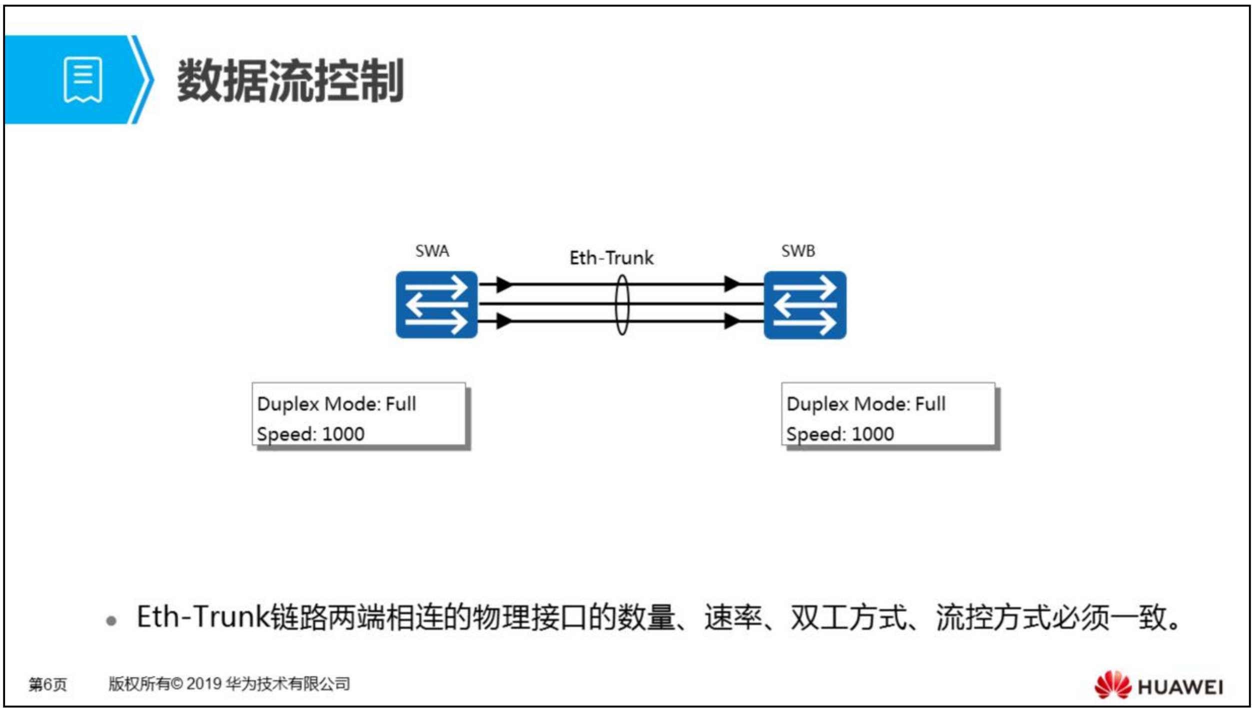
链路聚合

- 链路聚合能够提高链路带宽，增强网络可用性，支持负载分担。

- 链路聚合是把两台设备之间的多条物理链路聚合在一起，当做一条逻辑链路来使用。这两台设备可以是一对路由器，一对交换机，或者是一台路由器和一台交换机。一条聚合链路可以包含多条成员链路，在ARG3系列路由器和X7系列交换机上默认最多为8条。
- 链路聚合能够提高链路带宽。理论上，通过聚合几条链路，一个聚合口的带宽可以扩展为所有成员口带宽的总和，这样就有效地增加了逻辑链路的带宽。
- 链路聚合为网络提供了高可靠性。配置了链路聚合之后，如果一个成员接口发生故障，该成员口的物理链路会把流量切换到另一条成员链路上。
- 链路聚合还可以在一个聚合口上实现负载均衡，一个聚合口可以把流量分散到多个不同的成员口上，通过成员链路把流量发送到同一个目的地，将网络产生拥塞的可能性降到最低。



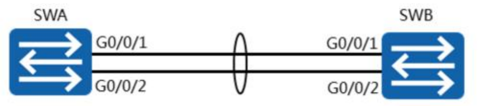
- 链路聚合包含两种模式：手动负载均衡模式和LACP（Link Aggregation Control Protocol）模式。
- 手工负载分担模式下，Eth-Trunk的建立、成员接口的加入由手工配置，没有链路聚合控制协议的参与。该模式下所有活动链路都参与数据的转发，平均分担流量，因此称为负载分担模式。如果某条活动链路故障，链路聚合组自动在剩余的活动链路中平均分担流量。当需要在两个直连设备间提供一个较大的链路带宽而设备又不支持LACP协议时，可以使用手工负载分担模式。ARG3系列路由器和X7系列交换机可以基于目的MAC地址，源MAC地址，或者基于源MAC地址和目的MAC地址，源IP地址，目的IP地址，或者基于源IP地址和目的IP地址进行负载均衡。
- 在LACP模式中，链路两端的设备相互发送LACP报文，协商聚合参数。协商完成后，两台设备确定活动接口和非活动接口。在LACP模式中，需要手动创建一个Eth-Trunk口，并添加成员口。LACP协商选举活动接口和非活动接口。LACP模式也叫M:N模式。M代表活动成员链路，用于在负载均衡模式中转发数据。N代表非活动链路，用于冗余备份。如果一条活动链路发生故障，该链路传输的数据被切换到一条优先级最高的备份链路上，这条备份链路转变为活动状态。
- 两种链路聚合模式的主要区别是：在LACP模式中，一些链路充当备份链路。在手动负载均衡模式中，所有的成员口都处于转发状态。



- 在一个聚合口中，聚合链路两端的物理口（即成员口）的所有参数必须一致，包括物理口的数量，传输速率，双工模式和流量控制模式。成员口可以是二层接口或三层接口。
- 数据流在聚合链路上传输，数据顺序必须保持不变。一个数据流可以看做是一组MAC地址和IP地址相同的帧。例如，两台设备间的SSH或SFTP连接可以看做一个数据流。如果未配置链路聚合，只是用一条物理链路来传输数据，那么一个数据流中的帧总是能按正确的顺序到达目的地。配置了链路聚合后，多条物理链路被绑定成一条聚合链路，一个数据流中的帧通过不同的物理链路传输。如果第一个帧通过一条物理链路传输，第二个帧通过另外一条物理链路传输，这样一来同一数据流的第二个数据帧就有可能比第一个数据帧先到达对端设备，从而产生接收数据包乱序的情况。
- 为了避免这种情况的发生，Eth-Trunk采用逐流负载分担的机制，这种机制把数据帧中的地址通过HASH算法生成HASH-KEY值，然后根据这个数值在Eth-Trunk转发表中寻找对应的出接口，不同的MAC或IP地址HASH得出的HASH-KEY值不同，从而出接口也就不同，这样既保证了同一数据流的帧在同一条物理链路转发，又实现了流量在聚合组内各物理链路上的负载分担，即逐流的负载分担。逐流负载分担能保证包的顺序，但不能保证带宽利用率。
- 负载分担的类型主要包括以下几种，用户可以根据具体应用选择不同的负载分担类型。
- 根据报文的源MAC地址进行负载分担；
- 根据报文的目的MAC地址进行负载分担；
- 根据报文的源IP地址进行负载分担；
- 根据报文的目的IP地址进行负载分担；

- 根据报文的源MAC地址和目的MAC地址进行负载分担；
- 根据报文的源IP地址和目的IP地址进行负载分担；
- 根据报文的VLAN、源物理端口等对L2、IPv4、IPv6和MPLS报文进行增强型负载分担。

二层链路聚合配置




```
[SWA]interface Eth-Trunk 1
[SWA-Eth-Trunk1]interface GigabitEthernet0/0/1
[SWA-GigabitEthernet0/0/1]eth-trunk 1
[SWA-GigabitEthernet0/0/1]interface GigabitEthernet0/0/2
[SWA-GigabitEthernet0/0/2]eth-trunk 1
```


- 本例中，通过执行interface Eth-trunk <trunk-id>命令配置链路聚合。这条命令创建了一个Eth-Trunk口，并且进入该Eth-Trunk口视图。trunk-id用来唯一标识一个Eth-Trunk口，该参数的取值可以是0到63之间的任何一个整数。如果指定的Eth-Trunk口已经存在，执行interface eth-trunk命令会直接进入该Eth-Trunk口视图。
- 配置Eth-Trunk口和成员口，需要注意以下规则：
- 只能删除不包含任何成员口的Eth-Trunk口。
- 把接口加入Eth-Trunk口时，二层Eth-Trunk口的成员口必须是二层接口，三层Eth-Trunk口的成员口必须是三层接口。
- 一个Eth-Trunk口最多可以加入8个成员口。
- 加入Eth-Trunk口的接口必须是hybrid接口（默认的接口类型）。
- 一个Eth-Trunk口不能充当其他Eth-Trunk口的成员口。
- 一个以太接口只能加入一个Eth-Trunk口。如果把一个以太接口加入另一个Eth-Trunk口，必须先把该以太接口从当前所属的Eth-Trunk口中删除。
- 一个Eth-Trunk口的成员口类型必须相同。例如，一个快速以太口（FE口）和一个千兆以太口（GE口）不能加入同一个Eth-Trunk。
- 位于不同接口板（LPU）上的以太口可以加入同一个Eth-Trunk口。如果一个对端接口直接和本端Eth-Trunk口的一个成员口相连，该对端接口也必须加入一个Eth-Trunk口。否则两端无法通信。
- 如果成员口的速率不同，速率较低的接口可能会拥塞，报文可能会被丢弃。

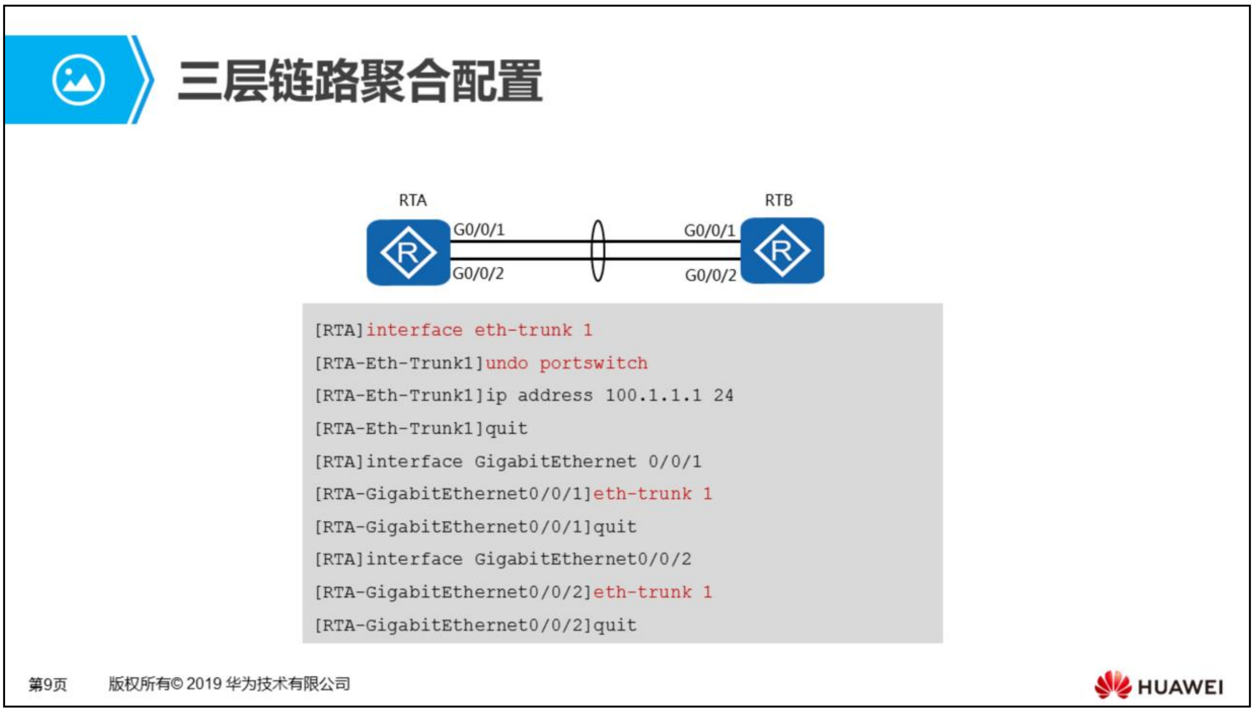
- 接口加入Eth-Trunk口后，Eth-Trunk口学习MAC地址，成员口不再学习。

查看链路聚合信息

```
[SWA]display interface eth-trunk 1
Eth-Trunk1 current state : UP
Line protocol current state : UP
……
----------------------------------------------------
PortName                      Status      Weight
----------------------------------------------------
GigabitEthernet0/0/1          UP          1
GigabitEthernet0/0/2          UP          1
----------------------------------------------------
The Number of Ports in Trunk : 2
The Number of UP Ports in Trunk : 2
```

- 两个成员接口已经被绑定到Eth-trunk 1。

- 执行display interface eth-trunk <trunk-id>命令，可以确认两台设备间是否已经成功实现链路聚合。也可以使用这条命令收集流量统计数据，定位接口故障。如果Eth-Trunk口处于UP状态，表明接口正常运行。如果接口处于Down状态，表明所有成员口物理层发生故障。如果管理员手动关闭端口，接口处于Administratively DOWN状态。可以通过接口状态的改变发现接口故障，所有接口正常情况下都应处于Up状态。

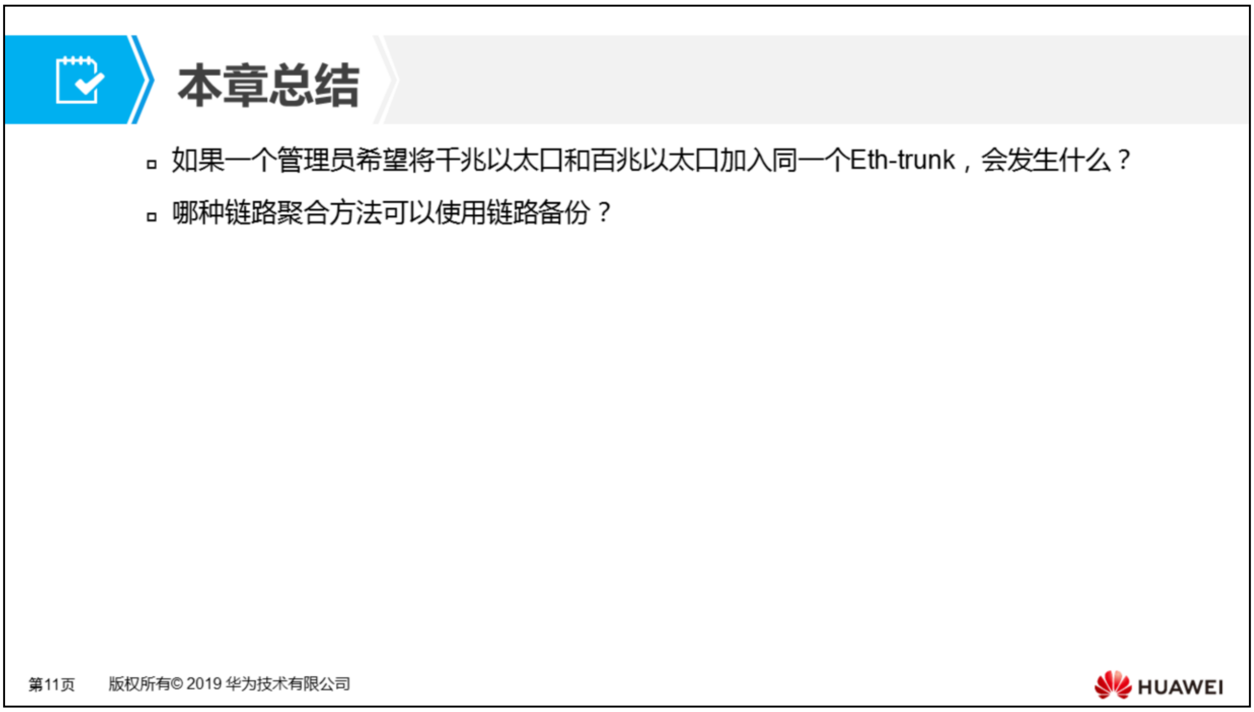


- 如果要在路由器上配置三层链路聚合，需要首先创建Eth-Trunk接口，然后在Eth-Trunk逻辑口上执行**undo portswitch**命令，把聚合链路从二层转为三层链路。执行**undo portswitch**命令后，可以为Eth-Trunk逻辑口分配一个IP地址。

查看链路聚合信息

```
[RTA]display interface eth-trunk 1
Eth-Trunk1 current state : UP
Line protocol current state : UP
……
------------------------------------------------------
PortName                  Status            Weight
------------------------------------------------------
GigabitEthernet0/0/1        UP                 1
GigabitEthernet0/0/2        UP                 1
------------------------------------------------------
The Number of Ports in Trunk : 2
The Number of UP Ports in Trunk : 2
```

- 两个成员接口已经被绑定到Eth-trunk 1。



- 一个快速以太口（FE口）和一个千兆以太口（GE口）不能加入同一个Eth-Trunk。如果将两个不同类型的接口加入到同一个Eth-Trunk口，设备会提示发生错误。
- 2. 只有LACP模式支持备份成员链路。如需建立备份链路，应使用LACP模式的链路聚合。


谢 谢
www.huawei.com

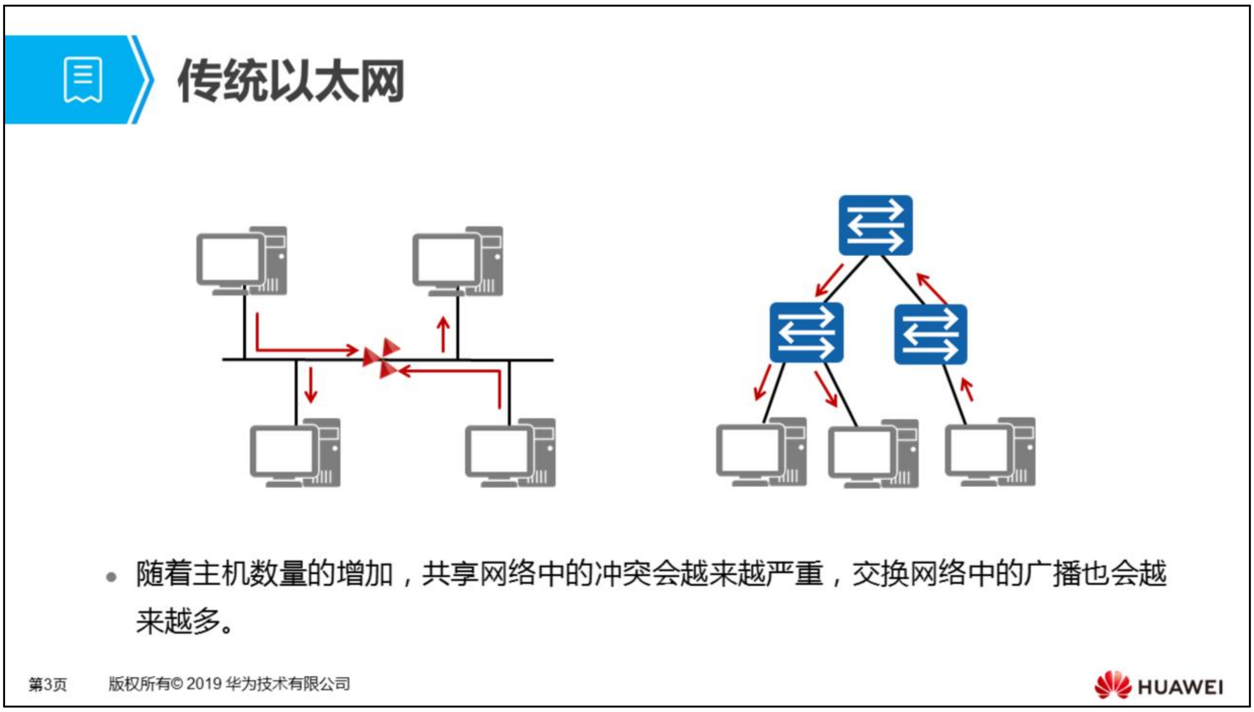
VLAN原理和配置

前言

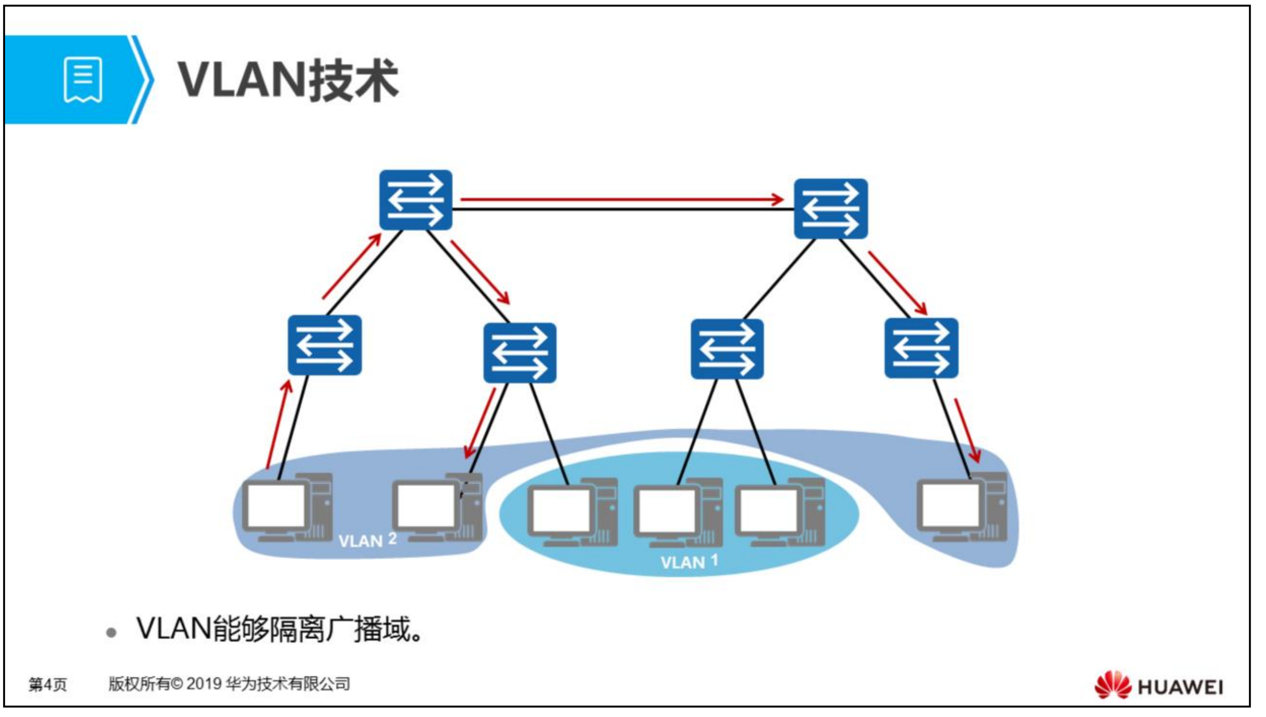
- 随着网络中计算机的数量越来越多，传统的以太网络开始面临冲突严重、广播泛滥以及安全性无法保障等各种问题。
- VLAN（Virtual Local Area Network）即虚拟局域网，是将一个物理的局域网在逻辑上划分成多个广播域的技术。通过在交换机上配置VLAN，可以实现在同一个VLAN内的用户可以进行二层互访，而不同VLAN间的用户被二层隔离。这样既能够隔离广播域，又能够提升网络的安全性。

目标

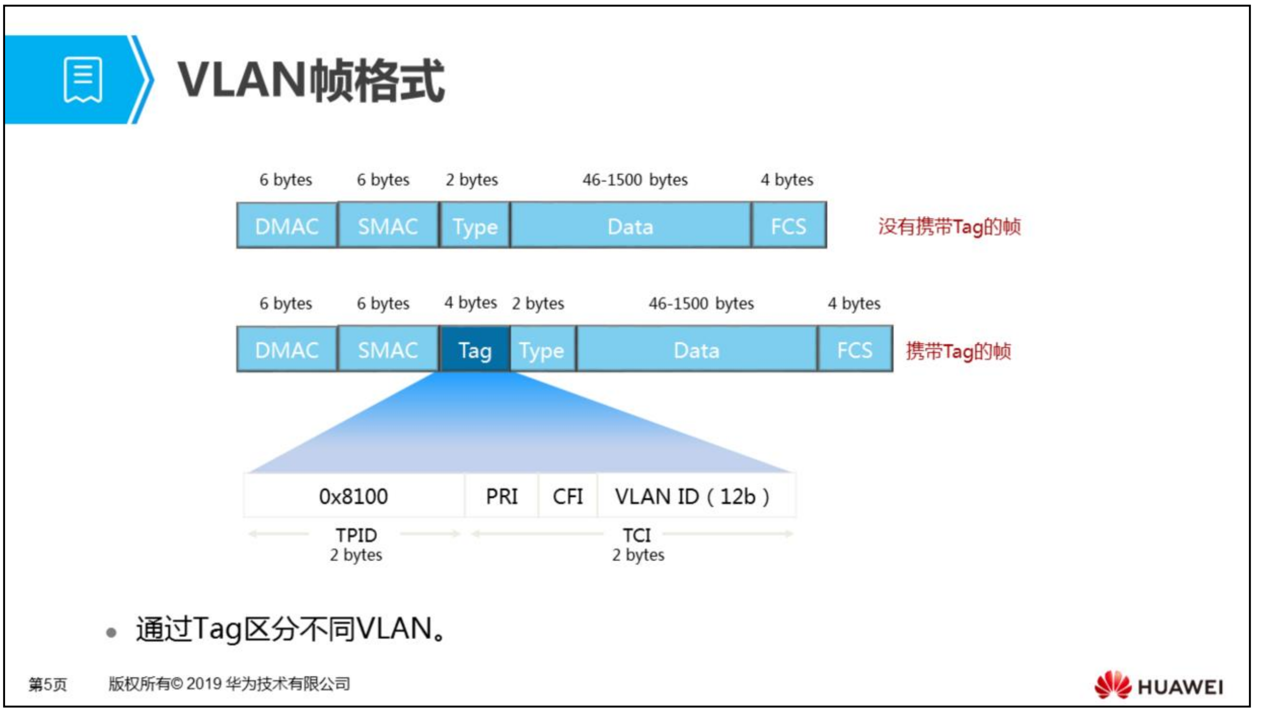
- 学完本课程后，您将能够：
  - 理解VLAN的工作原理
  - 掌握VLAN的基本配置



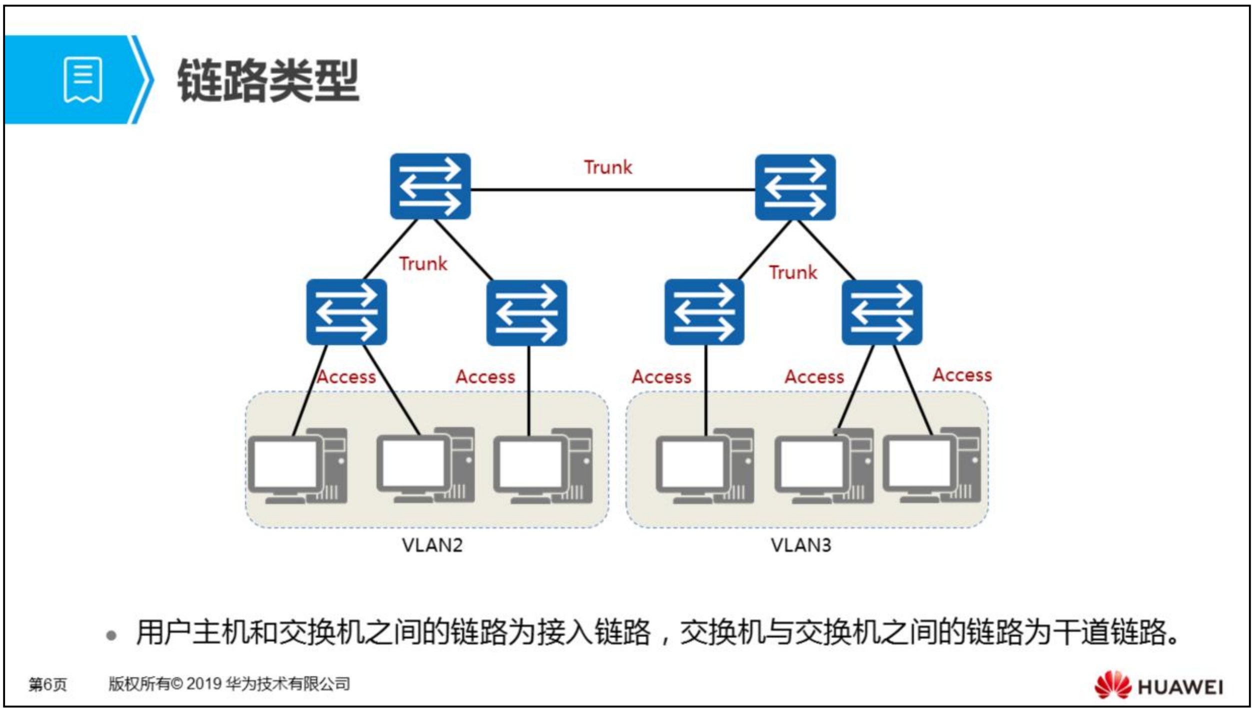
- 早期的局域网LAN技术是基于总线型结构的，它存在以下主要问题：
- 若某时刻有多个节点同时试图发送消息，那么它们将产生冲突。
- 从任意节点发出的消息都会被发送到其他节点，形成广播。
- 所有主机共享一条传输通道，无法控制网络中的信息安全。
- 这种网络构成了一个冲突域，网络中计算机数量越多，冲突越严重，网络效率越低。同时，该网络也是一个广播域，当网络中发送信息的计算机数量变多时，广播流量将会耗费大量带宽。
- 因此，传统局域网不仅面临冲突域太大和广播域太大两大难题，而且无法保障传输信息的安全。
- 为了扩展传统LAN，以接入更多计算机，同时避免冲突的恶化，出现了网桥和二层交换机，它们能有效隔离冲突域。网桥和交换机采用交换方式将来自入端口的信息转发到出端口上，克服了共享网络中的冲突问题。但是，采用交换机进行组网时，广播域和信息安全问题依旧存在。
- 为限制广播域的范围，减少广播流量，需要在没有二层互访需求的主机之间进行隔离。路由器是基于三层IP地址信息来选择路由和转发数据的，其连接两个网段时可以有效抑制广播报文的转发，但成本较高。因此，人们设想在物理局域网上构建多个逻辑局域网，即VLAN。



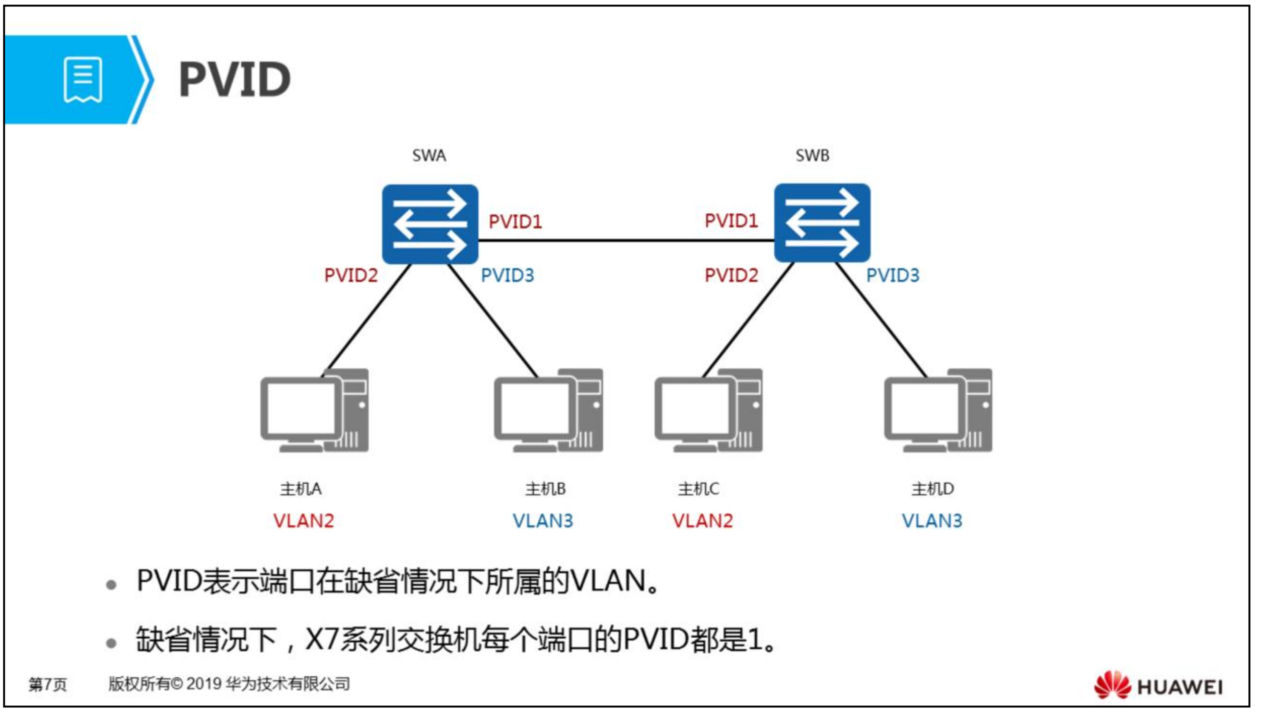
- VLAN技术可以将一个物理局域网在逻辑上划分成多个广播域，也就是多个VLAN。VLAN技术部署在数据链路层，用于隔离二层流量。同一个VLAN内的主机共享同一个广播域，它们之间可以直接进行二层通信。而VLAN间的主机属于不同的广播域，不能直接实现二层互通。这样，广播报文就被限制在各个相应的VLAN内，同时也提高了网络安全性。
- 本例中，原本属于同一广播域的主机被划分到了两个VLAN中，即，VLAN1和VLAN2。VLAN内部的主机可以直接在二层互相通信，VLAN1和VLAN2之间的主机无法直接实现二层通信。



- VLAN标签长4个字节，直接添加在以太网帧头中，IEEE802.1Q文档对VLAN标签作出了说明。
- TPID：Tag Protocol Identifier，2字节，固定取值，0x8100，是IEEE定义的新类型，表明这是一个携带802.1Q标签的帧。如果不支持802.1Q的设备收到这样的帧，会将其丢弃。
- TCI：Tag Control Information，2字节。帧的控制信息，详细说明如下：
    1. Priority：3比特，表示帧的优先级，取值范围为0～7，值越大优先级越高。当交换机阻塞时，优先发送优先级高的数据帧。
    2. CFI：Canonical Format Indicator，1比特。CFI表示MAC地址是否是经典格式。CFI为0说明是经典格式，CFI为1表示为非经典格式。用于区分以太网帧、FDDI（Fiber Distributed Digital Interface）帧和令牌环网帧。在以太网中，CFI的值为0。
    3. VLAN Identifier：VLAN ID，12比特，在X7系列交换机中，可配置的VLAN ID取值范围为0～4095，但是0和4095在协议中规定为保留的VLAN ID，不能给用户使用。
- 在现有的交换网络环境中，以太网的帧有两种格式：
- 没有加上VLAN标记的标准以太网帧（untagged frame）；有VLAN标记的以太网帧（tagged frame）。

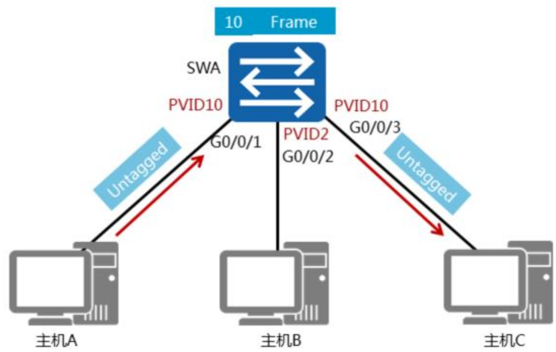


- VLAN链路分为两种类型：Access链路和Trunk链路。
- 接入链路（Access Link）：连接用户主机和交换机的链路称为接入链路。如本例所示，图中主机和交换机之间的链路都是接入链路。
- 干道链路（Trunk Link）：连接交换机和交换机的链路称为干道链路。如本例所示，图中交换机之间的链路都是干道链路。干道链路上通过的帧一般为带Tag的VLAN帧。



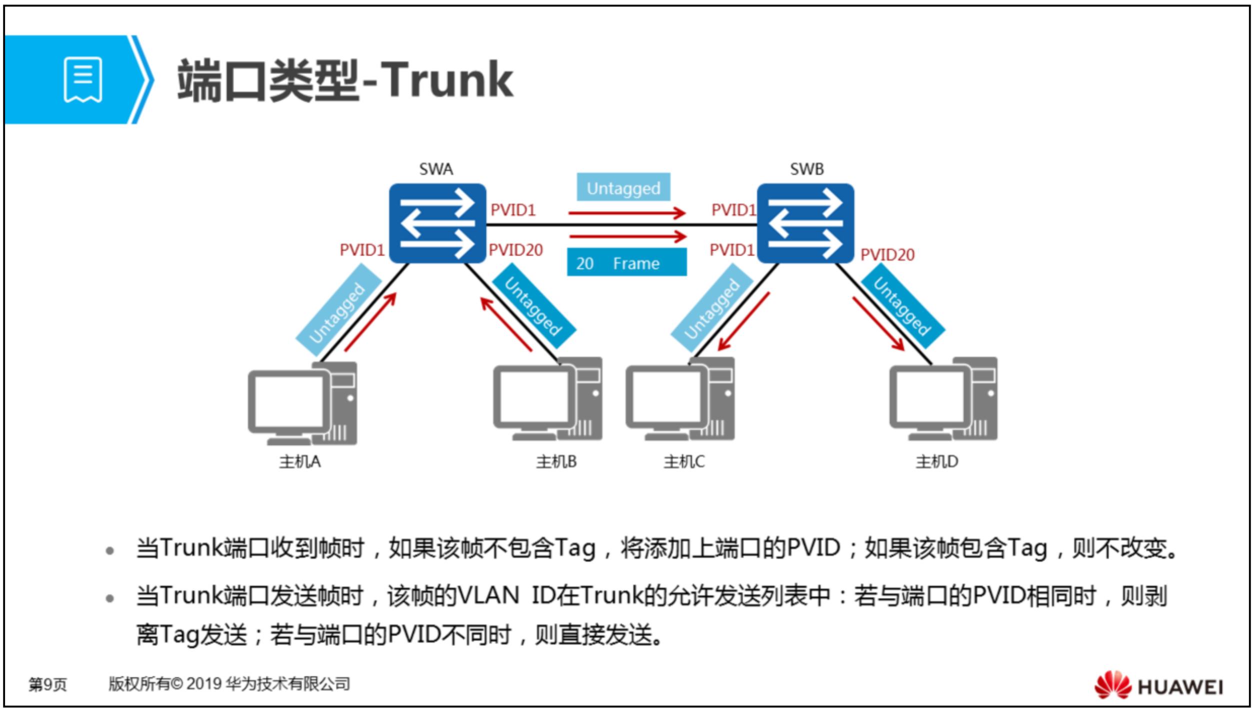
- PVID即Port VLAN ID，代表端口的缺省VLAN。交换机从对端设备收到的帧有可能是Untagged的数据帧，但所有以太网帧在交换机中都是以Tagged的形式来被处理和转发的，因此交换机必须给端口收到的Untagged数据帧添加上Tag。为了实现此目的，必须为交换机配置端口的缺省VLAN。当该端口收到Untagged数据帧时，交换机将给它加上该缺省VLAN的VLAN Tag。

端口类型-Access



- Access端口在收到数据后会添加VLAN Tag，VLAN ID和端口的PVID相同。
- Access端口在转发数据前会移除VLAN Tag。

- Access端口是交换机上用来连接用户主机的端口，它只能连接接入链路，并且只能允许唯一的VLAN ID通过本端口。
- Access端口收发数据帧的规则如下：
- 如果该端口收到对端设备发送的帧是untagged（不带VLAN标签），交换机将强制加上该端口的PVID。如果该端口收到对端设备发送的帧是tagged（带VLAN标签），交换机会检查该标签内的VLAN ID。当VLAN ID与该端口的PVID相同时，接收该报文。当VLAN ID与该端口的PVID不同时，丢弃该报文。
- Access端口发送数据帧时，总是先剥离帧的Tag，然后再发送。Access端口发往对端设备的以太网帧永远是不带标签的帧。
- 在本示例中，交换机的G0/0/1，G0/0/2，G0/0/3端口分别连接三台主机，都配置为Access端口。主机A把数据帧（未加标签）发送到交换机的G0/0/1端口，再由交换机发往其他目的地。收到数据帧之后，交换机根据端口的PVID给数据帧打上VLAN标签10，然后决定从G0/0/3端口转发数据帧。G0/0/3端口的PVID也是10，与VLAN标签中的VLAN ID相同，交换机移除标签，把数据帧发送到主机C。连接主机B的端口的PVID是2，与VLAN10不属于同一个VLAN，因此此端口不会接收到VLAN10的数据帧。

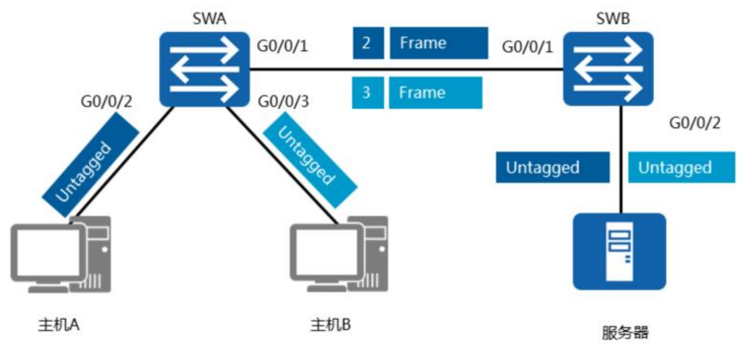


- Trunk端口是交换机上用来和其他交换机连接的端口，它只能连接干道链路。Trunk端口允许多个VLAN的帧（带Tag标记）通过。
- Trunk端口收发数据帧的规则如下：

1. 当接收到对端设备发送的不带Tag的数据帧时，会添加该端口的PVID，如果PVID在允许通过的VLAN ID列表中，则接收该报文，否则丢弃该报文。当接收到对端设备发送的带Tag的数据帧时，检查VLAN ID是否在允许通过的VLAN ID列表中。如果VLAN ID在接口允许通过的VLAN ID列表中，则接收该报文。否则丢弃该报文。
2. 端口发送数据帧时，当VLAN ID与端口的PVID相同，且是该端口允许通过的VLAN ID时，去掉Tag，发送该报文。当VLAN ID与端口的PVID不同，且是该端口允许通过的VLAN ID时，保持原有Tag，发送该报文。

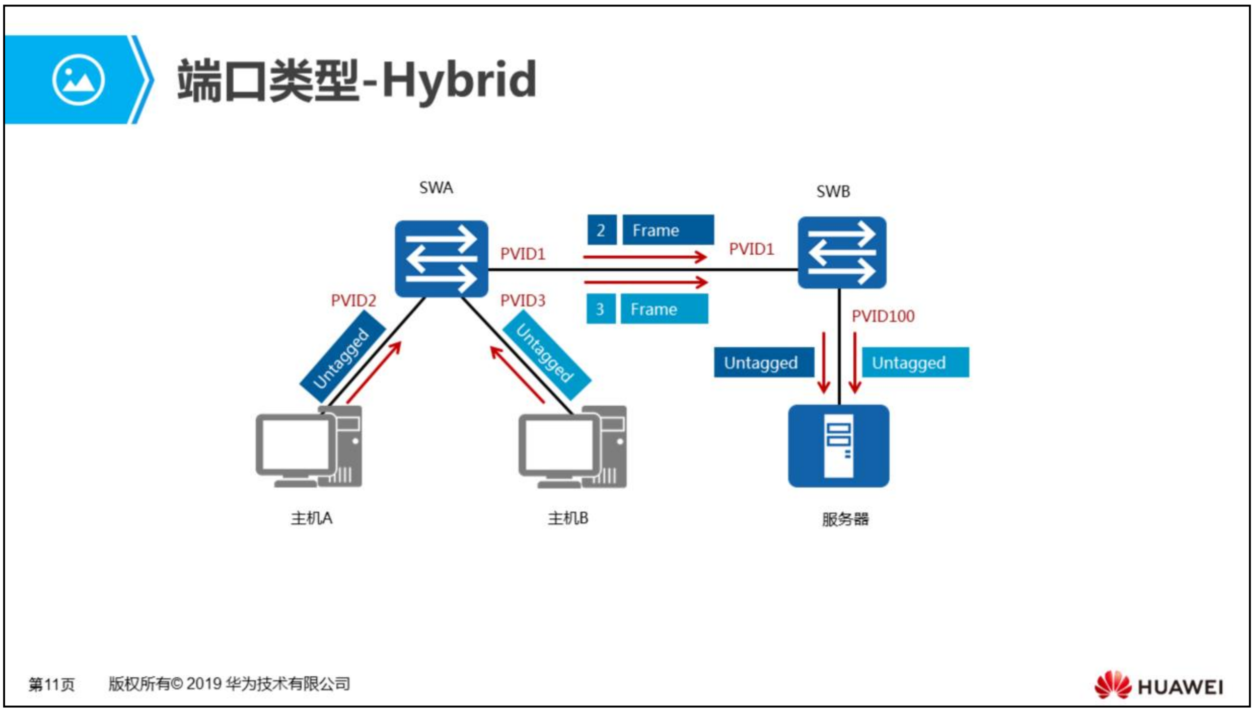
- 在本示例中，SWA和SWB连接主机的端口为Access端口，PVID如图所示。SWA和SWB互连的端口为Trunk端口，PVID都为1，此Trunk链路允许所有VLAN的流量通过。当SWA转发VLAN1的数据帧时会剥离VLAN标签，然后发送到Trunk链路上。而在转发VLAN20的数据帧时，不剥离VLAN标签直接转发到Trunk链路上。

端口类型-Hybrid



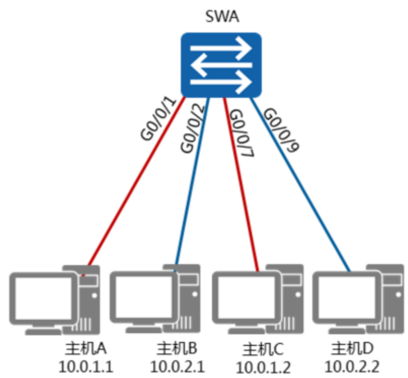
- Hybrid端口既可以连接主机，又可以连接交换机。
- Hybrid端口可以以Tagged 或Untagged方式加入VLAN 。

- Access端口发往其他设备的报文，都是Untagged数据帧，而Trunk端口仅在一种特定情况下才能发出untagged数据帧，其它情况发出的都是Tagged数据帧。
- Hybrid端口是交换机上既可以连接用户主机，又可以连接其他交换机的端口。Hybrid端口既可以连接接入链路又可以连接干道链路。Hybrid端口允许多个VLAN的帧通过，并可以在出端口方向将某些VLAN帧的Tag剥掉。华为设备默认的端口类型是Hybrid。
- 在本示例中，要求主机A和主机B都能访问服务器，但是它们之间不能互相访问。此时交换机连接主机和服务器的端口，以及交换机互连的端口都配置为Hybrid类型。交换机连接主机A的端口的PVID是2，连接主机B的端口的PVID是3，连接服务器的端口的PVID是100。



- Hybrid端口收发数据帧的规则如下：
- 当接收到对端设备发送的不带Tag的数据帧时，会添加该端口的PVID，如果PVID在允许通过的VLAN ID列表中，则接收该报文，否则丢弃该报文。当接收到对端设备发送的带Tag的数据帧时，检查VLAN ID是否在允许通过的VLAN ID列表中。如果VLAN ID在接口允许通过的VLAN ID列表中，则接收该报文，否则丢弃该报文。
- Hybrid端口发送数据帧时，将检查该接口是否允许该VLAN数据帧通过。如果允许通过，则可以通过命令配置发送时是否携带Tag。
- 配置port hybrid tagged vlan vlan-id命令后，接口发送该vlan-id的数据帧时，不剥离帧中的VLAN Tag，直接发送。该命令一般配置在连接交换机的端口上。
- 配置port hybrid untagged vlan vlan-id命令后，接口在发送vlan-id的数据帧时，会将帧中的VLAN Tag剥离掉再发送出去。该命令一般配置在连接主机的端口上。
- 本例介绍了主机A和主机B发送数据给服务器的情况。在SWA和SWB互连的端口上配置了port hybrid tagged vlan 2 3 100命令后，SWA和SWB之间的链路上传输的都是带Tag标签的数据帧。在SWB连接服务器的端口上配置了port hybrid untagged vlan 2 3，主机A和主机B发送的数据会被剥离VLAN标签后转发到服务器。

VLAN划分方法



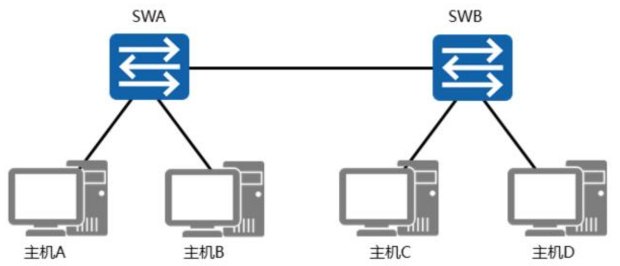
	VLAN 5	VLAN 10
基于端口	G0/0/1, G0/0/7	G0/0/2 G0/0/9
基于MAC地址	00-01-02-03-04-AA 00-01-02-03-04-CC	00-01-02-03-04-BB 00-01-02-03-04-DD
基于IP子网划分	10.0.1.*	10.0.2.*
基于协议划分	IP	IPX
基于策略	10.0.1.* + G0/0/1+ 00-01-02-03-04-AA	10.0.2.* + G0/0/2 + 00-01-02-03-04-BB

- 基于端口的VLAN划分方法在实际中最为常见。

- VLAN的划分包括如下5种方法：

1. 基于端口划分：根据交换机的端口编号来划分VLAN。通过为交换机的每个端口配置不同的PVID，来将不同端口划分到VLAN中。初始情况下，X7系列交换机的端口处于VLAN1中。此方法配置简单，但是当主机移动位置时，需要重新配置VLAN。
2. 基于MAC地址划分：根据主机网卡的MAC地址划分VLAN。此划分方法需要网络管理员提前配置网络中的主机MAC地址和VLAN ID的映射关系。如果交换机收到不带标签的数据帧，会查找之前配置的MAC地址和VLAN映射表，根据数据帧中携带的MAC地址来添加相应的VLAN标签。在使用此方法配置VLAN时，即使主机移动位置也不需要重新配置VLAN。
3. 基于IP子网划分：交换机在收到不带标签的数据帧时，根据报文携带的IP地址给数据帧添加VLAN标签。
4. 基于协议划分：根据数据帧的协议类型（或协议族类型）、封装格式来分配VLAN ID。网络管理员需要首先配置协议类型和VLAN ID之间的映射关系。
5. 基于策略划分：使用几个条件的组合来分配VLAN标签。这些条件包括IP子网、端口和IP地址等。只有当所有条件都匹配时，交换机才为数据帧添加VLAN标签。另外，针对每一条策略都是需要手工配置的。

VLAN配置




```
[SWA]vlan 10
[SWA-vlan10]quit
[SWA]vlan batch 2 to 3
Info: This operation may take a few seconds. Please wait for a
moment...done.
```


- 在交换机上划分VLAN时，需要首先创建VLAN。在交换机上执行**vlan** <vlan-id>命令，创建VLAN。如本例所示，执行**vlan** 10命令后，就创建了VLAN 10，并进入了VLAN 10视图。VLAN ID的取值范围是1到4094。如需创建多个VLAN，可以在交换机上执行**vlan batch** { *vlan-id1* [**to** *vlan-id2*] }命令，以创建多个连续的VLAN。也可以执行**vlan batch** { *vlan-id1 vlan-id2* }命令，创建多个不连续的VLAN，VLAN号之间需要有空格。

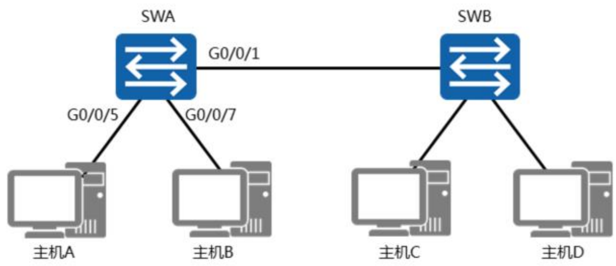
配置验证

```
[SWA]display vlan
The total number of vlans is : 4
------------------------------------------------------------
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
--------------------------------------------------------------
VID  Type    Ports
--------------------------------------------------------------
1    common  UT:GE0/0/1(U) ……
2    common
3    common
10   common
……
```

  HUAWEI

- 创建VLAN后，可以执行**display vlan**命令验证配置结果。如果不指定任何参数，则该命令将显示所有VLAN的简要信息。
- 执行**display vlan** [*vlan-id* [**verbose**]]命令，可以查看指定VLAN的详细信息，包括VLAN ID、类型、描述、VLAN的状态、VLAN中的端口、以及VLAN中端口的模式等。
- 执行**display vlan** *vlan-id* **statistics**命令，可以查看指定VLAN中的流量统计信息。
- 执行**display vlan summary**命令，可以查看系统中所有VLAN的汇总信息。

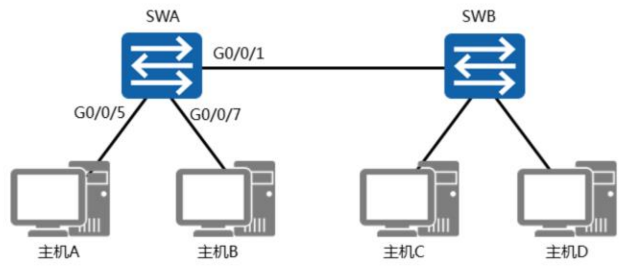
配置Access端口




```
[SWA]interface GigabitEthernet 0/0/5
[SWA-GigabitEthernet0/0/5]port link-type access
[SWA-GigabitEthernet0/0/5]interface GigabitEthernet 0/0/7
[SWA-GigabitEthernet0/0/7]port link-type access
```


- 华为X7系列交换机上，默认的端口类型是hybrid。
- 配置端口类型的命令是**port link-type** *<type>*，*type*可以配置为Access，Trunk或Hybrid。需要注意的是，如果查看端口配置时没有发现端口类型信息，说明端口使用了默认的hybrid端口链路类型。当修改端口类型时，必须先恢复端口的默认VLAN配置，使端口属于缺省的VLAN 1。

添加端口到VLAN




```
[SWA]vlan 2
[SWA-vlan2]port GigabitEthernet 0/0/7
[SWA-vlan2]quit
[SWA]interface GigabitEthernet0/0/5
[SWA-GigabitEthernet0/0/5]port default vlan 3
```


- 可以使用两种方法把端口加入到VLAN。

1. 第一种方法是进入到VLAN视图，执行port <interface>命令，把端口加入VLAN。
2. 第二种方法是进入到接口视图，执行port default vlan <vlan-id>命令，把端口加入VLAN。vlan-id是指端口要加入的VLAN。

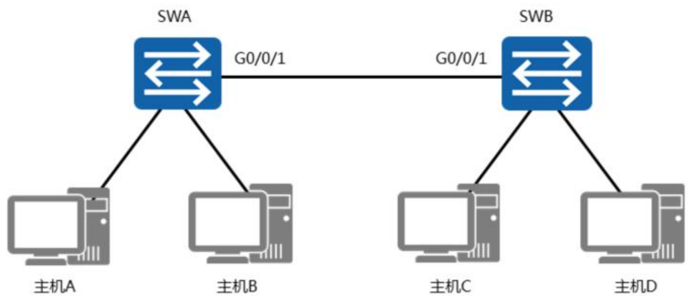
配置验证

```
[SWA]display vlan
The total number of vlans is : 4
------------------------------------------------------------
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
--------------------------------------------------------------
VID  Type    Ports
--------------------------------------------------------------
1    common  UT:GE0/0/1(U)  ……
2    common  UT:GE0/0/7(U)
3    common  UT:GE0/0/5(U)
10   common
……
```

  HUAWEI

- 执行**display vlan**命令，可以确认端口是否已经加入到VLAN中。在本示例中，端口GigabitEthernet0/0/5和GigabitEthernet0/0/7分别加入了VLAN 3和VLAN 2。**UT**表明该端口发送数据帧时，会剥离VLAN标签，即此端口是一个Access端口或不带标签的Hybrid端口。**U**或**D**分别表示链路当前是Up状态或Down状态。

配置Trunk端口




```
[SWA-GigabitEthernet0/0/1]port link-type trunk
[SWA-GigabitEthernet0/0/1]port trunk allow-pass vlan 2 3
```

  HUAWEI

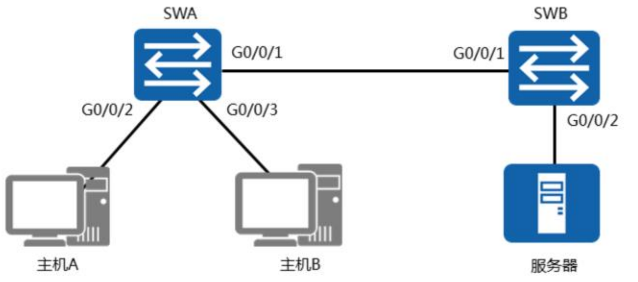
- 配置Trunk时，应先使用**port link-type trunk**命令修改端口的类型为Trunk，然后再配置Trunk端口允许哪些VLAN的数据帧通过。执行**port trunk allow-pass vlan** { { *vlan-id1* [**to** *vlan-id2*] } | **all** }命令，可以配置端口允许的VLAN，all表示允许所有VLAN的数据帧通过。
- 执行**port trunk pvid vlan** *vlan-id*命令，可以修改Trunk端口的PVID。修改Trunk端口的PVID之后，需要注意：缺省VLAN不一定是端口允许通过的VLAN。只有使用命令**port trunk allow-pass vlan** { { *vlan-id1* [**to** *vlan-id2*] } | **all** }允许缺省VLAN数据通过，才能转发缺省VLAN的数据帧。交换机的所有端口默认允许VLAN1的数据通过。
- 在本示例中，将SWA的G0/0/1端口配置为Trunk端口，该端口PVID默认为1。配置**port trunk allow-pass vlan** 2 3命令之后，该Trunk允许VLAN 2和VLAN 3的数据流量通过。

配置验证

```
[SWA]display vlan
The total number of vlans is : 4
-------------------------------------------------------------
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping;
ST:Vlan-stacking; #: ProtocolTransparent-vlan; *:Management-
vlan;
---------------------------------------------------------------
VID  Type    Ports
---------------------------------------------------------------
1    common  UT:GE0/0/1(U) ……
2    common  UT:GE0/0/7(D) TG:GE0/0/1(U)
3    common  UT:GE0/0/5(U) TG:GE0/0/1(U)
10   common
……
```


- 执行**display vlan**命令可以查看修改后的配置。TG表明该端口在转发对应VLAN的数据帧时，不会剥离标签，直接进行转发，该端口可以是Trunk端口或带标签的Hybrid端口。本示例中，GigabitEthernet0/0/1在转发VLAN 2和VLAN3的流量时，不剥离标签，直接转发。

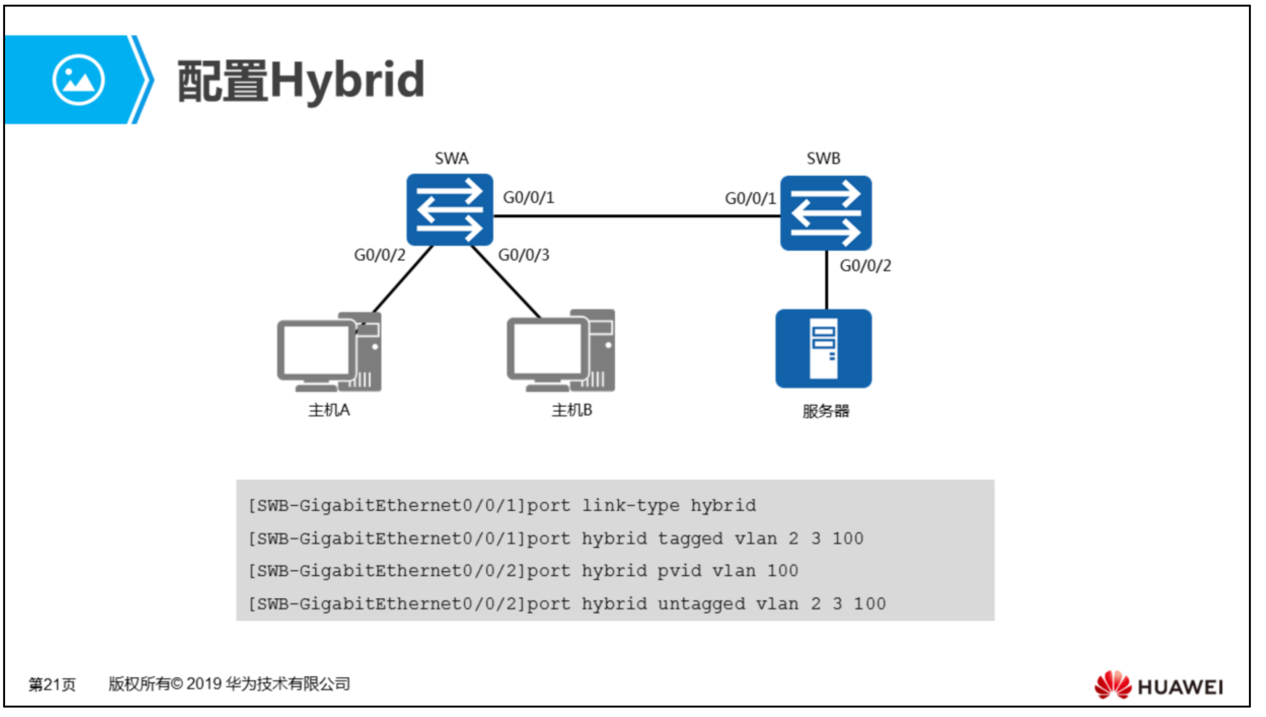
配置Hybrid端口




```
[SWA-GigabitEthernet0/0/1]port link-type hybrid
[SWA-GigabitEthernet0/0/1]port hybrid tagged vlan 2 3 100
[SWA-GigabitEthernet0/0/2]port hybrid pvid vlan 2
[SWA-GigabitEthernet0/0/2]port hybrid untagged vlan 2 100
[SWA-GigabitEthernet0/0/3]port hybrid pvid vlan 3
[SWA-GigabitEthernet0/0/3]port hybrid untagged vlan 3 100
```

  HUAWEI

- **port link-type hybrid**命令的作用是将端口的类型配置为Hybrid。默认情况下，X7系列交换机的端口类型是Hybrid。因此，只有在把Access口或Trunk口配置成Hybrid时，才需要执行此命令。
- **port hybrid tagged vlan**{ { *vlan-id1* [**to** *vlan-id2*] } | **all** }命令用来配置允许哪些VLAN的数据帧以Tagged方式通过该端口。
- **port hybrid untagged vlan** { { *vlan-id1* [**to** *vlan-id2*] } | **all** }命令用来配置允许哪些VLAN的数据帧以Untagged方式通过该端口。
- 在本示例中，要求主机A和主机B都能访问服务器，但是它们之间不能互相访问。此时通过命令port link-type hybrid配置交换机连接主机和服务器的端口，以及交换机互连的端口都为Hybrid类型。通过命令port hybrid pvid vlan 2配置交换机连接主机A的端口的PVID是2。类似地，连接主机B的端口的PVID是3，连接服务器的端口的PVID是100。
- 通过在G0/0/1端口下使用命令port hybrid tagged vlan 2 3 100，配置VLAN2,VLAN3和VLAN100的数据帧在通过该端口时都携带标签。在G0/0/2端口下使用命令port hybrid untagged vlan 2 100，配置VLAN2和VLAN100的数据帧在通过该端口时都不携带标签。在G0/0/3端口下使用命令port hybrid untagged vlan 3 100，配置VLAN3和VLAN100的数据帧在通过该端口时都不携带标签。



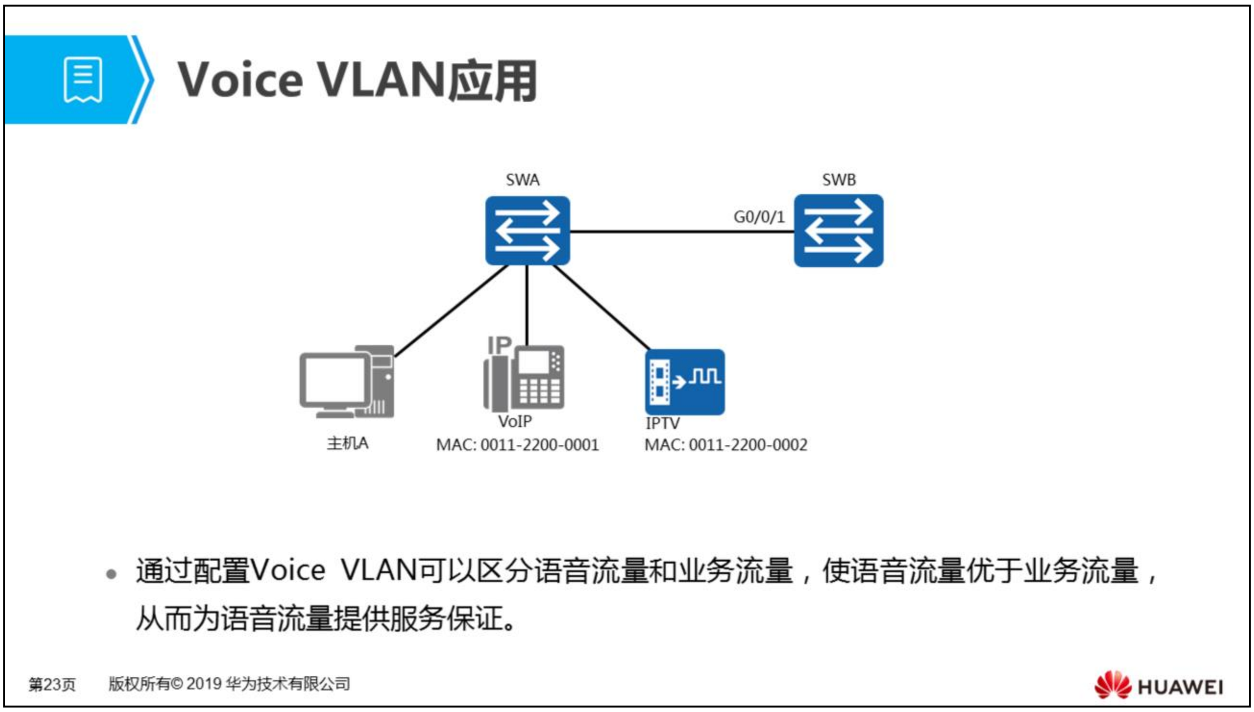

```
[SWB-GigabitEthernet0/0/1]port link-type hybrid
[SWB-GigabitEthernet0/0/1]port hybrid tagged vlan 2 3 100
[SWB-GigabitEthernet0/0/2]port hybrid pvid vlan 100
[SWB-GigabitEthernet0/0/2]port hybrid untagged vlan 2 3 100
```

- 在SWB上继续进行配置，在G0/0/1端口下使用命令**port link-type hybrid**配置端口类型为Hybrid。
- 在G0/0/1端口下使用命令**port hybrid tagged vlan** 2 3 100，配置VLAN2，VLAN3和VLAN100的数据帧在通过该端口时都携带标签。
- 在G0/0/2端口下使用命令**port hybrid untagged vlan** 2 3 100，配置VLAN2，VLAN3和VLAN100的数据帧在通过该端口时都不携带标签。

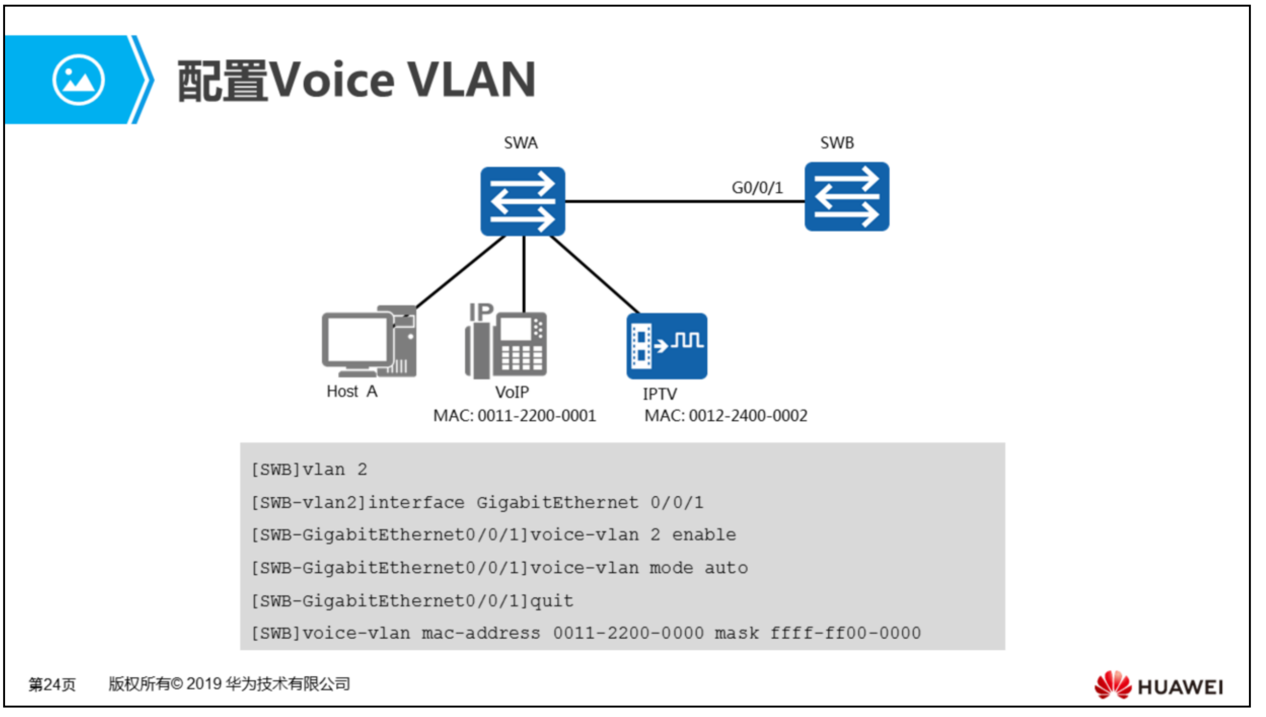
配置验证

```
[SWA]display vlan
The total number of vlans is : 4
------------------------------------------------------------
U:Up; D:Down; TG:Tagged; UT:Untagged; MP:Vlan-mapping; ST:Vlan-
stacking; #: ProtocolTransparent-vlan; *:Management-vlan;
1    common    UT:GE0/0/1(U)……
2    common    UT:GE0/0/2(U)
               TG:GE0/0/1(U)
3    common    UT:GE0/0/3(U)
               TG:GE0/0/1(U)
100   common   UT:GE0/0/2(U)      GE0/0/3(U)
               TG:GE0/0/1(U)
```

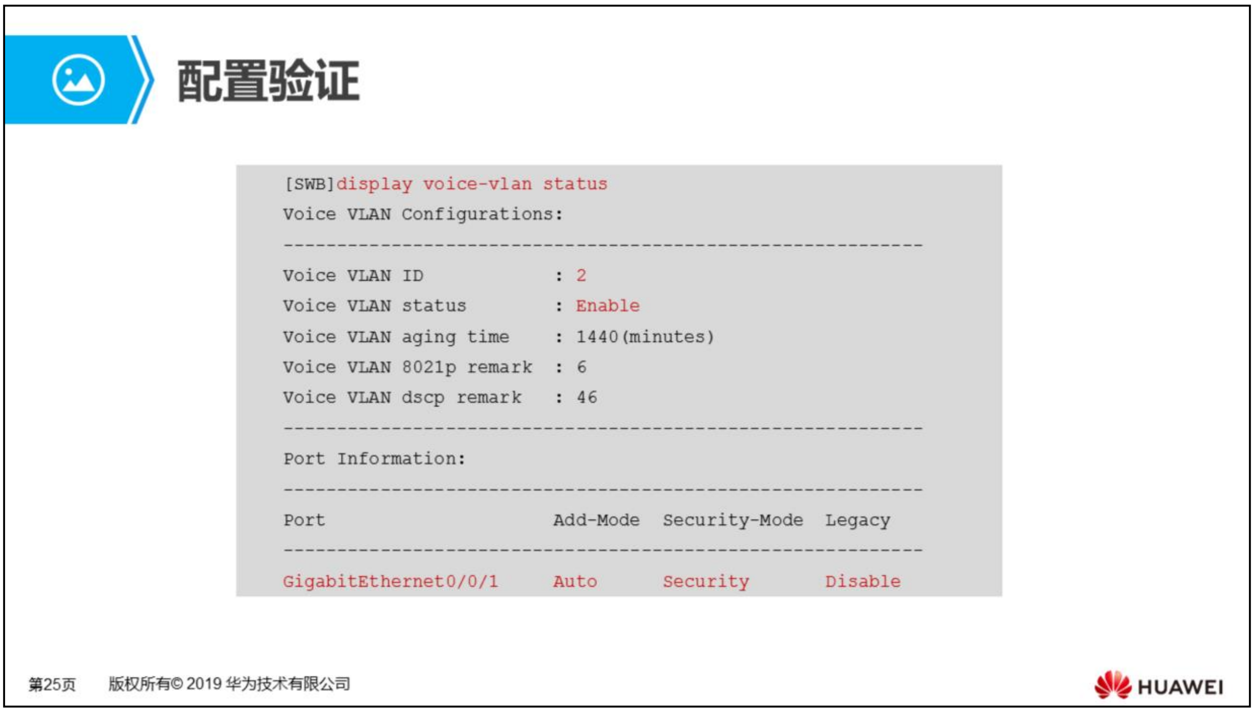

- 在SWA上执行display vlan命令，可以查看hybrid端口的配置。在本示例中，GigabitEthernet 0/0/2在发送VLAN2和VLAN100的数据帧时会剥离标签。GigabitEthernet 0/0/3在发送VLAN3和VLAN100的数据帧时会剥离标签。GigabitEthernet 0/0/1允许VLAN 2，VLAN 3和VLAN 100的带标签的数据帧通过。此配置满足了多个VLAN可以访问特定VLAN，而其他VLAN间不允许互相访问的需求。



- 随着IP网络的融合，TCP/IP网络可以为高速上网HSI（High Speed Internet）业务、VoIP（Voice over IP）业务、IPTV（Internet Protocol Television）业务提供服务。
- 语音数据在传输时需要具有比其他业务数据更高的优先级，以减少传输过程中可能产生的时延和丢包现象。
- 为了区分语音数据流，可在交换机上部署Voice VLAN功能，把VoIP的电话流量进行VLAN隔离，并配置更高的优先级，从而能够保证通话质量。



- 执行voice-vlan <vlan-id> enable命令，可以把VLAN 2到VLAN 4094之间的任一VLAN配置成语音VLAN。
- 执行voice-vlan mode <mode>命令，可以配置端口加入语音VLAN的模式。
- 端口加入Voice VLAN的模式有两种：
- 自动模式：使能Voice VLAN功能的端口根据进入端口的数据流中的源MAC地址字段来判断该数据流是否为语音数据流。源MAC地址符合系统设置的语音设备OUI（Organizationally Unique Identifier）地址的报文认为是语音数据流。接收到语音数据流的端口将自动加入Voice VLAN中传输，并通过老化机制维护Voice VLAN内的端口数量。
- 手动模式：当接口使能Voice VLAN功能后，必须通过手工将连接语音设备的端口加入或退出Voice VLAN中，这样才能保证Voice VLAN功能生效。
- 执行voice-vlan mac-address mac-address mask oui-mask [description text]命令，用来配置Voice VLAN的OUI地址。OUI地址表示一个MAC地址段。交换机将48位的MAC地址和掩码的对应位做“与”运算可以确定出OUI地址。接入设备的MAC地址和OUI地址匹配的位数，由掩码中全“1”的长度决定。例如，MAC地址为0001–0001–0001，掩码为FFFF-FF00–0000，那么将MAC地址与其相应掩码位执行“与”运算的结果就是OUI地址0001–0000–0000。只要接入设备的MAC地址前24位和OUI地址的前24位匹配，那么使能Voice VLAN功能的端口将认为此数据流是语音数据流，接入的设备是语音设备。



配置验证

```
[SWB]display voice-vlan status
Voice VLAN Configurations:
-----------------------------------------------------------
Voice VLAN ID              : 2
Voice VLAN status          : Enable
Voice VLAN aging time      : 1440(minutes)
Voice VLAN 8021p remark    : 6
Voice VLAN dscp remark     : 46
-----------------------------------------------------------
Port Information:
-----------------------------------------------------------
Port                     Add-Mode  Security-Mode  Legacy
-----------------------------------------------------------
GigabitEthernet0/0/1     Auto      Security       Disable
```

第25页　版权所有© 2019 华为技术有限公司　HUAWEI

- 执行**display voice-vlan status**命令，可以查看语音VLAN的信息，包括状态、工作模式、老化时间、以及使能了语音VLAN功能的端口信息。
- **Add-Mode**字段表明语音VLAN的添加模式。自动模式中，使能了语音VLAN功能后，端口可以自动加入到语音VLAN。如果语音设备发送的报文的MAC地址匹配了OUI，连接该语音设备的端口也会加入语音VLAN。如果在老化时间内，端口没有收到语音设备的任何语音数据报文，端口自动会被删除。手动模式中，在端口上使能了语音VLAN功能之后，必须手动把端口添加到语音VLAN中。
- **Security-Mode**字段表示Voice VLAN端口的工作模式，有两种：

1. 正常模式：可以传输语音数据和业务数据，但是容易受到恶意数据流量的攻击。
2. 安全模式：只允许传输语音数据流。安全模式可以防止Voice VLAN受到恶意数据流量的攻击，但是检查报文的工作会占用一定的系统资源。

- **Legacy**字段表明端口是否开启与其他厂商语音设备互通的功能，Enable表示开启，Disable表示关闭。

本章总结

- 如果一个Trunk链路PVID是5，且端口下配置port trunk allow-pass vlan 2 3，那么哪些VLAN的流量可以通过该Trunk链路进行传输？
- PVID为2的Access端口收到一个不带标记的帧会采取什么样的动作？

- 执行了port trunk allow-pass vlan 2 3命令后，VLAN 5的数据帧不能在此链路上进行传输。VLAN 1的数据默认也可以通过Trunk链路进行传输。所以VLAN 1，VLAN 2和VLAN 3的数据帧可以在Trunk链路上传输。
- 收到不带标签的数据帧后，PVID为2的Access端口会给数据帧打上VLAN 2的标签。然后交换机会根据标签和目的MAC地址将数据帧发送到其他端口。


谢 谢
www.huawei.com

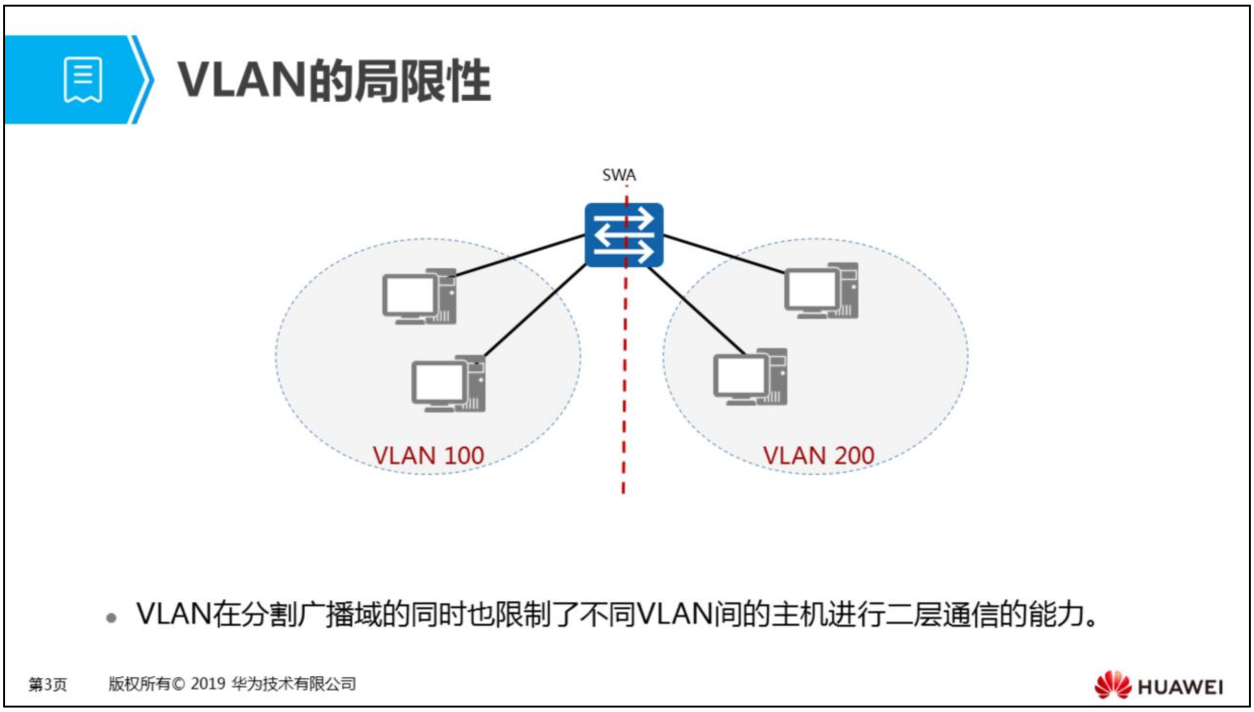
VLAN间路由

前言

- 部署了VLAN的传统交换机不能实现不同VLAN间的二层报文转发，因此必须引入路由技术来实现不同VLAN间的通信。VLAN路由可以通过二层交换机配合路由器来实现，也可以通过三层交换机来实现。

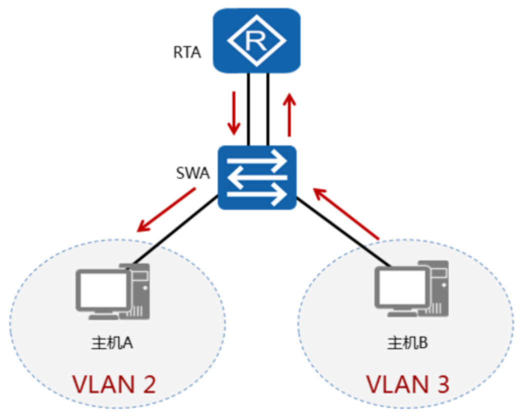
目标

- 学完本课程后，您将能够：
  - 掌握VLAN路由的应用场景
  - 掌握VLAN路由的工作原理
  - 掌握VLAN路由的基本配置



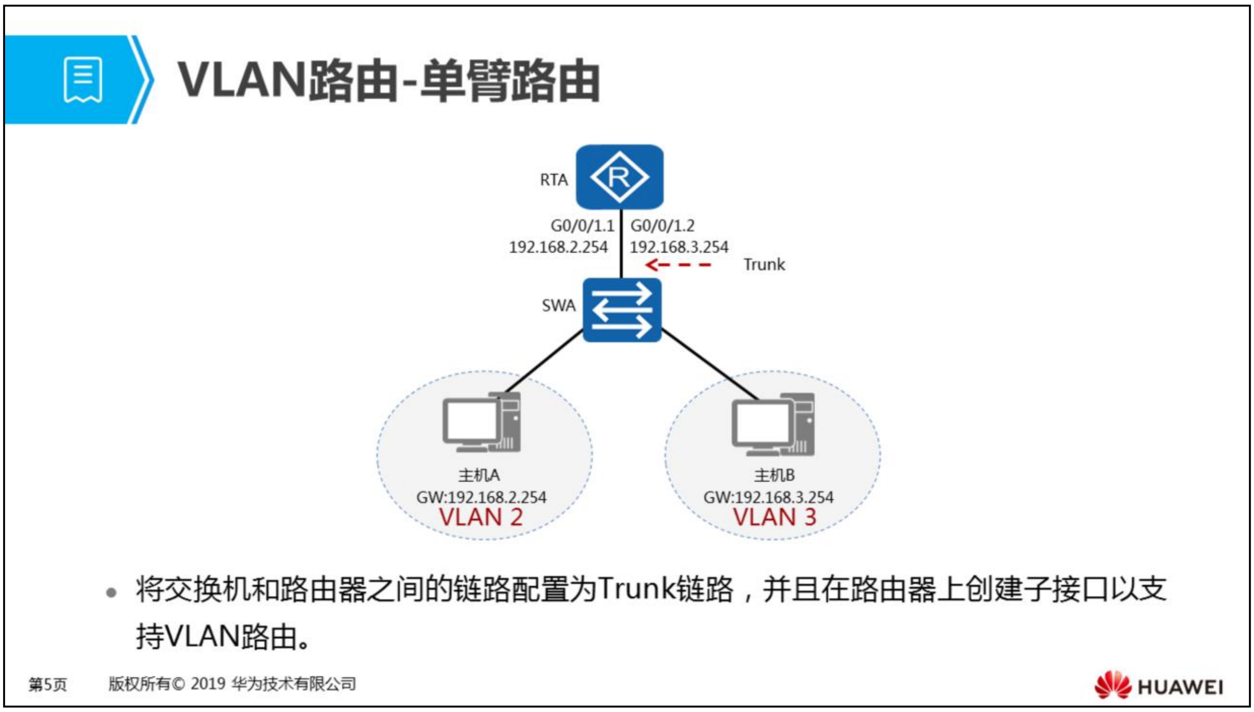
- VLAN隔离了二层广播域，也严格地隔离了各个VLAN之间的任何二层流量，属于不同VLAN的用户之间不能进行二层通信。

VLAN路由-每个VLAN一个物理连接



- 在二层交换机上配置VLAN，每一个VLAN使用一条独占的物理链路连接到路由器的一个接口上。

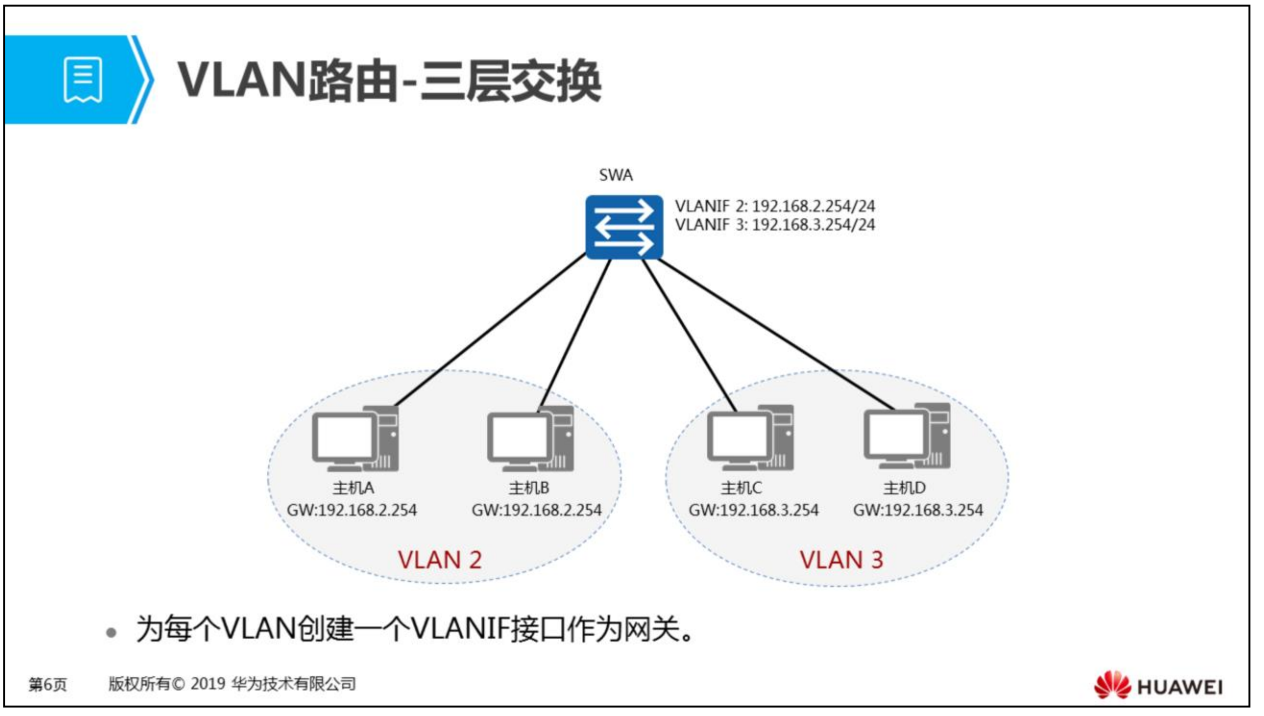
- 因为不同VLAN之间的主机是无法实现二层通信的，所以必须通过三层路由才能将报文从一个VLAN转发到另外一个VLAN。
- 解决VLAN间通信问题的第一种方法是：在路由器上为每个VLAN分配一个单独的接口，并使用一条物理链路连接到二层交换机上。当VLAN间的主机需要通信时，数据会经由路由器进行三层路由，并被转发到目的VLAN内的主机，这样就可以实现VLAN之间的相互通信。
- 然而，随着每个交换机上VLAN数量的增加，这样做必然需要大量的路由器接口，而路由器的接口数量是极其有限的。并且，某些VLAN之间的主机可能不需要频繁进行通信，如果这样配置的话，会导致路由器的接口利用率很低。因此，实际应用中一般不会采用这种方案来解决VLAN间的通信问题。



- 解决VLAN间通信问题的第二种方法是：
- 在交换机和路由器之间仅使用一条物理链路连接。在交换机上，把连接到路由器的端口配置成Trunk类型的端口，并允许相关VLAN的帧通过。在路由器上需要创建子接口，逻辑上把连接路由器的物理链路分成了多条。一个子接口代表了一条归属于某个VLAN的逻辑链路。配置子接口时，需要注意以下几点：

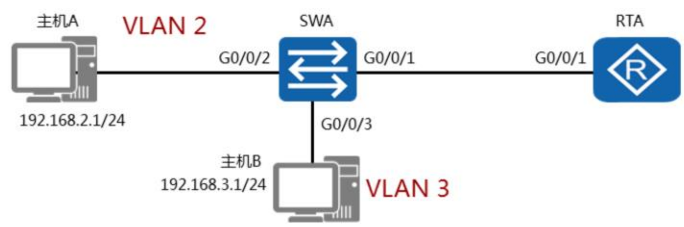
1. 必须为每个子接口分配一个IP地址。该IP地址与子接口所属VLAN位于同一网段。
2. 需要在子接口上配置802.1Q封装，来剥掉和添加VLAN Tag，从而实现VLAN间互通。
3. 在子接口上执行命令**arp broadcast enable**使能子接口的ARP广播功能。

- 本例中，主机A发送数据给主机B时，RTA会通过G0/0/1.1子接口收到此数据，然后查找路由表，将数据从G0/0/1.2子接口发送给主机B，这样就实现了VLAN2和VLAN3之间的主机通信。



- 解决VLAN间通信问题的第三种方法是：
- 在三层交换机上配置VLANIF接口来实现VLAN间路由。如果网络上有多个VLAN，则需要给每个VLAN配置一个VLANIF接口，并给每个VLANIF接口配置一个IP地址。用户设置的缺省网关就是三层交换机中VLANIF接口的IP地址。

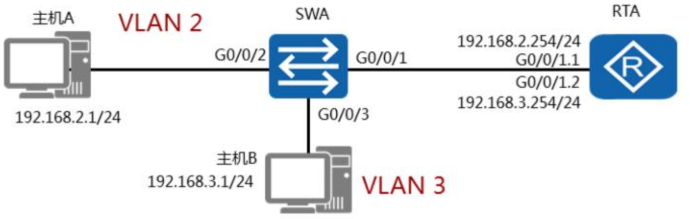
配置单臂路由




```
[SWA]vlan batch 2 3
[SWA-GigabitEthernet0/0/1]port link-type trunk
[SWA-GigabitEthernet0/0/1]port trunk allow-pass vlan 2 3
[SWA-GigabitEthernet0/0/2]port link-type access
[SWA-GigabitEthernet0/0/2]port default vlan 2
[SWA-GigabitEthernet0/0/3]port link-type access
[SWA-GigabitEthernet0/0/3]port default vlan 3
```


- 执行**port link-type trunk**命令，配置SWA的G0/0/1端口为Trunk类型的端口。
- 执行**port trunk allow-pass vlan 2 3**命令，配置SWA的G0/0/1端口允许VLAN 2和VLAN 3的数据通过。

配置单臂路由




```
[RTA]interface GigabitEthernet0/0/1.1
[RTA-GigabitEthernet0/0/1.1]dot1q termination vid 2
[RTA-GigabitEthernet0/0/1.1]ip address 192.168.2.254 24
[RTA-GigabitEthernet0/0/1.1]arp broadcast enable
[RTA]interface GigabitEthernet0/0/1.2
[RTA-GigabitEthernet0/0/1.2]dot1q termination vid 3
[RTA-GigabitEthernet0/0/1.2]ip address 192.168.3.254 24
[RTA-GigabitEthernet0/0/1.2]arp broadcast enable
```


- **interface** *interface-type interface-number.sub-interface number*命令用来创建子接口。*sub-interface number*代表物理接口内的逻辑接口通道。
- **dot1q termination** *vid*命令用来配置子接口dot1q封装的单层VLAN ID。缺省情况，子接口没有配置dot1q封装的单层VLAN ID。本命令执行成功后，终结子接口对报文的处理如下：接收报文时，剥掉报文中携带的Tag后进行三层转发。转发出去的报文是否带Tag由出接口决定。发送报文时，将相应的VLAN信息添加到报文中再发送。
- **arp broadcast enable**命令用来使能终结子接口的ARP广播功能。缺省情况下，终结子接口没有使能ARP广播功能。终结子接口不能转发广播报文，在收到广播报文后它们直接把该报文丢弃。为了允许终结子接口能转发广播报文，可以通过在子接口上执行此命令。

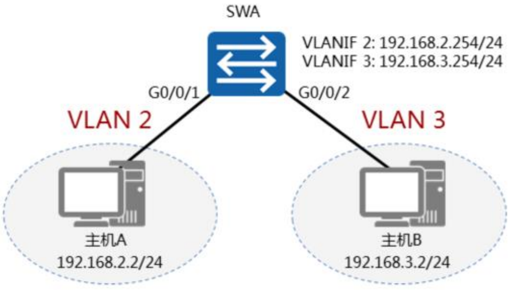
配置验证

```
Host A>ping 192.168.3.1

Ping 192.168.3.1: 32 data bytes, Press Ctrl_C to break
From 192.168.3.1: bytes=32 seq=1 ttl=127 time=15 ms
From 192.168.3.1: bytes=32 seq=2 ttl=127 time=15 ms
From 192.168.3.1: bytes=32 seq=3 ttl=127 time=32 ms
From 192.168.3.1: bytes=32 seq=4 ttl=127 time=16 ms
From 192.168.3.1: bytes=32 seq=5 ttl=127 time=31 ms

--- 192.168.3.1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 15/21/32 ms
```

  HUAWEI

- 配置完成单臂路由后，可以使用**ping**命令来验证主机之间的连通性。如上所示，VLAN2中的主机A(IP地址：192.168.2.1）可以Ping通VLAN3中的主机B(IP地址：192.168.3.1)。

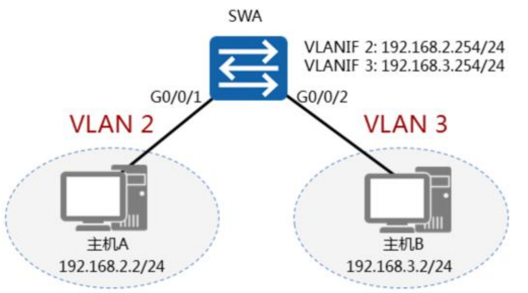
配置三层交换




```
[SWA]vlan batch 2 3
[SWA-GigabitEthernet0/0/1]port link-type access
[SWA-GigabitEthernet0/0/1]port default vlan 2
[SWA-GigabitEthernet0/0/2]port link-type access
[SWA-GigabitEthernet0/0/2]port default vlan 3
```


- 在三层交换机上配置VLAN路由时，首先创建VLAN，并将端口加入到VLAN中。

配置三层交换




```
[SWA]interface vlanif 2
[SWA-Vlanif2]ip address 192.168.2.254 24
[SWA-Vlanif2]quit
[SWA]interface vlanif 3
[SWA-Vlanif3]ip address 192.168.3.254 24
[SWA-Vlanif3]quit
```


- **interface vlanif** *vlan-id*命令用来创建VLANIF接口并进入到VLANIF接口视图。*vlan-id*表示与VLANIF接口相关联的VLAN编号。VLANIF接口的IP地址作为主机的网关IP地址，和主机的IP地址必须位于同一网段。

配置验证

```
Host A>ping 192.168.3.2

Ping 192.168.3.2: 32 data bytes, Press Ctrl_C to break
From 192.168.3.2: bytes=32 seq=1 ttl=127 time=15 ms
From 192.168.3.2: bytes=32 seq=2 ttl=127 time=15 ms
From 192.168.3.2: bytes=32 seq=3 ttl=127 time=32 ms
From 192.168.3.2: bytes=32 seq=4 ttl=127 time=16 ms
From 192.168.3.2: bytes=32 seq=5 ttl=127 time=31 ms

--- 192.168.3.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 15/21/32 ms
```

  HUAWEI

- 配置三层交换后，可以用**ping**命令验证主机之间的连通性。如上所示，VLAN2中的主机A(IP地址：192.168.2.2）可以Ping通VLAN3中的主机B(IP地址：192.168.3.2)。

本章总结

- 配置命令dot1q termination vid <vlan-id> 的目的是什么？
- 配置单臂路由时，交换机连接路由器的接口需要哪些配置？

  HUAWEI

- 1. dot1q termination vid vlan-id命令有两个功能。第一个功能是删除VLAN标签。接口在收到VLAN报文后，剥掉报文中携带的Tag后进行三层转发。第二个功能是添加VLAN标签。接口在发送报文时，将相应的VLAN信息添加到报文中再发送。
- 2. 必须把接口配置成Trunk口，并允许相应VLAN的数据通过。


谢 谢
www.huawei.com

HDLC&PPP原理与配置

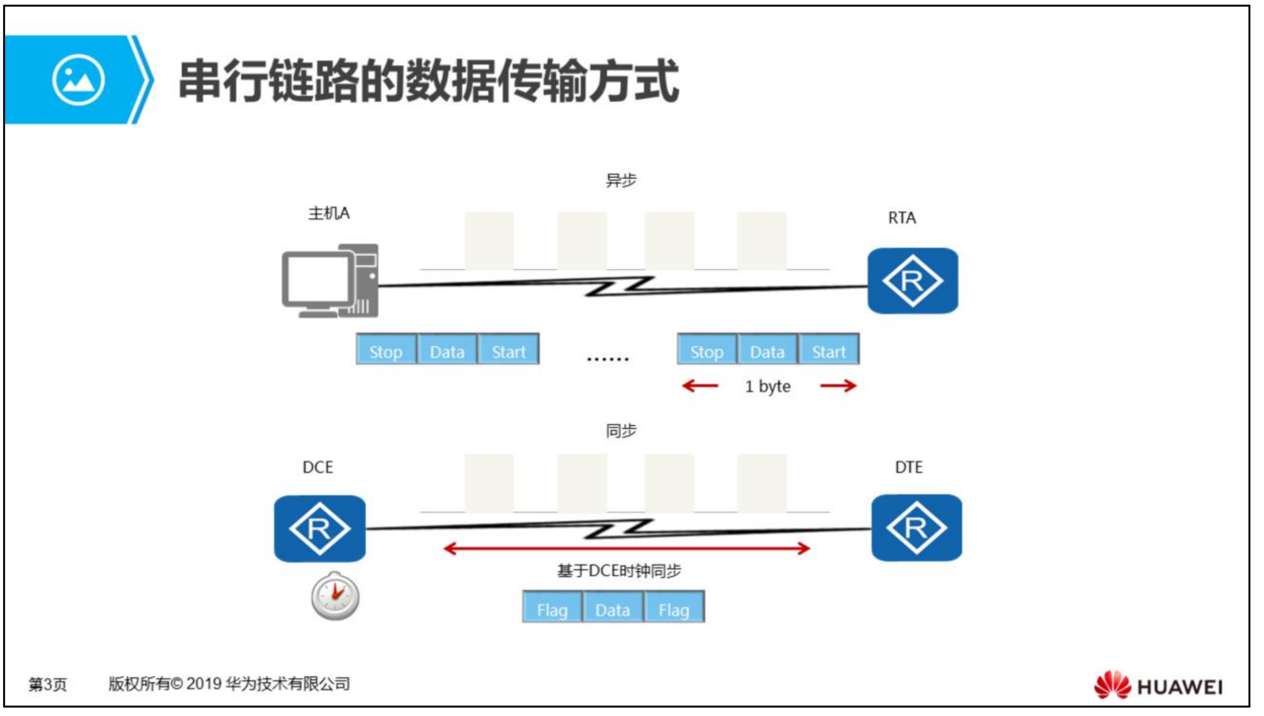
HUAWEI

前言

- 广域网中经常会使用串行链路来提供远距离的数据传输，高级数据链路控制HDLC（High-Level Data Link Control）和点对点协议PPP（ Point to Point Protocol）是两种典型的串口封装协议。

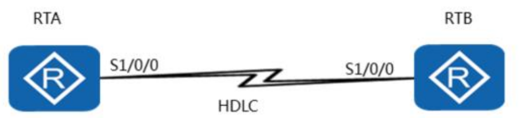
目标

- 学完本课程后，您将能够：
  - 掌握HDLC和PPP的工作原理
  - 掌握HDLC和PPP的基本配置



- 串行链路普遍用于广域网中。串行链路中定义了两种数据传输方式：异步和同步。
- 异步传输是以字节为单位来传输数据，并且需要采用额外的起始位和停止位来标记每个字节的开始和结束。起始位为二进制值0，停止位为二进制值1。在这种传输方式下，开始和停止位占据发送数据的相当大的比例，每个字节的发送都需要额外的开销。
- 同步传输是以帧为单位来传输数据，在通信时需要使用时钟来同步本端和对端的设备通信。DCE即数据通信设备，它提供了一个用于同步DCE设备和DTE设备之间数据传输的时钟信号。DTE即数据终端设备，它通常使用DCE产生的时钟信号。

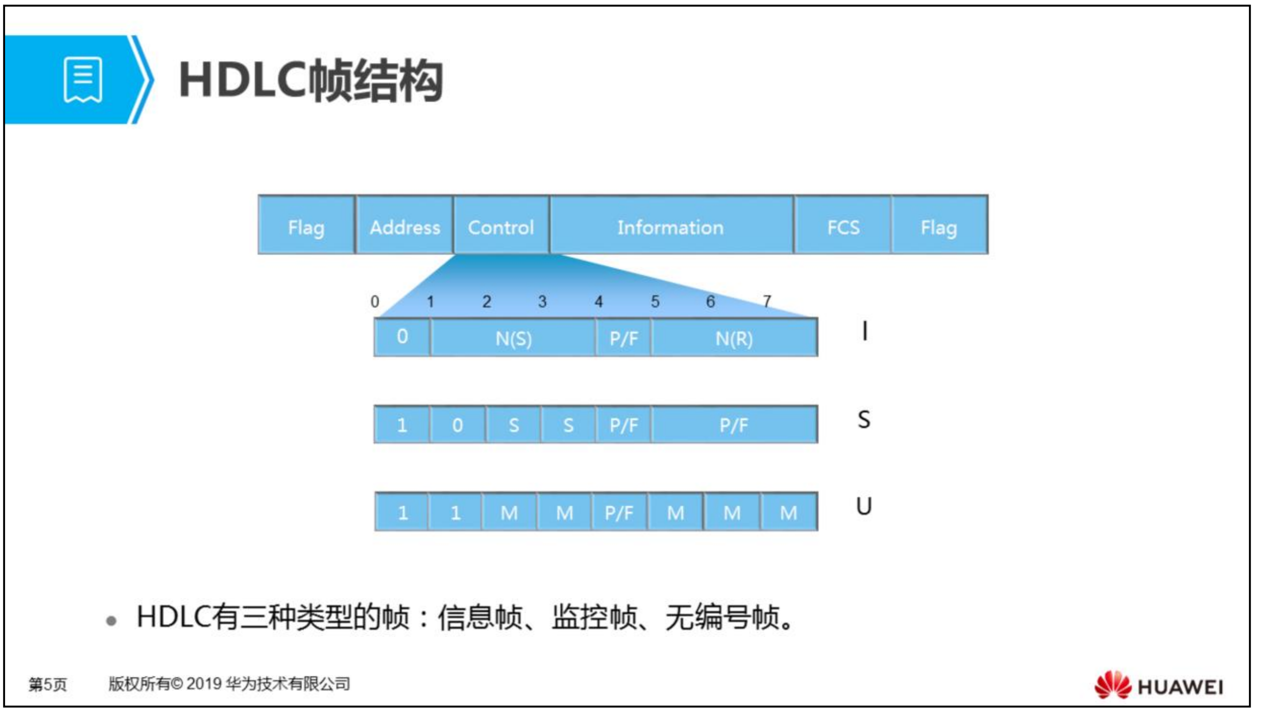
HDLC协议应用



- High-level Data Link Control，高级数据链路控制，简称HDLC，是一种面向比特的链路层协议。

- ISO制定的HDLC是一种面向比特的通信规则。HDLC传送的信息单位为帧。作为面向比特的同步数据控制协议的典型，HDLC具有如下特点：

1. 协议不依赖于任何一种字符编码集；
2. 数据报文可透明传输，用于透明传输的“0比特插入法”易于硬件实现；
3. 全双工通信，不必等待确认可连续发送数据，有较高的数据链路传输效率；
4. 所有帧均采用CRC校验，并对信息帧进行编号，可防止漏收或重收，传输可靠性高；
5. 传输控制功能与处理功能分离，具有较大的灵活性和较完善的控制功能。



- 完整的HDLC帧由标志字段（F）、地址字段（A）、控制字段（C）、信息字段（I）、帧校验序列字段（FCS）等组成。

1. 标志字段为01111110，用以标志帧的开始与结束，也可以作为帧与帧之间的填充字符。
2. 地址字段携带的是地址信息。
3. 控制字段用于构成各种命令及响应，以便对链路进行监视与控制。发送方利用控制字段来通知接收方来执行约定的操作；相反，接收方用该字段作为对命令的响应，报告已经完成的操作或状态的变化。
4. 信息字段可以包含任意长度的二进制数，其上限由FCS字段或通讯节点的缓存容量来决定，目前用得较多的是1000-2000比特，而下限可以是0，即无信息字段。监控帧中不能有信息字段。
5. 帧检验序列字段可以使用16位CRC对两个标志字段之间的内容进行校验。

- HDLC有三种类型的帧：

1. 信息帧（I帧）用于传送有效信息或数据，通常简称为I帧。
2. 监控帧（S帧）用于差错控制和流量控制，通常称为S帧。S帧的标志是控制字段的前两个比特位为“10”。S帧不带信息字段，只有6个字节即48个比特。

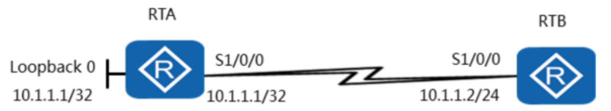
3. 无编号帧（U帧）简称U帧。U帧用于提供对链路的建立、拆除以及多种控制功能。

HDLC基本配置


```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol hdlc
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address 10.0.1.1 30
```


- 用户只需要在串行接口视图下运行**link-protocol** hdlc命令就可以使能接口的HDLC协议。华为设备上的串行接口默认运行PPP协议。用户必须在串行链路两端的端口上配置相同的链路协议，双方才能通信。

HDLC接口地址借用




```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol hdlc
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address unnumbered interface loopBack 0
[RTA]ip route-static 10.1.1.0 24 Serial 1/0/0
```

- 串行接口可以借用Loopback接口的IP地址和对端建立连接。

  HUAWEI

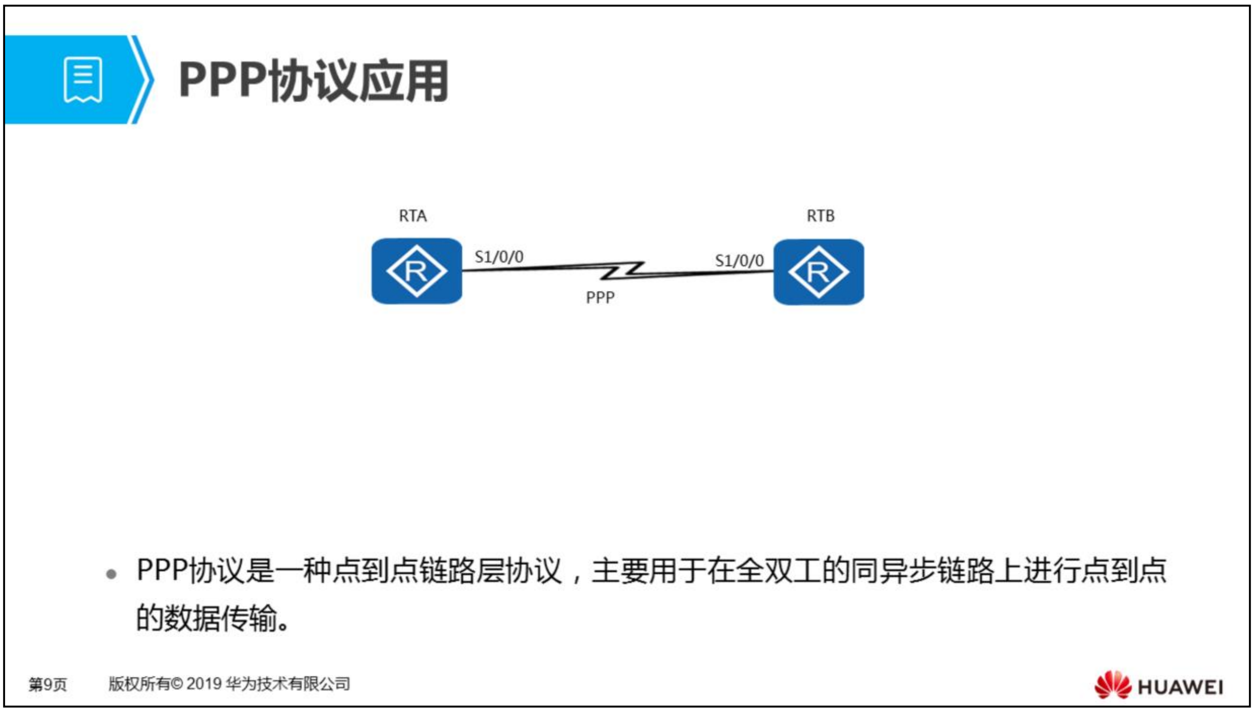
- 一个接口如果没有IP地址就无法生成路由，也就无法转发报文。IP地址借用允许一个没有IP地址的接口从其它接口借用IP地址。这样可以避免一个接口独占IP地址，节省IP地址资源。一般建议借用loopback接口的IP地址，因为这类接口总是处于活跃（active）状态，因而能提供稳定可用的IP地址。
- 本例中，在RTA的S1/0/0接口配置完接口地址借用之后，还需要在RTA上配置静态路由，以使RTA能够转发数据到10.1.1.0/24网络。

配置验证

```
[RTA]display ip interface brief
*down: administratively down ^down: standby (l): loopback
(s): spoofing
……
Interface              IP Address/Mask    Physical   Protocol
LoopBack0              10.1.1.1/32          up         up(s)
Serial1/0/0            10.1.1.1/32          up         up
Serial1/0/1            unassigned           up         down
```

  HUAWEI

- 执行**display ip interface brief**命令可以查看路由器接口简要信息。如果有IP地址被借用，该IP地址会显示在多个接口上，说明借用loopback接口的IP地址成功。



- PPP协议是一种点到点链路层协议，主要用于在全双工的同异步链路上进行点到点的数据传输。PPP协议有如下优点：

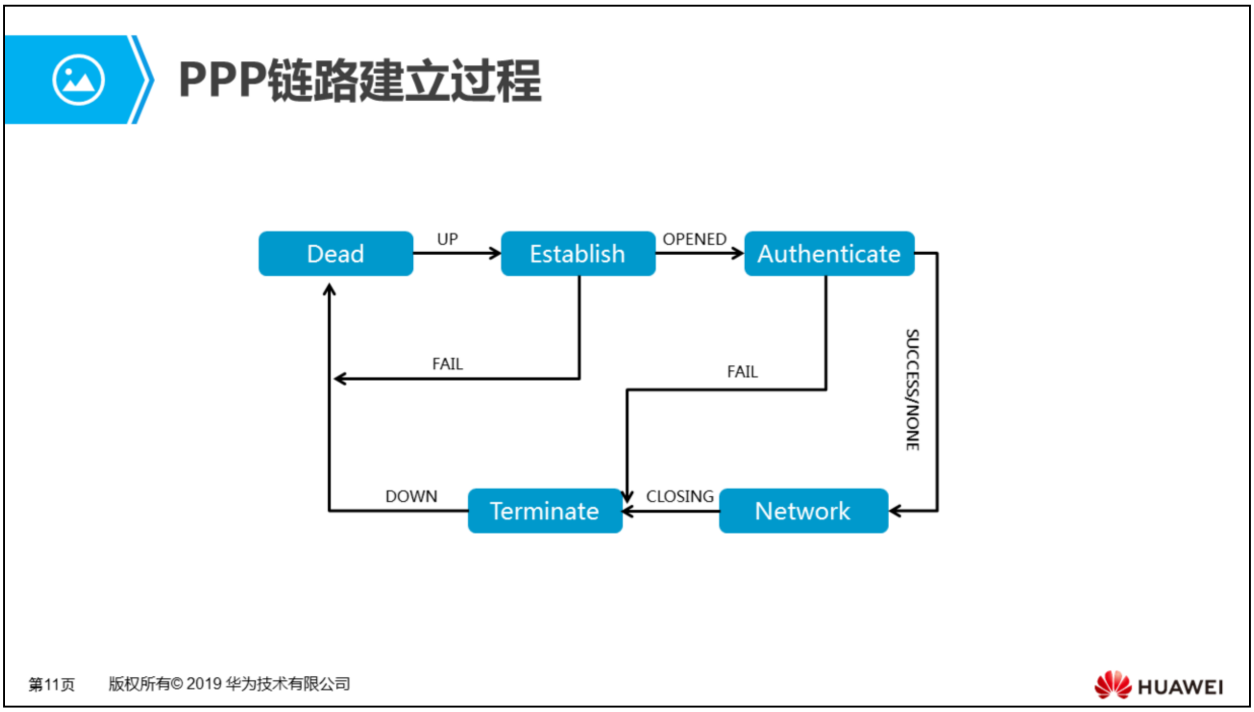
1. PPP既支持同步传输又支持异步传输，而X.25、FR（Frame Relay）等数据链路层协议仅支持同步传输，SLIP仅支持异步传输。
2. PPP协议具有很好的扩展性，例如，当需要在以太网链路上承载PPP协议时，PPP可以扩展为PPPoE。
3. PPP提供了LCP（Link Control Protocol）协议，用于各种链路层参数的协商。
4. PPP提供了各种NCP（Network Control Protocol）协议（如IPCP、IPXCP），用于各网络层参数的协商，更好地支持了网络层协议。
5. PPP提供了认证协议：CHAP（Challenge-Handshake Authentication Protocol）、PAP（Password Authentication Protocol），更好的保证了网络的安全性。
6. 无重传机制，网络开销小，速度快。

PPP组件

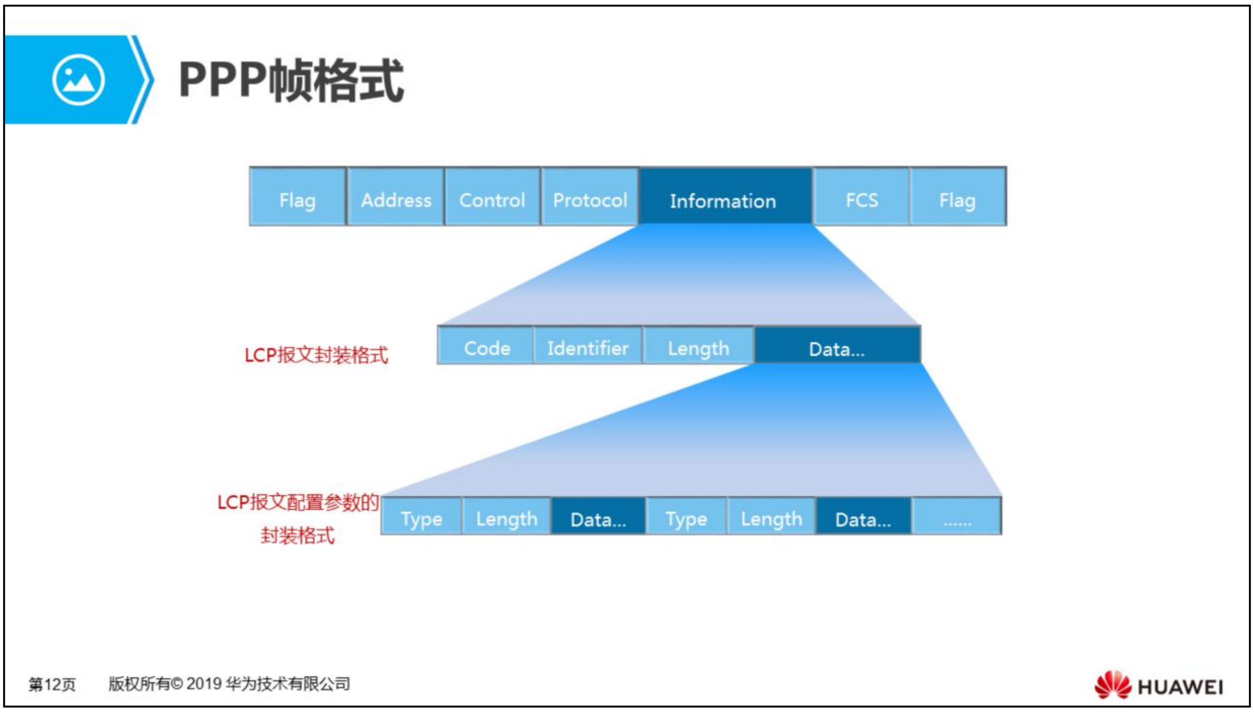
名称	作用
链路控制协议 Link Control Protocol	用来建立、拆除和监控PPP数据链路
网络层控制协议 Network Control Protocol	用于对不同的网络层协议进行连接建立和参数协商

  HUAWEI

- PPP包含两个组件：链路控制协议LCP和网络层控制协议NCP。
- 为了能适应多种多样的链路类型，PPP定义了链路控制协议LCP。LCP可以自动检测链路环境，如是否存在环路；协商链路参数，如最大数据包长度，使用何种认证协议等等。与其他数据链路层协议相比，PPP协议的一个重要特点是可以提供认证功能，链路两端可以协商使用何种认证协议来实施认证过程，只有认证成功之后才会建立连接。
- PPP定义了一组网络层控制协议NCP，每一个NCP对应了一种网络层协议，用于协商网络层地址等参数，例如IPCP用于协商控制IP协议，IPXCP用于协商控制IPX协议等。



- 对于PPP链路建立过程的描述如下：
- Dead阶段也称为物理层不可用阶段。当通信双方的两端检测到物理线路激活时，就会从Dead阶段迁移至Establish阶段，即链路建立阶段。
- 在Establish阶段，PPP链路进行LCP参数协商。协商内容包括最大接收单元MRU、认证方式、魔术字（Magic Number）等选项。LCP参数协商成功后会进入Opened状态，表示底层链路已经建立。
- 多数情况下,链路两端的设备是需要经过认证阶段（Authenticate）后才能够进入到网络层协议阶段。PPP链路在缺省情况下是不要求进行认证的。如果要求认证，则在链路建立阶段必须指定认证协议。认证方式是在链路建立阶段双方进行协商的。如果在这个阶段再次收到了Configure-Request报文，则又会返回到链路建立阶段。
- 在Network阶段，PPP链路进行NCP协商。通过NCP协商来选择和配置一个网络层协议并进行网络层参数协商。只有相应的网络层协议协商成功后，该网络层协议才可以通过这条PPP链路发送报文。如果在这个阶段收到了Configure-Request报文，也会返回到链路建立阶段。
- NCP协商成功后，PPP链路将保持通信状态。PPP运行过程中，可以随时中断连接，例如物理链路断开、认证失败、超时定时器时间、管理员通过配置关闭连接等动作都可能导致链路进入Terminate阶段。
- 在Terminate阶段，如果所有的资源都被释放，通信双方将回到Dead阶段，直到通信双方重新建立PPP连接。



- PPP采用了与HDLC协议类似的帧格式：
- Flag域标识一个物理帧的起始和结束，该字节为二进制序列01111110（0X7E）。
- PPP帧的地址域跟HDLC帧的地址域有差异，PPP帧的地址域字节固定为11111111（0XFF），是一个广播地址。
- PPP数据帧的控制域默认为00000011(0X03)，表明为无序号帧。
- 帧校验序列（FCS）是个16位的校验和，用于检查PPP帧的完整性。
- 协议字段用来说明PPP所封装的协议报文类型，典型的字段值有：0XC021代表LCP报文，0XC023代表PAP报文，0XC223代表CHAP报文。
- 信息字段包含协议字段中指定协议的数据包。数据字段的默认最大长度（不包括协议字段）称为最大接收单元MRU（Maximum Receive Unit），MRU的缺省值为1500字节。
- 如果协议字段被设为0XC021，则说明通信双方正通过LCP报文进行PPP链路的协商和建立：
- Code字段，主要是用来标识LCP数据报文的类型。典型的报文类型有：配置信息报文（Configure Packets: 0x01)，配置成功信息报文(Configure-Ack: 0X02)，终止请求报文(Terminate-Request：0X05)。
- Identifier域为1个字节，用来匹配请求和响应。
- Length域的值就是该LCP报文的总字节数据。
- 数据字段则承载各种TLV（Type/Length/Value）参数用于协商配置选项，包括最大接收单元，认证协议等等。

LCP报文

报文类型	作用
Configure-Request	包含发送者试图与对端建立连接时使用的参数列表
Configure-Ack	表示完全接受对端发送的Configure-Request的参数取值
Configure-Nak	表示对端发送的Configure-Request中的某些参数取值在本端不被认可
Configure-Reject	表示对端发送的Configure-Request中的某些参数本端不能识别

- 此表格列出了LCP用于链路层参数协商所使用四种报文类型。

1. Configure-Request（配置请求）：链路层协商过程中发送的第一个报文，该报文表明点对点双方开始进行链路层参数的协商。
2. Configure-Ack（配置响应）：收到对端发来的Configure-Request报文，如果参数取值完全接受，则以此报文响应。
3. Configure-Nak（配置不响应）：收到对端发来的Configure-Request报文，如果参数取值不被本端认可，则发送此报文并且携带本端可接受的配置参数。
4. Configure-Reject（配置拒绝）：收到对端发来的Configure-Request报文，如果本端不能识别对端发送的Configure-Request中的某些参数，则发送此报文并且携带那些本端不能认别的配置参数。

LCP协商参数

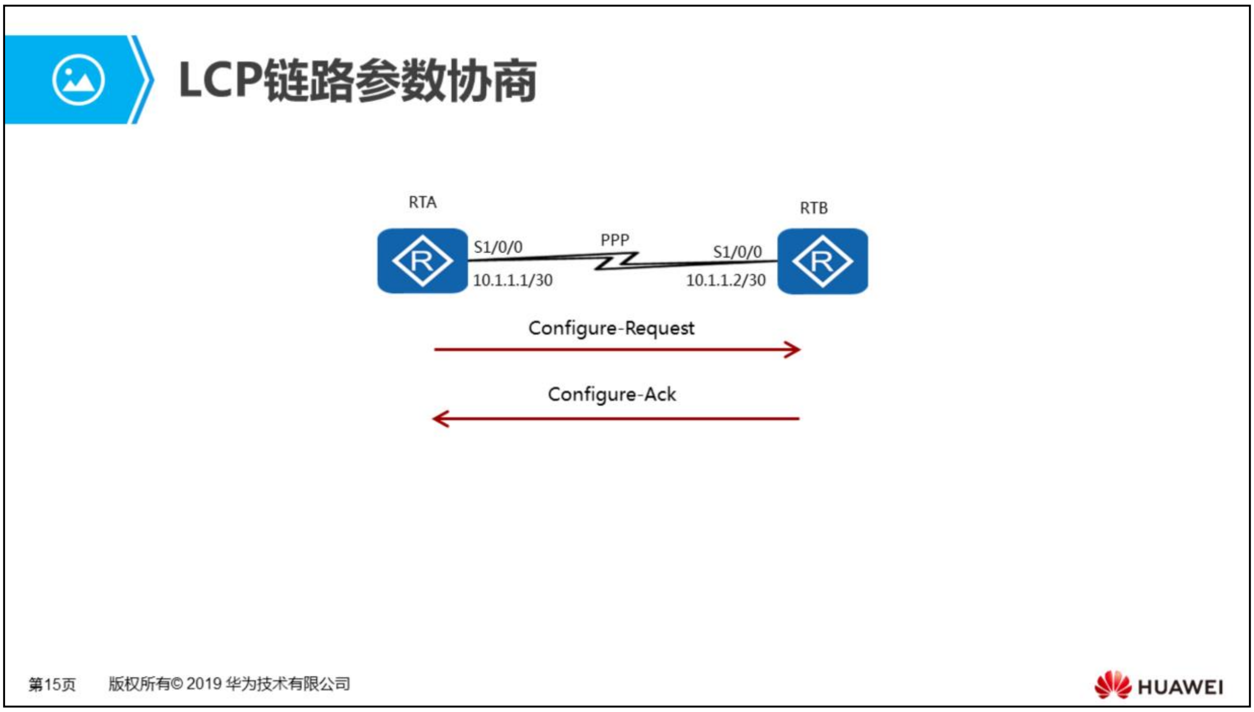
参数	作用	缺省值
最大接收单元MRU	PPP数据帧中Information字段和Padding字段的总长度	1500字节
认证协议	认证对端使用的认证协议	不认证
魔术字	魔术字为一个随机产生的数字，用于检测链路环路，如果收到的LCP报文中的魔术字和本端产生的魔术字相同，则认为链路有环路	启用

  HUAWEI

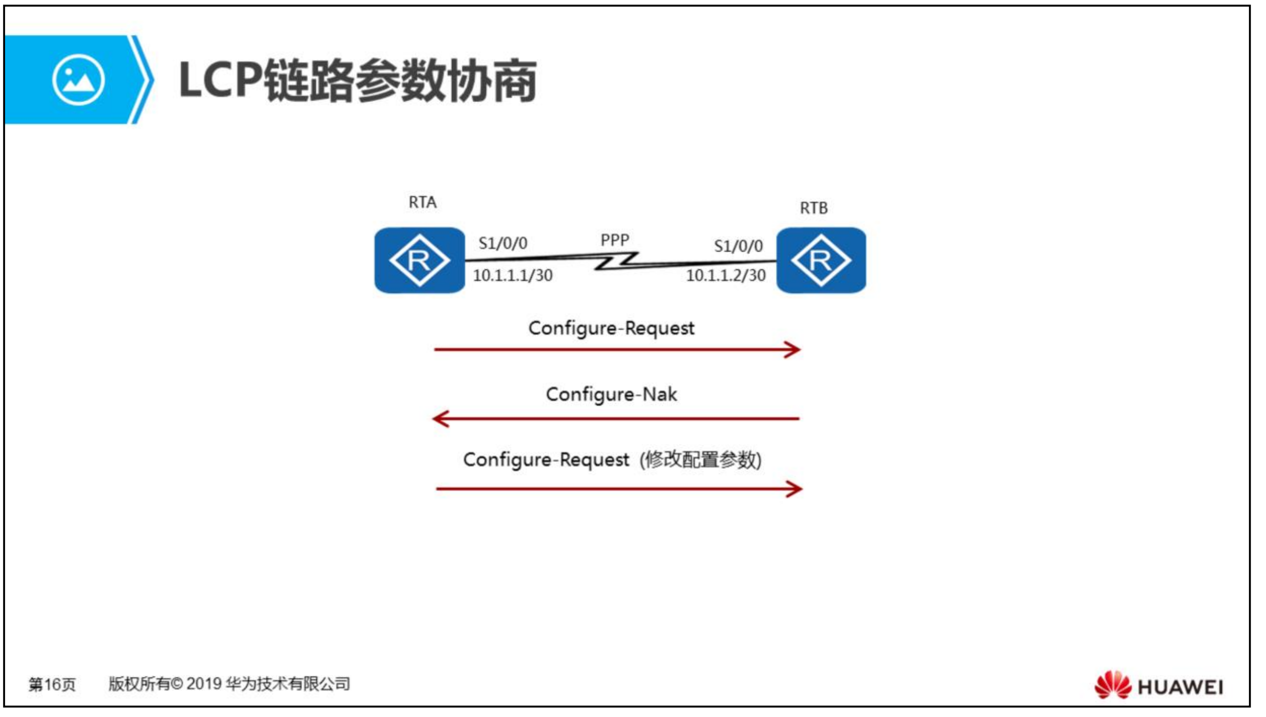
- LCP报文携带的一些常见的配置参数有MRU，认证协议，以及魔术字。

1. 在VRP平台上，MRU参数使用接口上配置的最大传输单元（MTU）值来表示。
2. 常用的PPP认证协议有PAP和CHAP，一条PPP链路的两端可以使用不同的认证协议认证对端，但是被认证方必须支持认证方要求使用的认证协议并正确配置用户名和密码等认证信息。
3. LCP使用魔术字来检测链路环路和其它异常情况。魔术字为随机产生的一个数字，随机机制需要保证两端产生相同魔术字的可能性几乎为0。

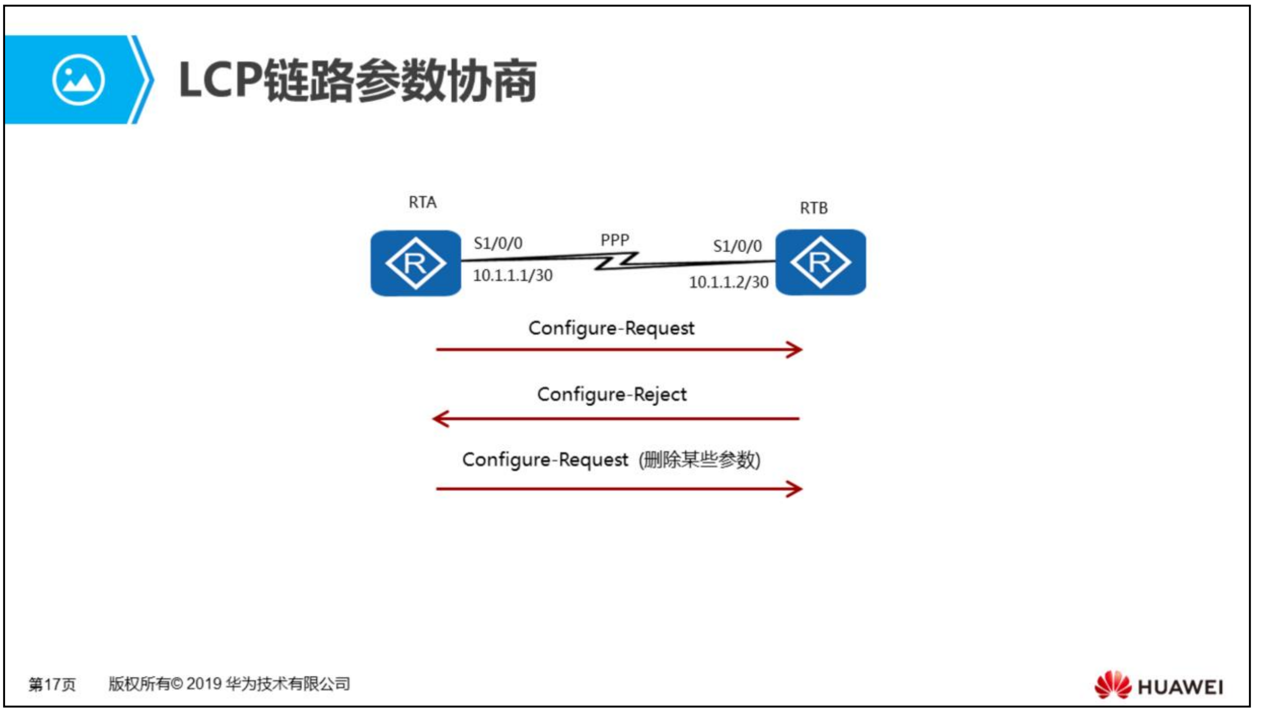
- 收到一个Configure-Request报文之后，其包含的魔术字需要和本地产生的魔术字做比较，如果不同，表示链路无环路，则使用Confugure-Ack报文确认（其它参数也协商成功），表示魔术字协商成功。在后续发送的报文中，如果报文含有魔术字字段，则该字段设置为协商成功的魔术字。



- 如图所示，RTA和RTB使用串行链路相连，运行PPP。当物理层链路变为可用状态之后，RTA和RTB使用LCP协商链路参数。本例中，RTA首先发送一个Configure-Request报文，此报文中包含RTA上配置的链路层参数。当RTB收到此Configure-Request报文之后，如果RTB能识别并接受此报文中的所有链路层参数，则向RTA回应一个Configure-Ack报文。
- RTA在没有收到Configure-Ack报文的情况下，会每隔3秒重传一次Configure-Request报文，如果连续10次发送Configure-Request报文仍然没有收到Configure-Ack报文，则认为对端不可用，停止发送Configure-Request报文。
- 注：完成上述过程只是表明RTB认为RTA上的链路参数配置是可接受的。RTB也需要向RTA发送Configure-Request报文，使RTA检测RTB上的链路参数是不是可接受的。

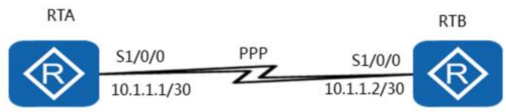


- 当RTB收到RTA发送的Configure-Request报文之后，如果RTB能识别此报文中携带的所有链路层参数，但是认为部分或全部参数的取值不能接受，即参数的取值协商不成功，则RTB需要向RTA回应一个Configure-Nak报文。
- 在这个Configure-Nak报文中，只包含不能接受的链路层参数，并且此报文所包含的链路层参数均被修改为RTB上可以接受的取值（或取值范围）。
- 在收到Configure-Nak报文之后，RTA需要根据此报文中的链路层参数重新选择本地配置的其它参数，并重新发送一个Configure-Request。



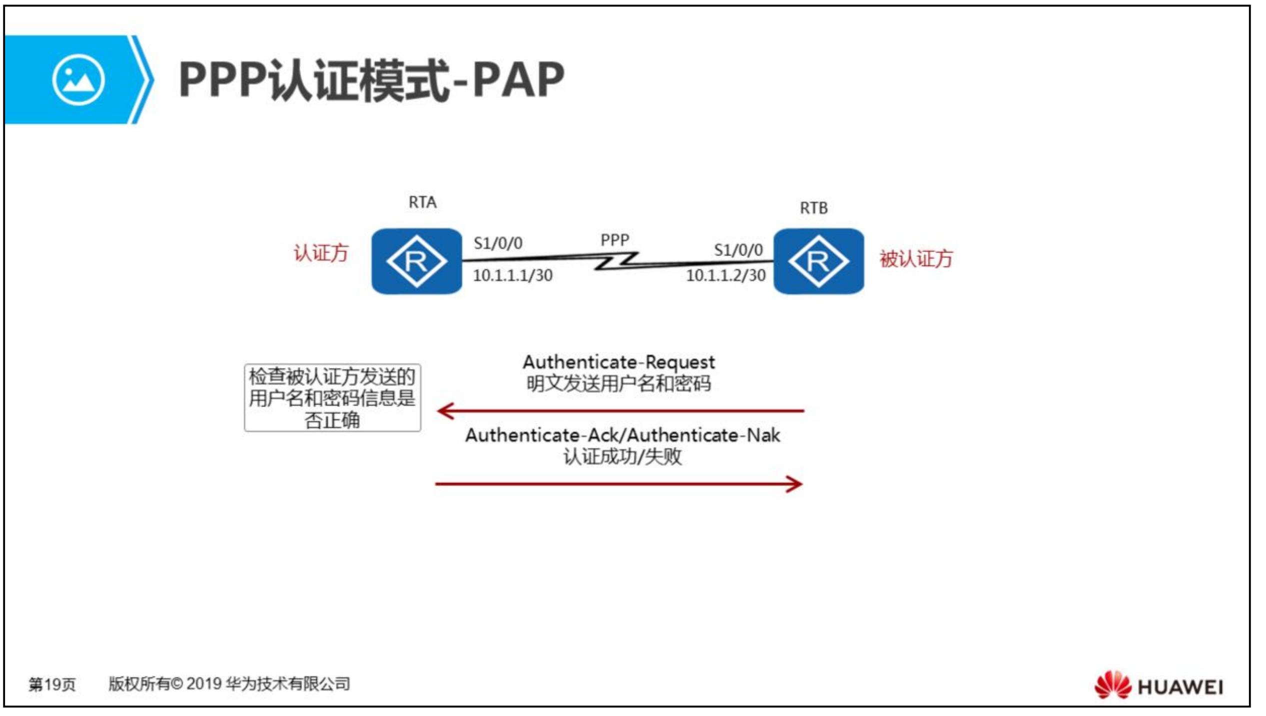
- 当RTB收到RTA发送的Configure-Request报文之后，如果RTB不能识别此报文中携带的部分或全部链路层参数，则RTB需要向RTA回应一个Configure-Reject报文。在此Configure-Reject报文中，只包含不能被识别的链路层参数。
- 在收到Configure-Reject报文之后，RTA需要向RTB重新发送一个Configure-Request报文，在新的Configure-Request报文中，不再包含不被对端（RTB）识别的参数。

PPP基本配置

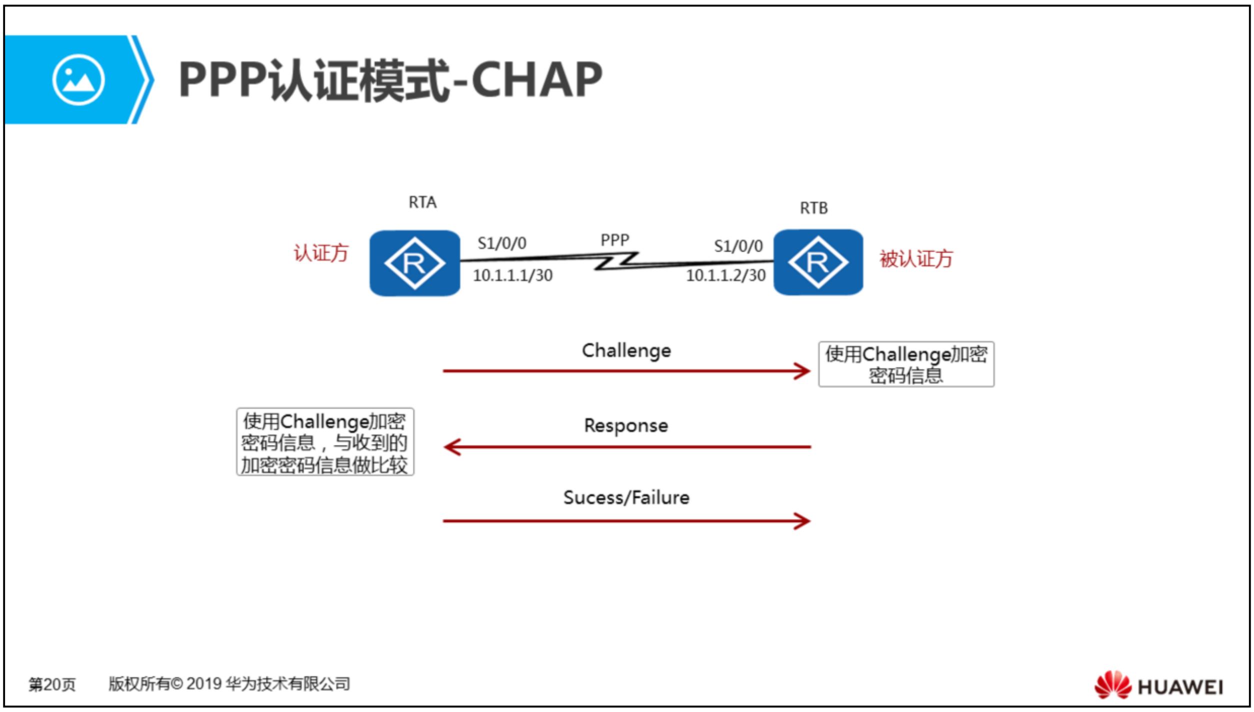



```
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol ppp
Warning: The encapsulation protocol of the link will be changed.
Continue? [Y/N]:y
[RTA-Serial1/0/0]ip address 10.1.1.1 30
```


- 建立PPP链路之前，必须先在串行接口上配置链路层协议。华为ARG3系列路由器默认在串行接口上使能PPP协议。如果接口运行的不是PPP协议，需要运行**link-protocol ppp**命令来使能数据链路层的PPP协议。



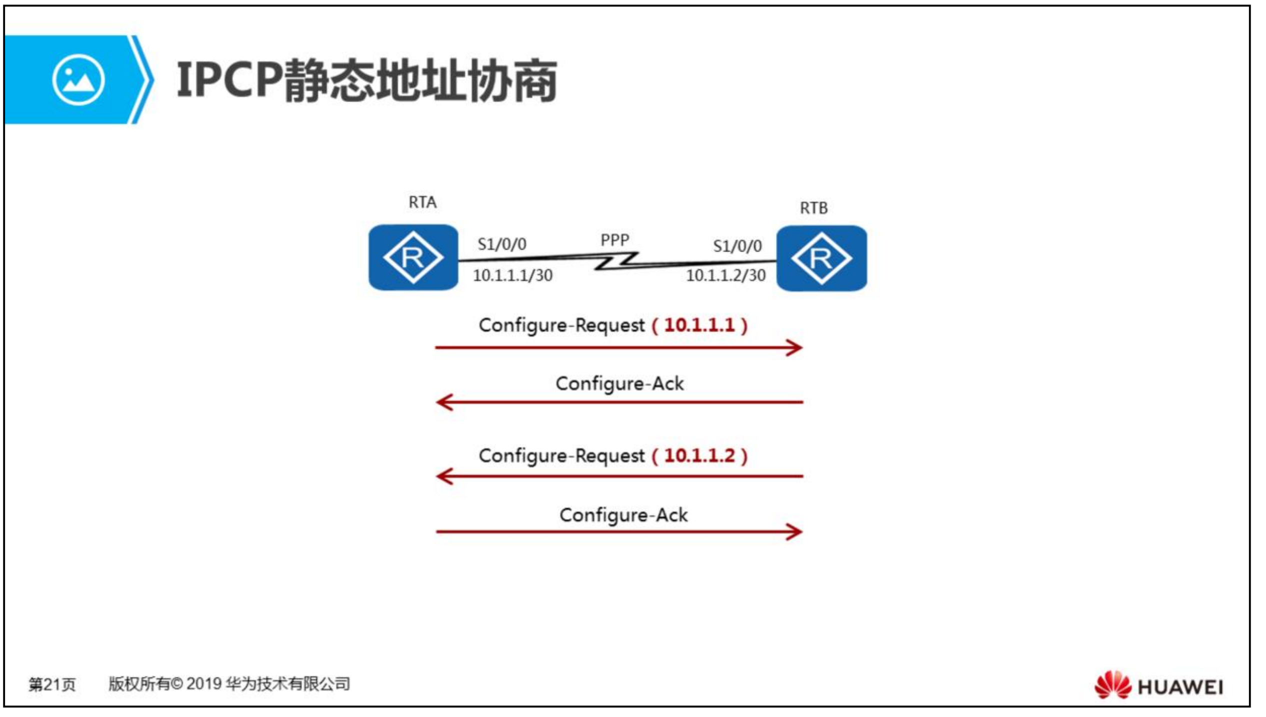
- PAP认证的工作原理较为简单。PAP认证协议为两次握手认证协议，密码以明文方式在链路上发送。
- LCP协商完成后，认证方要求被认证方使用PAP进行认证。
- 被认证方将配置的用户名和密码信息使用Authenticate-Request报文以明文方式发送给认证方。
- 认证方收到被认证方发送的用户名和密码信息之后，根据本地配置的用户名和密码数据库检查用户名和密码信息是否匹配，如果匹配，则返回Authenticate-Ack报文，表示认证成功。否则，返回Authenticate-Nak报文，表示认证失败。



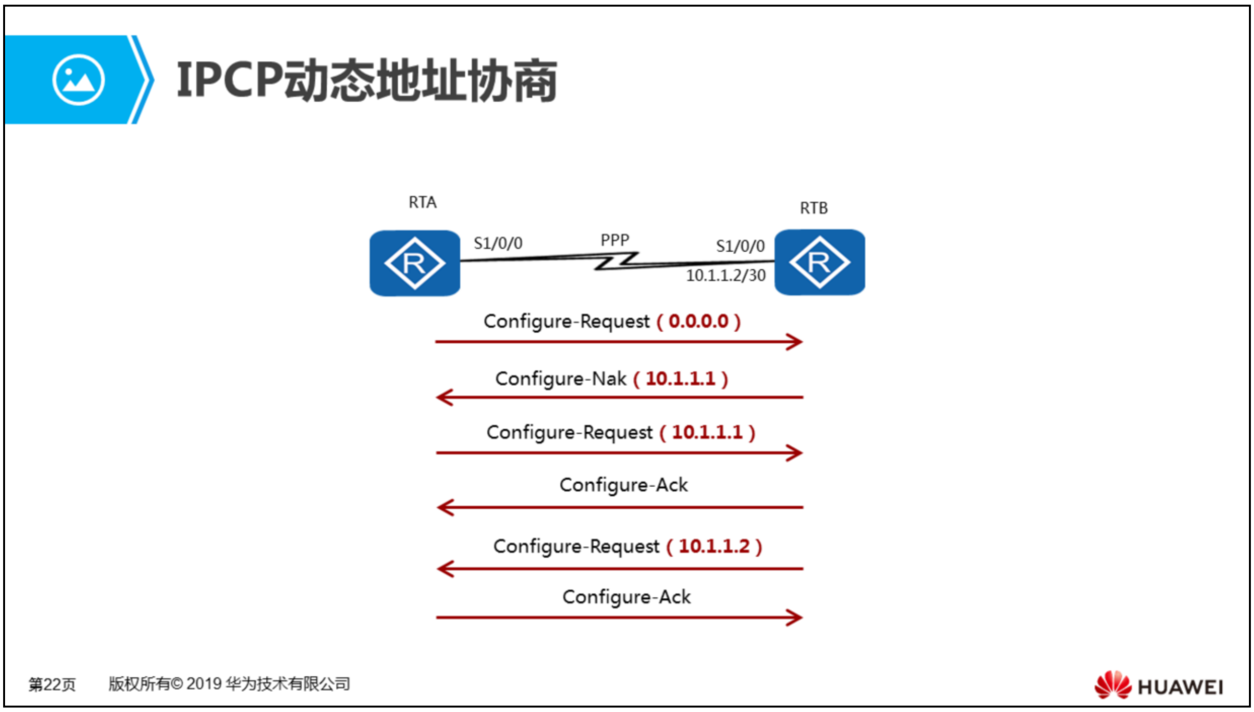
- CHAP认证过程需要三次报文的交互。为了匹配请求报文和回应报文，报文中含有Identifier字段，一次认证过程所使用的报文均使用相同的Identifier信息。

1. LCP协商完成后，认证方发送一个Challenge报文给被认证方，报文中含有Identifier信息和一个随机产生的Challenge字符串，此Identifier即为后续报文所使用的Identifier。
2. 被认证方收到此Challenge报文之后，进行一次加密运算，运算公式为MD5{ Identifier＋密码＋Challenge }，意思是将Identifier、密码和Challenge三部分连成一个字符串，然后对此字符串做MD5运算，得到一个16字节长的摘要信息，然后将此摘要信息和端口上配置的CHAP用户名一起封装在Response报文中发回认证方。
3. 认证方接收到被认证方发送的Response报文之后，按照其中的用户名在本地查找相应的密码信息，得到密码信息之后，进行一次加密运算，运算方式和被认证方的加密运算方式相同，然后将加密运算得到的摘要信息和Response报文中封装的摘要信息做比较，相同则认证成功，不相同则认证失败。

- 使用CHAP认证方式时，被认证方的密码是被加密后才进行传输的，这样就极大的提高了安全性。



- IPCP使用和LCP相同的协商机制、报文类型，但IPCP并非调用LCP，只是工作过程、报文等和LCP相同。
- IP地址协商包括两种方式：静态配置协商和动态配置协商。
- 如图所示，两端路由器配置的IP地址分别为10.1.1.1/30和10.1.1.2/30。
- 静态IP地址的协商过程如下：
- 1. 每一端都要发送Configure-Request报文，在此报文中包含本地配置的IP地址；
- 2. 每一端接收到此Configure-Request报文之后，检查其中的IP地址，如果IP地址是一个合法的单播IP地址，而且和本地配置的IP地址不同（没有IP冲突），则认为对端可以使用该地址，回应一个Configure-Ack报文。

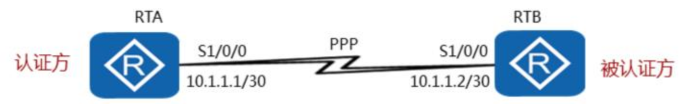


- 两端动态协商IP地址的过程如下：

1. RTA向RTB发送一个Configure-Request报文，此报文中会包含一个IP地址0.0.0.0，表示向对端请求IP地址；
2. RTB收到上述Configure-Request报文后，认为其中包含的地址（0.0.0.0）不合法，使用Configure-Nak回应一个新的IP地址10.1.1.1；
3. RTA收到此Configure-Nak报文之后，更新本地IP地址，并重新发送一个Configure-Request报文，包含新的IP地址10.1.1.1；
4. RTB收到Configure-Request报文后，认为其中包含的IP地址为合法地址，回应一个Configure-Ack报文。

- 同时，RTB也要向RTA发送Configure-Request报文请求使用地址10.1.1.2，RTA认为此地址合法，回应Configure-Ack报文。

PAP认证




```
[RTA]aaa
[RTA-aaa]local-user huawei password cipher huawei123
[RTA-aaa]local-user huawei service-type ppp
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol ppp
[RTA-Serial1/0/0]ppp authentication-mode pap
[RTA-Serial1/0/0]ip address 10.1.1.1 30
```

```
[RTB]interface Serial 1/0/0
[RTB-Serial1/0/0]link-protocol ppp
[RTB-Serial1/0/0]ppp pap local-user huawei password cipher huawei123
[RTB-Serial1/0/0]ip address 10.1.1.2 30
```

  HUAWEI

- **local-user huawei password cipher huawei123**命令用于创建一个本地用户，用户名为“huawei”，密码为“huawei123”，关键字“cipher”表示密码信息在配置文件中被加密。
- **local-user huawei service-type ppp**命令用于设置用户“huawei”为PPP用户。
- **ppp authentication-mode pap**命令用于在认证方开启PAP认证的功能，即要求对端使用PAP认证。
- **ppp pap local-user huawei password cipher huawei123**命令用于在被认证方配置PAP使用的用户名和密码信息。

配置验证

```
<RTB>debugging ppp pap all
Mar 20 2016 04:50:24.280.4+00:00 RTB PPP/7/debug2:
  PPP State Change:
      Serial1/0/0 PAP : Initial --> SendRequest
Mar 20 2016 04:50:24.290.3+00:00 RTB PPP/7/debug2:
  PPP State Change:
      Serial1/0/0 PAP : SendRequest --> ClientSuccess
……
```

配置CHAP认证模式


```
[RTA]aaa
[RTA-aaa]local-user huawei password cipher huawei123
[RTA-aaa]local-user huawei service-type ppp
[RTA]interface Serial 1/0/0
[RTA-Serial1/0/0]link-protocol ppp
[RTA-Serial1/0/0]ppp authentication-mode chap
```

```
[RTB]interface Serial 1/0/0
[RTB-Serial1/0/0]link-protocol ppp
[RTB-Serial1/0/0]ppp chap user huawei
[RTB-Serial1/0/0]ppp chap password cipher huawei123
```


- **local-user huawei password cipher huawei123**命令用于创建一个本地用户，用户名为“huawei”，密码为“huawei123”；关键字“cipher”表示密码信息在配置文件中加密保存。
- **local-user huawei service-type ppp**命令用于设置用户“huawei”为PPP用户。
- **ppp authentication-mode chap**命令用于在认证方开启CHAP认证的功能，即要求对端使用CHAP认证。
- **ppp chap user huawei**命令用于在被认证方设置CHAP使用的用户名为“huawei”。
- **ppp chap password cipher huawei123**命令用于在被认证方设置CHAP使用的密码为“huawei123”。

配置验证

```
<RTB>debugging ppp chap all
Mar 20 2016 05:15:54.230.1+00:00 RTB PPP/7/debug2:
PPP State Change:
      Serial1/0/0 CHAP : Initial --> ListenChallenge
Mar 20 2016 05:15:54.230.7+00:00 RTB PPP/7/debug2:
  PPP State Change:
      Serial1/0/0 CHAP : ListenChallenge --> SendResponse
Mar 20 2016 05:15:54.250.3+00:00 RTB PPP/7/debug2:
  PPP State Change:
      Serial1/0/0 CHAP : SendResponse --> ClientSuccess
……
```


本章总结

- 发送端在发送Configure-Request之后，收到哪个消息才能表示PPP链路建立成功？
- CHAP认证方式需要交互几次报文？

  HUAWEI

- 如果使用PPP作为链路层封装协议，需要建立PPP链路的两端设备都必须发送Configure-Request报文，当每个设备均已收到对端发来的Configure-Ack报文后，就表示链路的建立过程已成功完成。
- CHAP认证协议为三次握手认证协议，需要交互三次报文来认证对方身份。


谢 谢
www.huawei.com

PPPoE原理和配置

HUAWEI

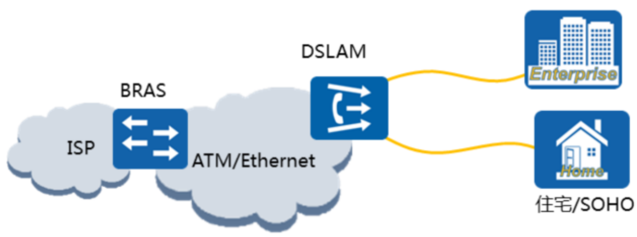
前言

- 数字用户线路DSL(Digital Subscriber Line)是以电话线为传输介质的传输技术，人们通常把所有的DSL技术统称为xDSL，x代表不同种类的数字用户线路技术。目前比较流行的宽带接入方式为ADSL，ADSL是非对称DSL技术，使用的是PPPoE（PPP over Ethernet）协议。
- PPPoE协议通过在以太网上提供点到点的连接，建立PPP会话，使得以太网中的主机能够连接到远端的宽带接入服务器上。PPPoE具有适用范围广、安全性高、计费方便等特点。

目标

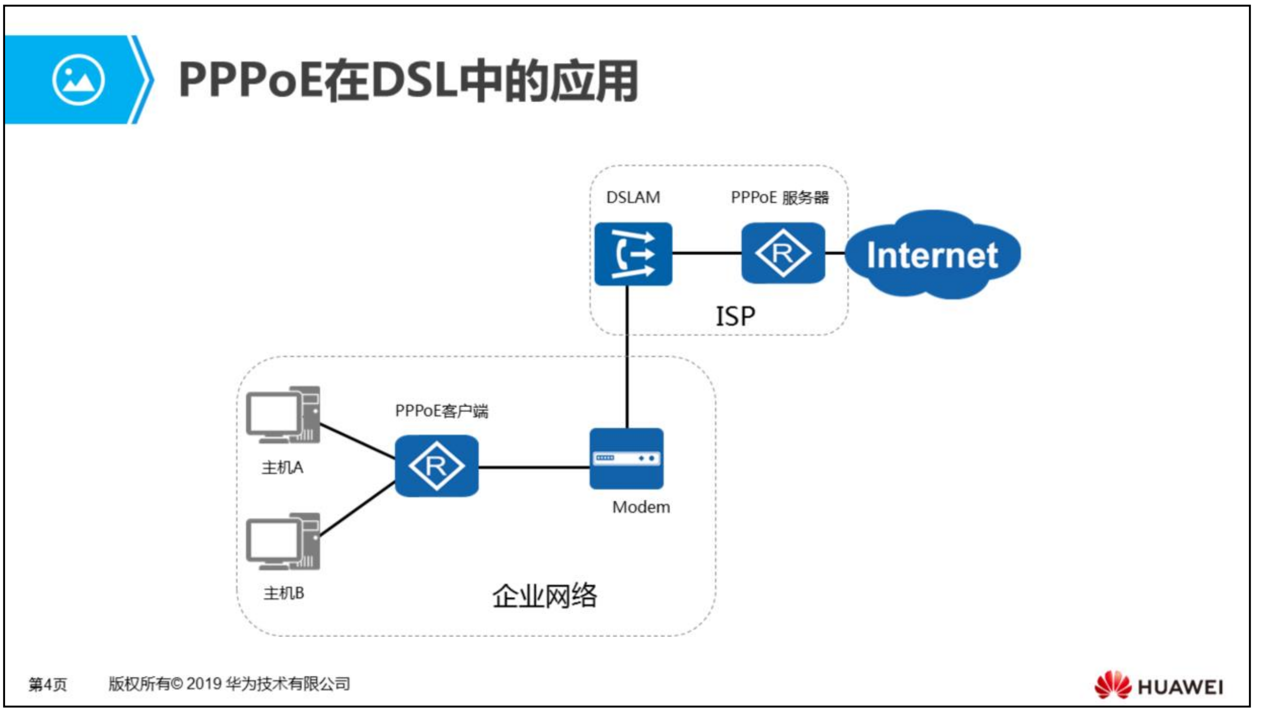
- 学完本课程后，您将能够：
  - 掌握PPPoE连接协商过程
  - 掌握PPPoE的配置

DSL应用场景



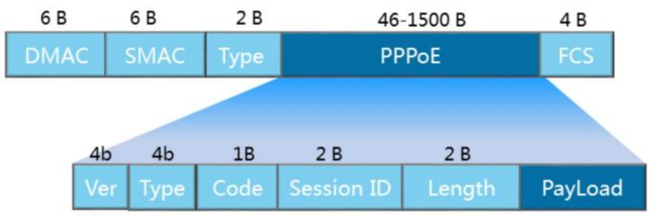
- 数字用户线路DSL是以电话线为传输介质的传输技术。

- DSL是一种利用现有电话网络实现数据通信的宽带技术。在使用DSL接入网络时，用户侧会安装调制解调器，然后通过现有的电话线与数字用户线路接入复用器（DSLAM）相连。DSLAM是各种DSL系统的局端设备，属于最后一公里接入设备。
- 然后，DSLAM通过高速ATM网络或者以太网将用户的数据流量转发给宽带远程接入服务器（BRAS）。BRAS是面向宽带网络应用的接入网关，位于骨干网的边缘层。



- 运营商希望通过同一台接入设备来连接远程的多个主机，同时接入设备能够提供访问控制和计费功能。在众多的接入技术中，把多个主机连接到接入设备的最经济的方法就是以太网，而PPP协议可以提供良好的访问控制和计费功能，于是产生了在以太网上传输PPP报文的技术，即PPPoE。
- PPPoE利用以太网将大量主机组成网络，通过一个远端接入设备连入因特网，并运用PPP协议对接入的每个主机进行控制，具有适用范围广、安全性高、计费方便的特点。
- PPPoE技术解决了用户上网收费等实际应用问题，得到了宽带接入运营商的认可并被广泛应用。

PPPoE报文



- PPPoE报文是使用Ethernet格式来进行封装的。

  HUAWEI

- PPPoE报文是使用Ethernet格式进行封装的，Ethernet中各字段解释如下：

1. DMAC：表示目的设备的MAC地址，通常为以太网单播目的地址或者以太网广播地址（0xFFFFFFFF）。
2. SMAC：表示源设备的以太网MAC地址。
3. Type：表示协议类型字段，当值为0x8863时表示承载的是PPPoE发现阶段的报文。当值为0x8864时表示承载的是PPPoE会话阶段的报文。

- PPPoE字段中的各个字段解释如下：

1. VER：表示PPPoE版本号，值为0x01。
2. Type：表示类型，值为0x01。
3. Code：表示PPPoE报文类型，不同取值标识不同的PPPoE报文类型。
4. PPPoE会话ID，与以太网SMAC和DMAC一起定义了一个PPPoE会话。
5. Length：表示PPPoE报文的Payload长度，不包括以太网头部和PPPoE头部的长度。

PPPoE会话建立过程

阶段	描述
发现阶段	获取对方以太网地址，以及确定唯一的PPPoE会话。
会话阶段	包含两部分：PPP协商阶段和PPP报文传输阶段。
会话终结阶段	会话建立以后的任意时刻，发送报文结束PPPoE会话。

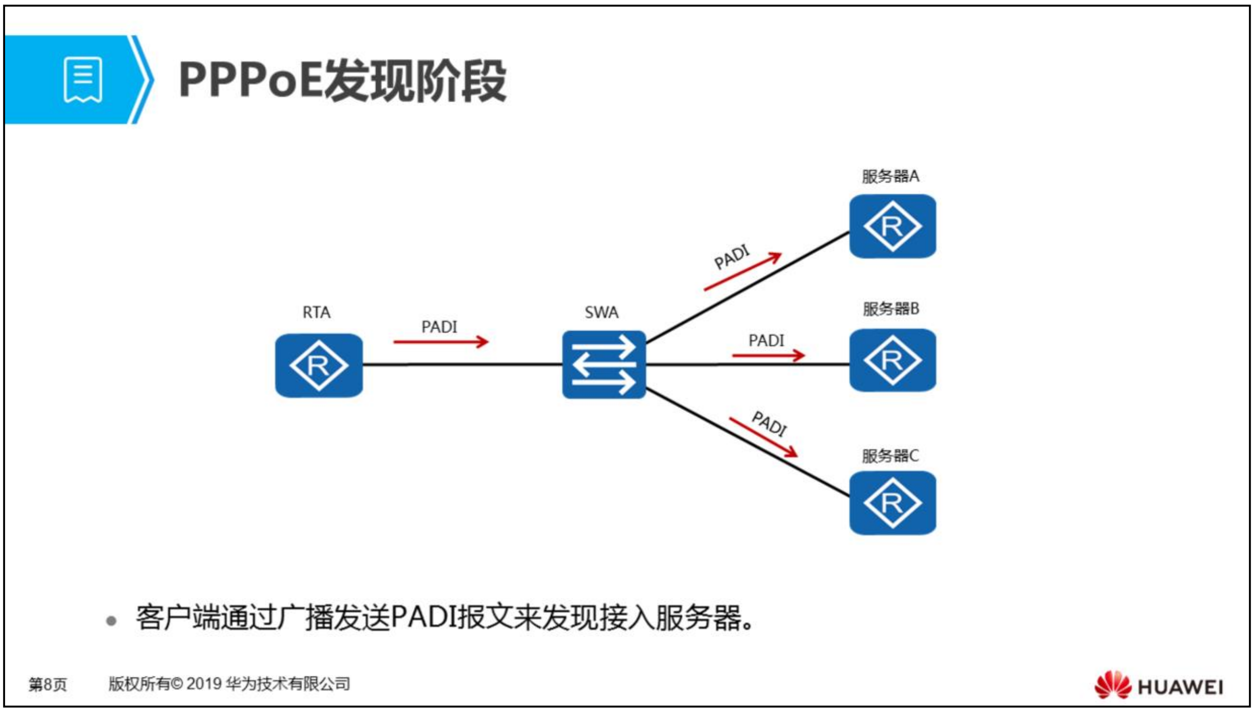
- PPPoE可分为三个阶段，即发现阶段、会话阶段和会话终结阶段。

PPPoE协议报文

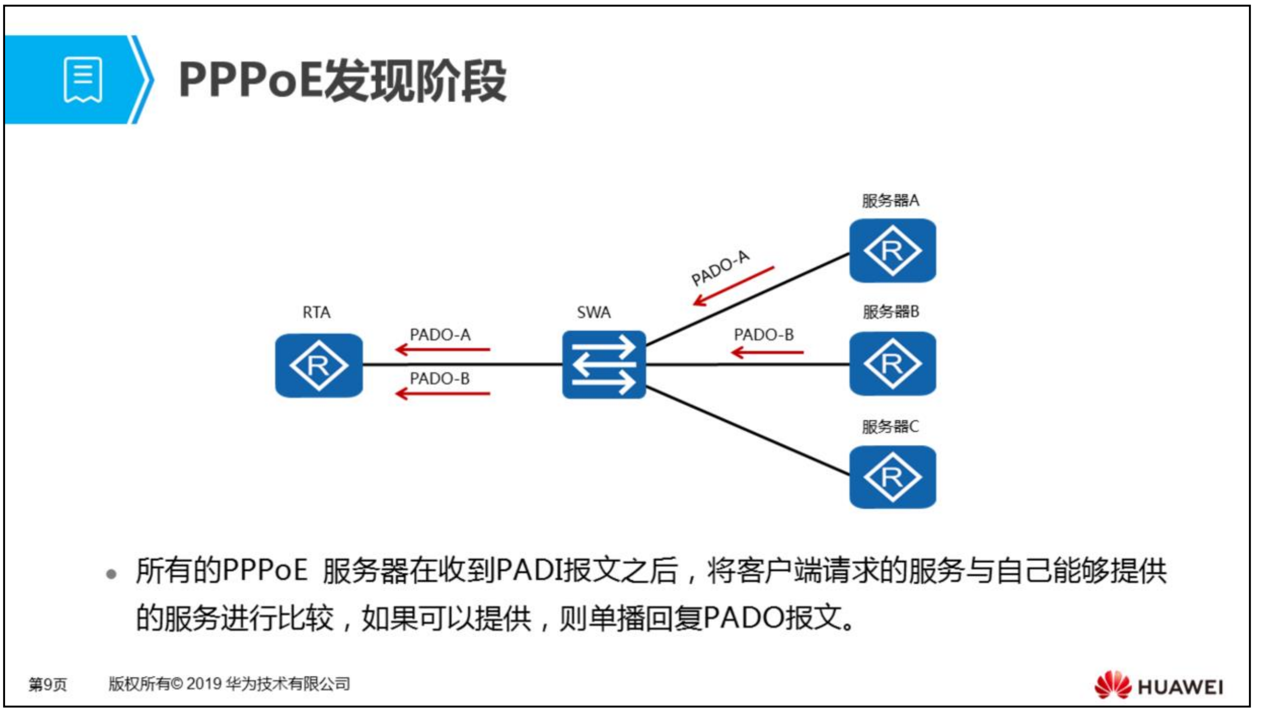
类型	描述
PADI	PPPoE发现初始报文
PADO	PPPoE发现提供报文
PADR	PPPoE发现请求报文
PADS	PPPoE发现会话确认报文
PADT	PPPoE发现终止报文

- PPPoE通过这五种类型的报文来建立和终结PPPoE会话。

1. PADI（PPPoE Active Discovery Initiation）报文：用户主机发起的PPPoE服务器探测报文，目的MAC地址为广播地址。
2. PADO（PPPoE Active Discovery Offer）报文：PPPoE服务器收到PADI报文之后的回应报文，目的MAC地址为客户端主机的MAC地址。
3. PADR（PPPoE Active Discovery Request）报文：用户主机收到PPPoE服务器回应的PADO报文后，单播发起的请求报文，目的地址为此用户选定的那个PPPoE服务器的MAC地址。
4. PADS（PPPoE Active Discovery Session Configuration）报文：PPPoE服务器分配一个唯一的会话进程ID，并通过PADS报文发送给主机。
5. PADT（PPPoE Active Discovery Terminate）报文：当用户或者服务器需要终止会话时，可以发送这种PADT报文。

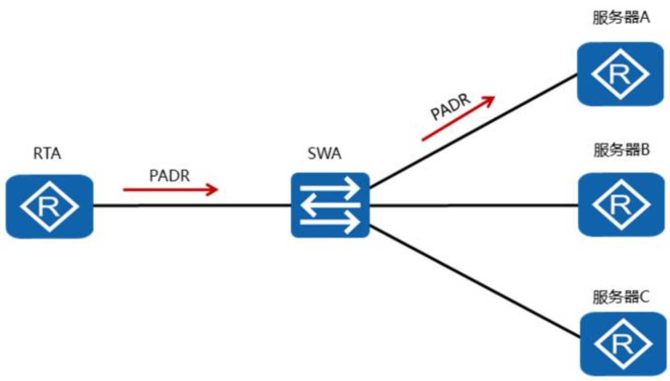


- 在发现阶段，PPPoE客户端在本地以太网中广播一个PADI报文，此PADI报文中包含了客户端需要的服务信息。在PADI报文中，目的MAC地址是一个广播地址，Code字段为0x09，Session ID字段为0x0000。所有PPPoE服务器收到PADI报文之后，会将报文中所请求的服务与自己能够提供的服务进行比较。



- 如果服务器可以提供客户端请求的服务，就会回复一个PADO报文。客户端（RTA）可能会收到多个PPPoE服务器发送的PADO报文。在PADO报文中，目的地址是发送PADI报文的客户端MAC地址，Code字段为0x07，Session ID字段为0x0000。

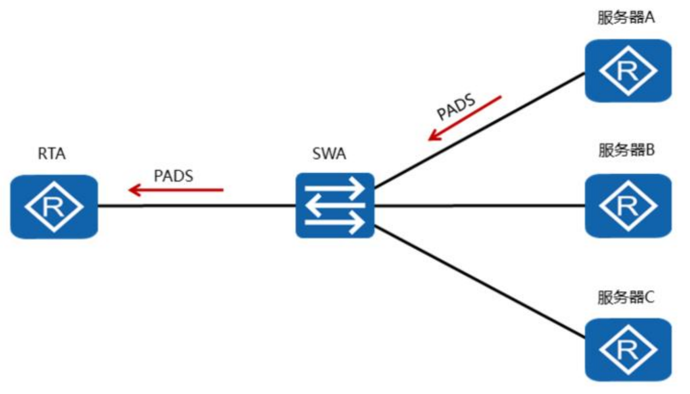
PPPoE发现阶段



- PPPoE客户端选择最先收到的PADO报文对应的PPPoE服务器，并单播发送一个PADR报文。

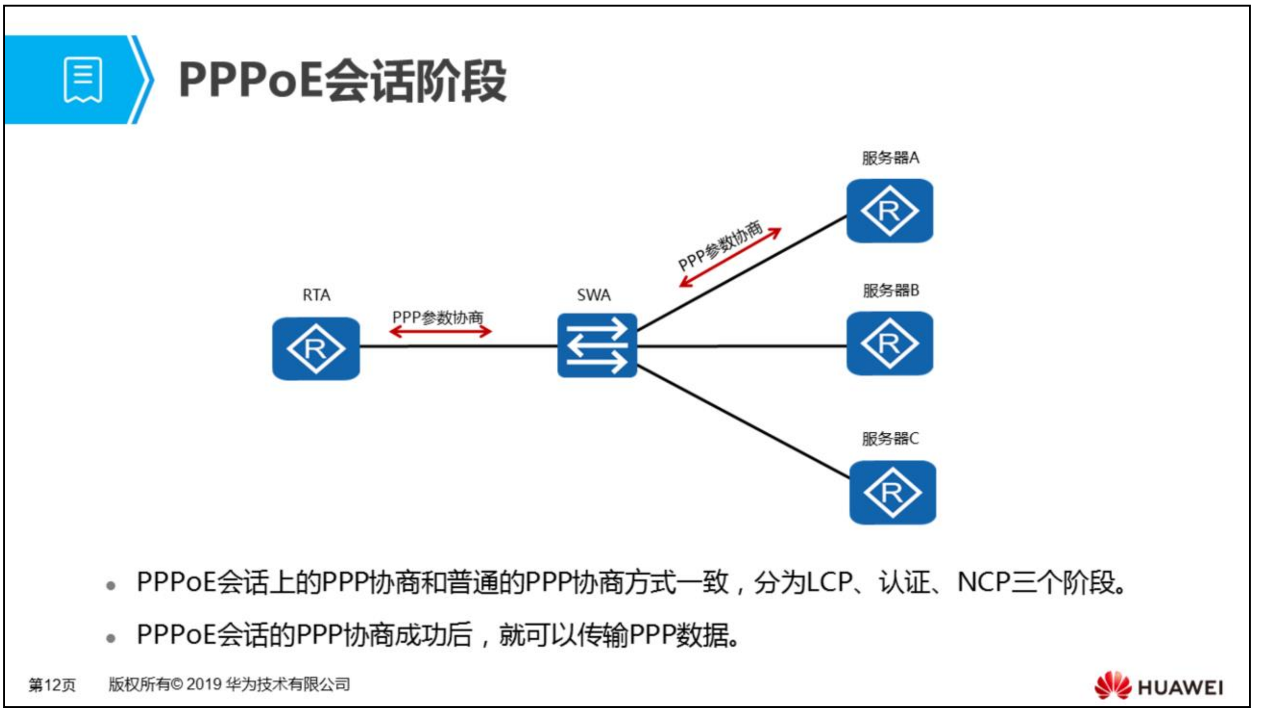
- 因为PPPoE客户端是以广播的形式发送PADI报文，所以客户端可能会收到多个PADO报文。在接收到的所有PADO报文中，PPPoE客户端选择最先收到的PADO报文对应的PPPoE服务器，并发送一个PADR报文给这个服务器。在PADR报文中，目的地址是选中的服务器的MAC地址，Code字段为0x19，Session ID字段为0x0000。

PPPoE发现阶段



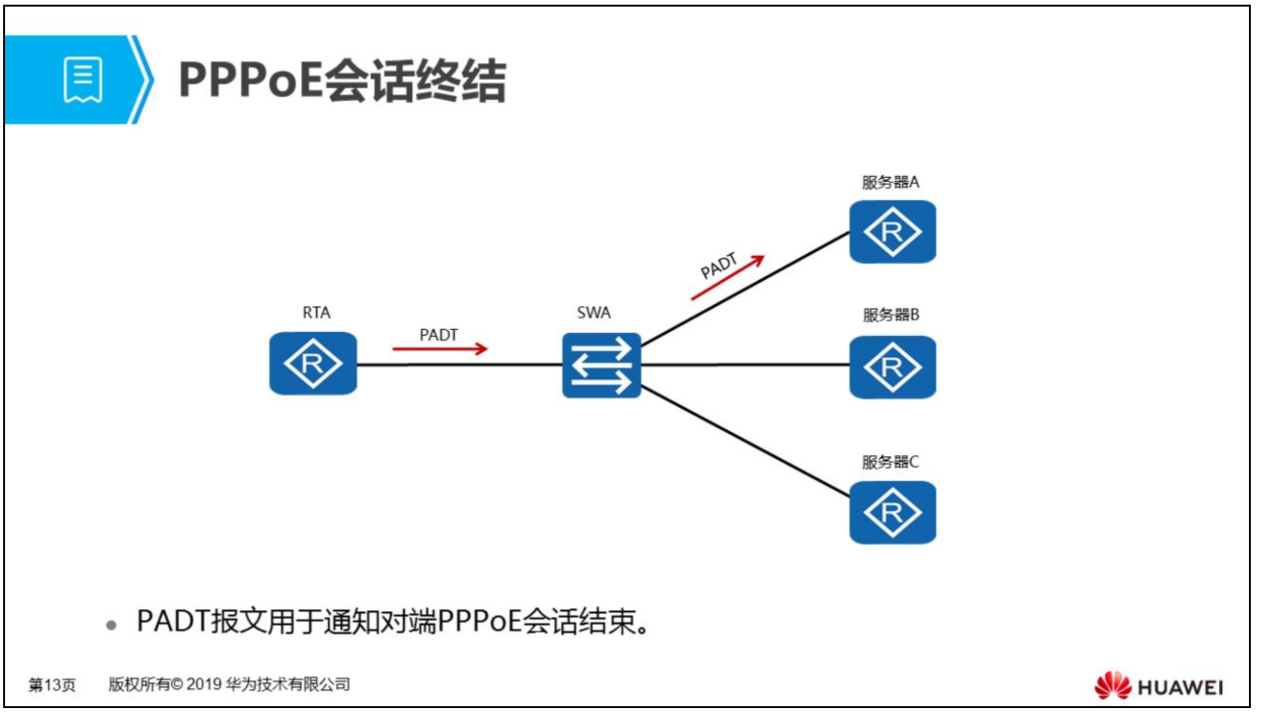
- PPPoE服务器生成唯一的PPPoE Session ID，并发送PADS报文给客户端，会话建立成功。

- PPPoE服务器收到PADR报文后，会生成一个唯一的Session ID来标识和PPPoE客户端的会话，并通过一个PADS报文把Session ID发送给PPPoE客户端。在PADS报文中，目的地址是PPPoE客户端的MAC地址，Code字段为0x65，Session ID字段是PPPoE服务器为本PPPoE会话产生的Session ID。会话建立成功后，PPPoE客户端和服务器进入PPPoE会话阶段。

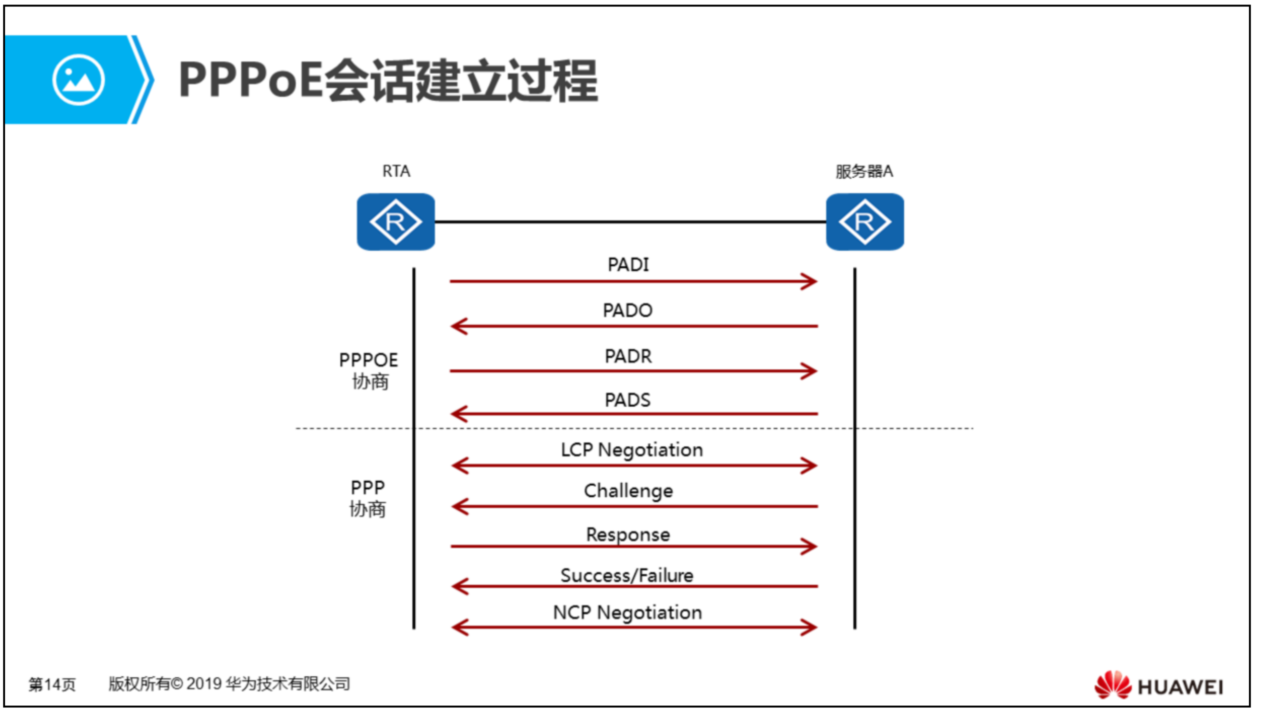


- PPPoE会话阶段可分为两部分：PPP协商阶段和PPP报文传输阶段。

1. PPPoE Session上的PPP协商和普通的PPP协商方式一致，分为LCP、认证、NCP三个阶段。LCP阶段主要完成建立、配置和检测数据链路连接。LCP协商成功后，开始进行认证，认证协议类型由LCP协商结果决定。认证成功后，PPP进入NCP阶段，NCP是一个协议族，用于配置不同的网络层协议，常用的是IP控制协议（IPCP），它负责配置用户的IP地址和DNS服务器地址等。
2. PPPoE Session的PPP协商成功后，就可以承载PPP数据报文。在这一阶段传输的数据包中必须包含在发现阶段确定的Session ID并保持不变。

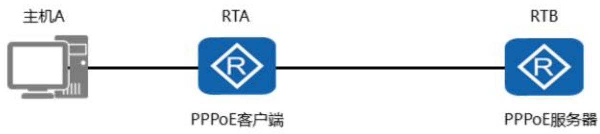


- 当PPPOE客户端希望关闭连接时，可以向PPPOE服务器端发送一个PADT报文。同样，如果PPPOE服务器端希望关闭连接时，也可以向PPPOE客户端发送一个PADT报文，此报文用于关闭连接。
- 在PADT报文中，目的MAC地址为单播地址，Session ID为希望关闭的连接的Session ID。一旦收到一个PADT报文之后，连接随即关闭。



- 用户客户端向服务器发送一个PADI报文，开始PPPOE接入。
- 服务器向客户端发送PADO报文。
- 客户端根据回应，发起PADR请求给服务器。
- 服务器产生一个Session ID，通过PADS发给客户端。
- 客户端和服务器之间进行PPP的LCP协商，建立链路层通信。同时，协商使用CHAP认证方式。
- 服务器通过Challenge报文发送给认证客户端，提供一个128bit的Challenge。
- 客户端收到Challenge报文后，并将密码和Challenge做MD5算法运算后，在Response回应报文中把结果发送给服务器。
- 服务器根据用户发送的信息判断用户是否合法，然后回应认证成功/失败报文，将认证结果返回给客户端。
- 进行NCP（如IPCP）协商，通过服务器获取到规划的IP地址等参数。

PPPoE配置




```
[RTA]dialer-rule
[RTA-dialer-rule]dialer-rule 1 ip permit
[RTA-dialer-rule]quit
[RTA]interface dialer 1
[RTA-Dialer1]dialer user enterprise
[RTA-Dialer1]dialer-group 1
[RTA-Dialer1]dialer bundle 1
[RTA-Dialer1]ppp chap user enterprise@huawei
[RTA-Dialer1]ppp chap password cipher huawei123
[RTA-Dialer1]ip address ppp-negotiate
```

  HUAWEI

- PPPoE客户端配置包括三个步骤。
- 首先需要配置一个拨号接口。
- **dialer-rule**命令用于进入Dialer-rule视图，在该视图下，可以通过拨号规则来配置发起PPPoE会话的条件。
- **interface dialer** *number*命令用来创建并进入Dialer接口。
- **dialer user** *user-name*命令用于配置对端用户名，这个用户名必须与对端服务器上的PPP用户名相同。
- **dialer-group** *group-number*命令用来将接口置于一个拨号访问组。
- **dialer bundle** *number*命令用来指定Dialer接口使用的Dialer bundle。设备通过Dialer bundle将物理接口与拨号接口关联起来。

PPPoE配置


```
[RTA]interface GigabitEthernet 0/0/1
[RTA-GigabitEthernet0/0/1]pppoe-client dial-bundle-number 1 on-demand
[RTA-GigabitEthernet0/0/1]quit
[RTA]ip route-static 0.0.0.0 0 dialer 1
```

  HUAWEI

- 第二个步骤是在接口上将Dialer Bundle和接口绑定：
- **pppoe-client dial-bundle-number** *number*命令来实现Dialer Bundle和物理接口的绑定，用来指定PPPoE会话对应的Dialer Bundle，其中*number*是与PPPoE会话相对应的Dialer Bundle编号。**on-demand**表示PPPoE会话工作在按需拨号模式。AR2200支持报文触发方式的按需拨号。目前ARG3系列路由器支持的按需拨号方式为报文触发方式，即当物理线路Up后，设备不会立即发起PPPoE呼叫，只有当有数据需要传送时，设备才会发起PPPoE呼叫，建立PPPoE会话。
- 第三个步骤是配置一条缺省静态路由，该路由允许在路由表中没有相应匹配表项的流量都能通过拨号接口发起PPPoE会话。

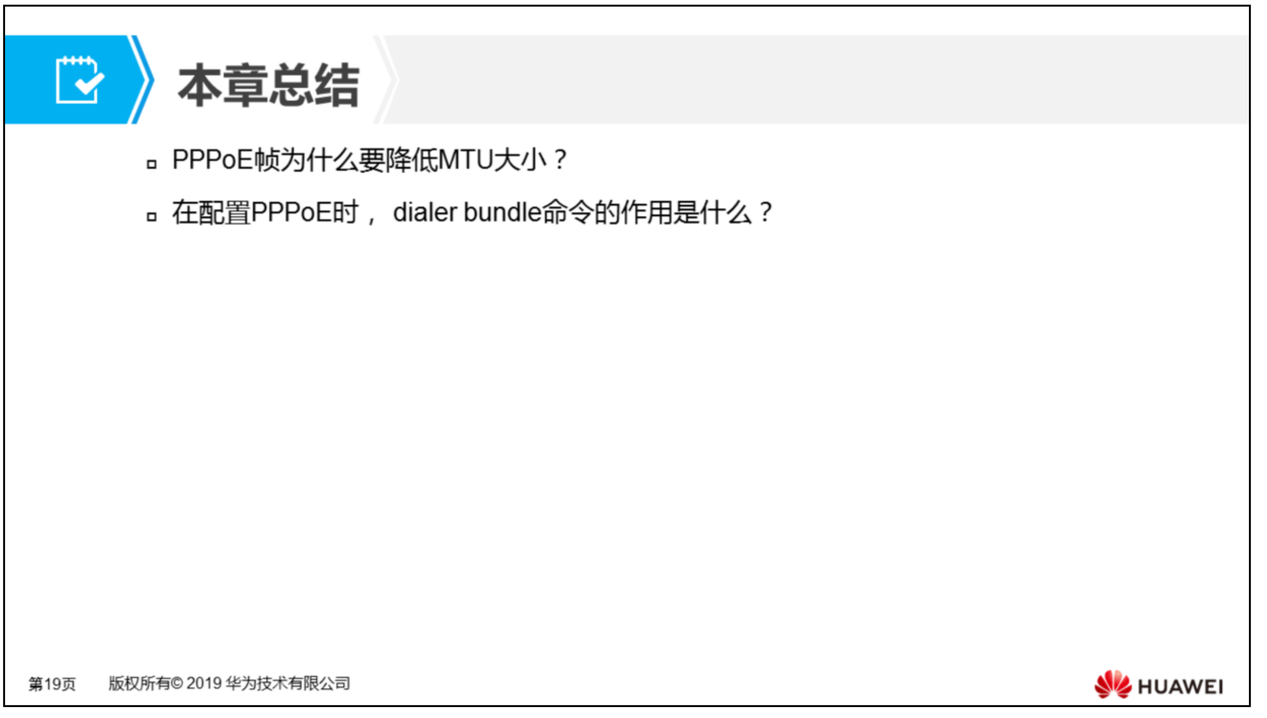
- **display interface dialer**[*number*]命令用于查看拨号接口的配置，便于定位拨号接口的故障。
- **LCP opened, IPCP opened**表示链路的状态完全正常。

配置验证

```
[RTA]display pppoe-client session summary
PPPoE Client Session:
ID    Bundle  Dialer  Intf           Client-MAC    Server-MAC    State
0     1       1       GE0/0/1        54899876830c  000000000000  IDLE
```

```
[RTA]display pppoe-client session summary
PPPoE Client Session:
ID    Bundle  Dialer  Intf           Client-MAC    Server-MAC    State
1     1       1       GE0/0/1        00e0fc0308f6  00e0fc036781  UP
```


- **display pppoe-client session summary**命令用于查看PPPoE客户端的PPPoE会话状态和统计信息。
- 本节给出了两个例子来说明不同的PPPoE会话状态。
- **ID**表示PPPoE会话ID，Bundle ID和Dialer ID的值与拨号参数配置有关。
- **Intf**表示客户端侧协商时的物理接口。
- **State**表示PPPoE会话的状态，包括以下四种：
- 1. **IDLE**表示当前会话状态为空闲。
- 2. **PADI**表示PPPoE会话处于发现阶段，并已经发送PADI报文。
- 3. **PADR**表示PPPoE会话处于发现阶段，并已经发送PADR报文。
- 4. **UP**表示PPPoE会话建立成功。



- 以太网中默认最大支持1500字节的有效载荷。PPPoE头部长度为6字节，PPP协议ID长度为2字节，所以PPPoE帧中的MTU不能超过1492字节。
- 2. dialer bundle命令用来指定Dialer接口使用的Dialer bundle。设备通过Dialer bundle将物理接口与拨号接口关联起来。


谢 谢
www.huawei.com

网络地址转换

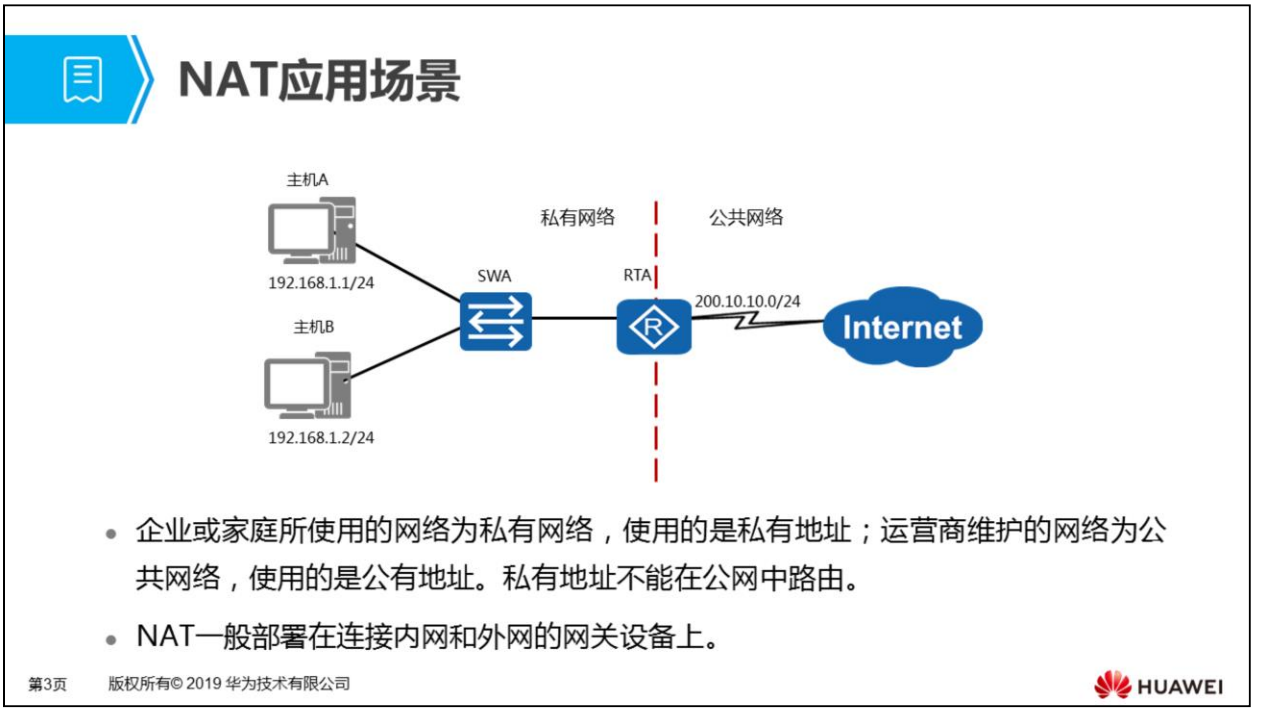
前言

- 随着Internet的发展和网络应用的增多，IPv4地址枯竭已经成为制约网络发展的瓶颈。尽管IPv6可以从根本上解决IPv4地址空间不足的问题，但目前众多的网络设备和网络应用仍是基于IPv4的，因此在IPv6广泛应用之前，一些过渡技术的使用是解决这个问题的主要技术手段。
- 网络地址转换技术NAT（Network Address Translation）主要用于实现位于内部网络的主机访问外部网络的功能。当局域网内的主机需要访问外部网络时，通过NAT技术可以将其私网地址转换为公网地址，并且多个私网用户可以共用一个公网地址，这样既可保证网络互通，又节省了公网地址。

  HUAWEI

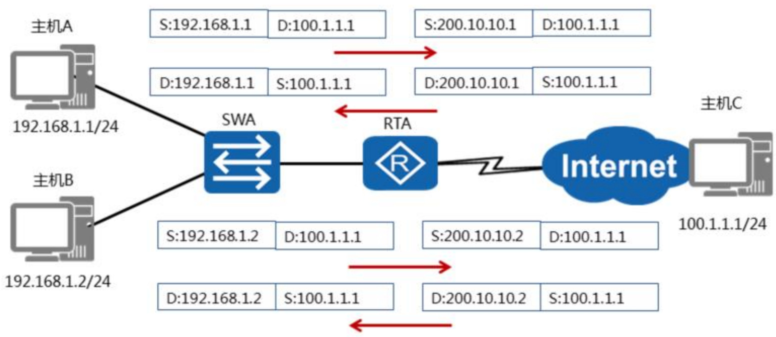
目标

- 学完本课程后，您将能够：
  - 掌握NAT的工作原理
  - 掌握NAT的基本配置



- 随着网络设备的数量不断增长，对IPv4地址的需求也不断增加，导致可用IPv4地址空间逐渐耗尽。解决IPv4地址枯竭问题的权宜之计是分配可重复使用的各类私网地址段给企业内部或家庭使用。但是，私有地址不能在公网中路由，即私网主机不能与公网通信，也不能通过公网与另外一个私网通信。
- NAT是将IP数据报文头部中的IP地址转换为另一个IP地址的过程，主要用于实现内部网络（私有IP地址）访问外部网络（公有IP地址）的功能。NAT一般部署在连接内网和外网的网关设备上。当收到的报文源地址为私网地址、目的地址为公网地址时，NAT可以将源私网地址转换成一个公网地址。这样公网目的地就能够收到报文，并做出响应。此外，网关上还会创建一个NAT映射表，以便判断从公网收到的报文应该发往的私网目的地址。

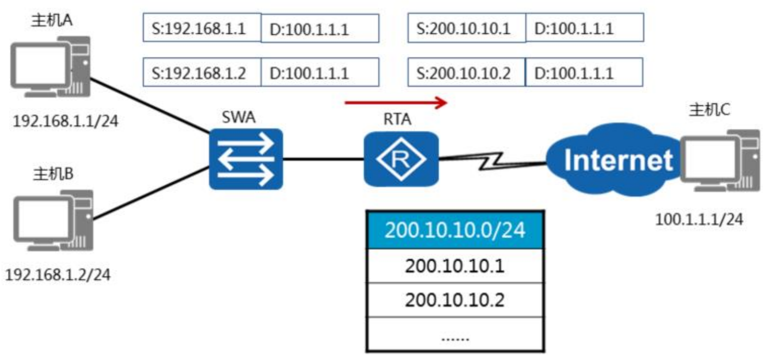
静态NAT



- 静态NAT实现了私有地址和公有地址的一对一映射。
- 一个公网IP只会分配给唯一且固定的内网主机。

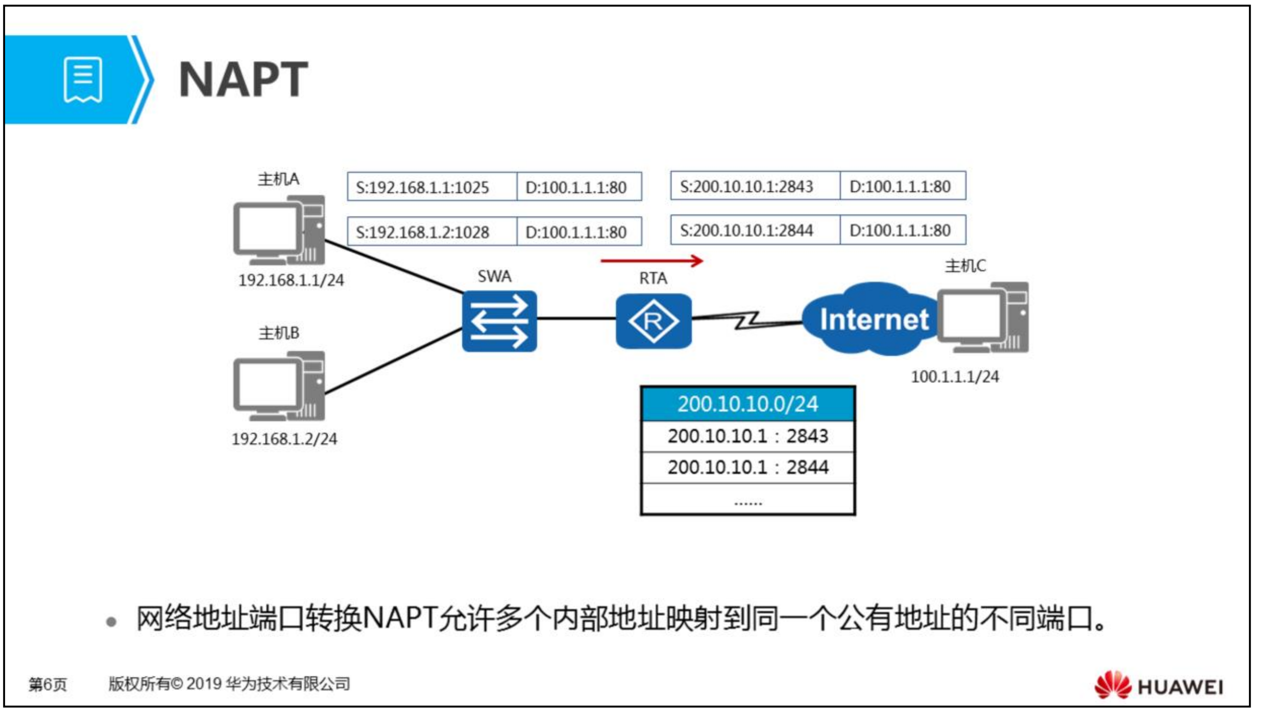
- NAT的实现方式有多种，适用于不同的场景。
- 静态NAT实现了私有地址和公有地址的一对一映射。如果希望一台主机优先使用某个关联地址，或者想要外部网络使用一个指定的公网地址访问内部服务器时，可以使用静态NAT。但是在大型网络中，这种一对一的IP地址映射无法缓解公用地址短缺的问题。
- 在本示例中，源地址为192.168.1.1的报文需要发往公网地址100.1.1.1。在网关RTA上配置了一个私网地址192.168.1.1到公网地址200.10.10.1的映射。当网关收到主机A发送的数据包后，会先将报文中的源地址192.168.1.1转换为200.10.10.1，然后转发报文到目的设备。目的设备回复的报文目的地址是200.10.10.1。当网关收到回复报文后，也会执行静态地址转换，将200.10.10.1转换成192.168.1.1，然后转发报文到主机A。和主机A在同一个网络中其他主机，如主机B，访问公网的过程也需要网关RTA做静态NAT转换。

动态NAT



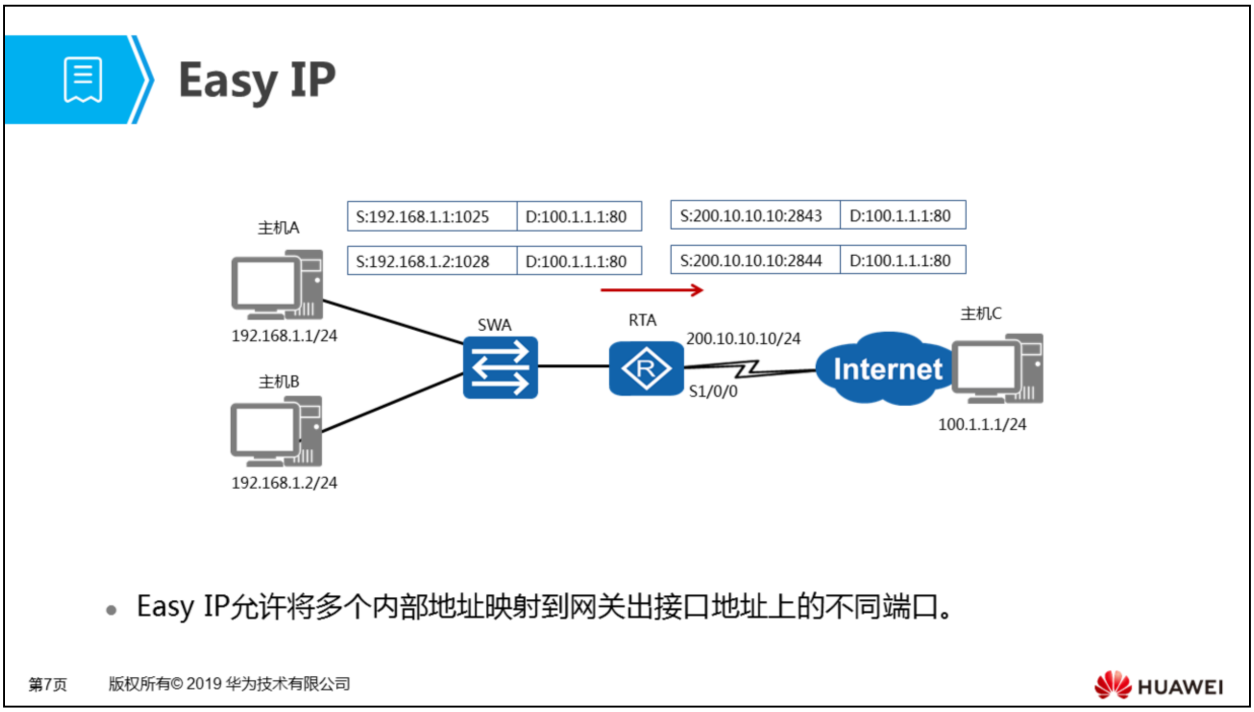
- 动态NAT基于地址池来实现私有地址和公有地址的转换。

- 动态NAT通过使用地址池来实现。
- 本示例中，当内部主机A和主机B需要与公网中的目的主机通信时，网关RTA会从配置的公网地址池中选择一个未使用的公网地址与之做映射。每台主机都会分配到地址池中的一个唯一地址。当不需要此连接时，对应的地址映射将会被删除，公网地址也会被恢复到地址池中待用。当网关收到回复报文后，会根据之前的映射再次进行转换之后转发给对应主机。
- 动态NAT地址池中的地址用尽以后，只能等待被占用的公用IP被释放后，其他主机才能使用它来访问公网。



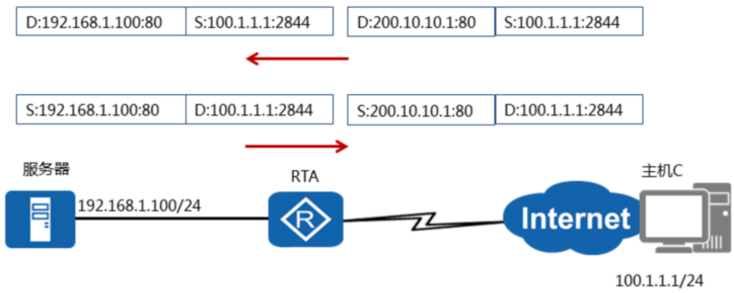
网络地址端口转换NAPT（Network Address Port Translation）允许多个内部地址映射到同一个公有地址的不同端口。

- 本例中，RTA收到一个私网主机发送的报文，源IP地址是192.168.1.1，源端口号是1025，目的IP地址是100.1.1.1，目的端口是80。RTA会从配置的公网地址池中选择一个空闲的公网IP地址和端口号，并建立相应的NAPT表项。这些NAPT表项指定了报文的私网IP地址和端口号与公网IP地址和端口号的映射关系。之后，RTA将报文的源IP地址和端口号转换成公网地址200.10.10.1和端口号2843，并转发报文到公网。当网关RTA收到回复报文后，会根据之前的映射表再次进行转换之后转发给主机A。主机B同理。



- Easy IP适用于小规模局域网中的主机访问Internet的场景。小规模局域网通常部署在小型的网吧或者办公室中，这些地方内部主机不多，出接口可以通过拨号方式获取一个临时公网IP地址。Easy IP可以实现内部主机使用这个临时公网IP地址访问Internet。
- 本示例说明了Easy IP的实现过程。RTA收到一个主机A访问公网的请求报文，报文的源IP地址是192.168.1.1，源端口号是1025。RTA会建立Easy IP表项，这些表项指定了源IP地址和端口号与出接口的公网IP地址和端口号的映射关系。之后，根据匹配的Easy IP表项，将报文的源IP地址和端口号转换成出接口的IP地址和端口号，并转发报文到公网。报文的源IP地址转换成200.10.10.10/24，相应的端口号是2843。
- 路由器收到回复报文后，会根据报文的目的IP地址和端口号，查询Easy IP表项。路由器根据匹配的Easy IP表项，将报文的目的IP地址和端口号转换成私网主机的IP地址和端口号，并转发报文到主机。

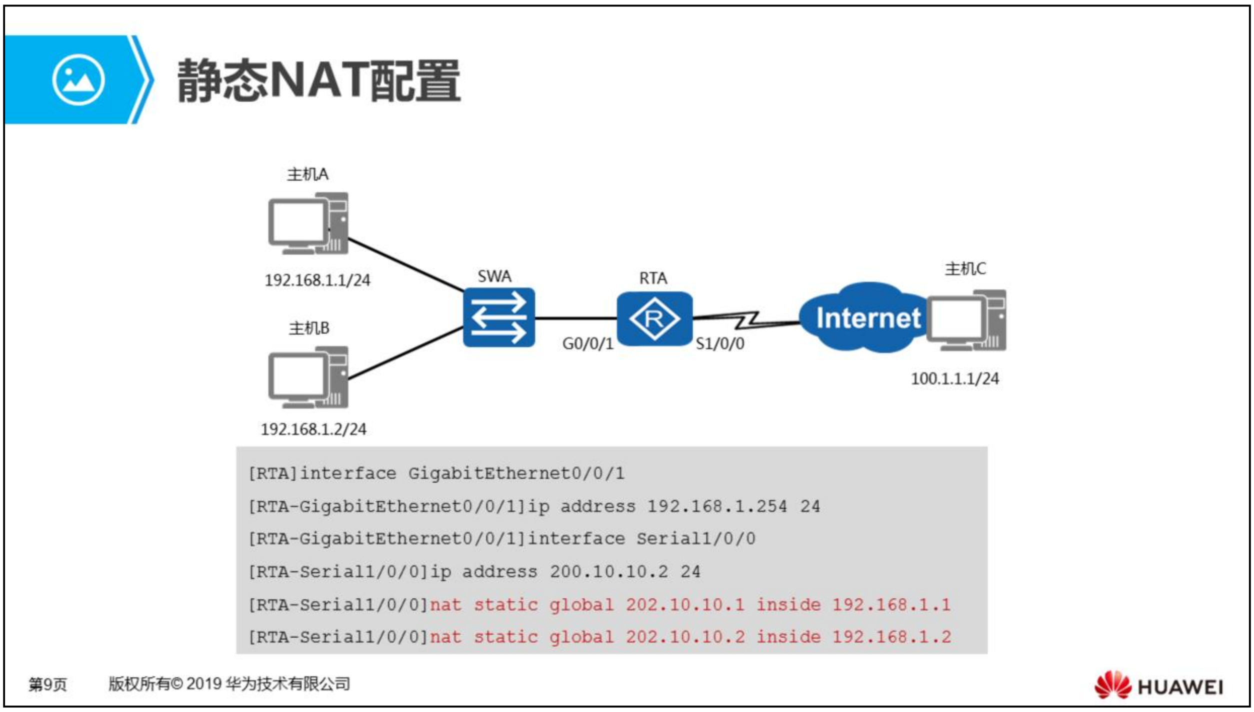
NAT服务器



- 通过配置NAT服务器,可以使外网用户访问内网服务器。

  HUAWEI

- NAT在使内网用户访问公网的同时，也屏蔽了公网用户访问私网主机的需求。当一个私网需要向公网用户提供Web和SFTP服务时，私网中的服务器必须随时可供公网用户访问。
- NAT服务器可以实现这个需求，但是需要配置服务器私网IP地址和端口号转换为公网IP地址和端口号并发布出去。路由器在收到一个公网主机的请求报文后，根据报文的目的IP地址和端口号查询地址转换表项。路由器根据匹配的地址转换表项，将报文的目的IP地址和端口号转换成私网IP地址和端口号，并转发报文到私网中的服务器。
- 本例中，主机C需要访问私网服务器，发送报文的目的IP地址是200.10.10.1，目的端口号是80。RTA收到此报文后会查找地址转换表项，并将目的IP地址转换成192.168.1.1，目的端口号保持不变。服务器收到报文后会进行响应，RTA收到私网服务器发来的响应报文后，根据报文的源IP地址192.168.1.1和端口号80查询地址转换表项。然后，路由器根据匹配的地址转换表项，将报文的源IP地址和端口号转换成公网IP地址200.10.10.1和端口号80，并转发报文到目的公网主机。



- **nat static global** { *global-address*} **inside** {*host-address* } 命令用于创建静态NAT。

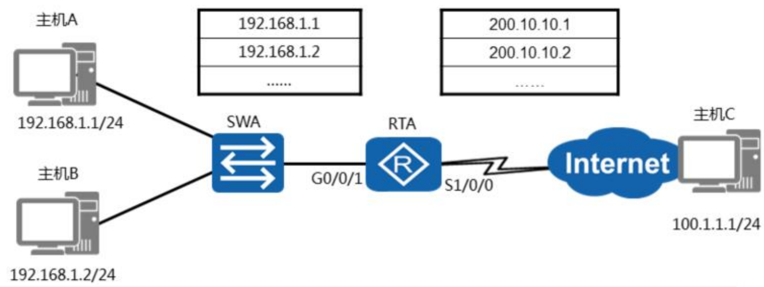
1. global参数用于配置外部公网地址。
2. inside参数用于配置内部私有地址。

配置验证

```
[RTA]display nat static
  Static Nat Information:
  Interface  : Serial1/0/0
    Global IP/Port     : 202.10.10.1/----
    Inside IP/Port     : 192.168.1.1/----
……
 Global IP/Port     : 202.10.10.2/----
    Inside IP/Port     : 192.168.1.2/----
……

  Total :    2
```


- 命令**display nat static**用于查看静态NAT的配置。
- Global IP/Port表示公网地址和服务端口号。
- Inside IP/Port表示私有地址和服务端口号。

动态NAT配置




```
[RTA]nat address-group 1 200.10.10.1 200.10.10.200
[RTA]acl 2000
[RTA-acl-basic-2000]rule 5 permit source 192.168.1.0 0.0.0.255
[RTA-acl-basic-2000]quit
[RTA]interface serial1/0/0
[RTA-Serial1/0/0]nat outbound 2000 address-group 1 no-pat
```


- **nat outbound**命令用来将一个访问控制列表ACL和一个地址池关联起来，表示ACL中规定的地址可以使用地址池进行地址转换。ACL用于指定一个规则，用来过滤特定流量。后续将会介绍有关ACL的详细信息。
- **nat address-group**命令用来配置NAT地址池。
- 本示例中使用**nat outbound**命令将ACL 2000与待转换的192.168.1.0/24网段的流量关联起来，并使用地址池1（**address-group 1**）中的地址进行地址转换。**no-pat**表示只转换数据报文的地址而不转换端口信息。

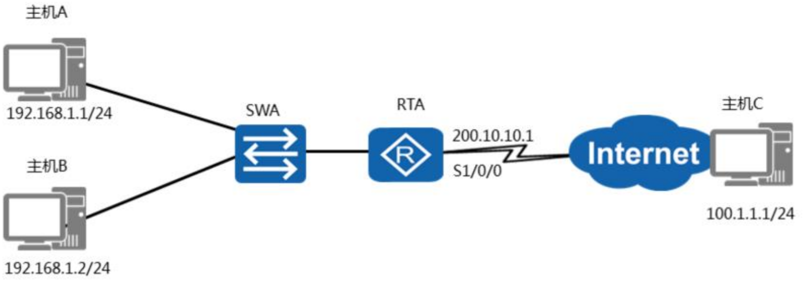
配置验证

```
[RTA]display nat address-group 1
 NAT Address-Group Information:
 --------------------------------------
 Index    Start-address      End-address
 1        200.10.10.1        200.10.10.200

[RTA]display nat outbound
 NAT Outbound Information:
 ----------------------------------------------------------------
 Interface           Acl     Address-group/IP/Interface      Type
 ----------------------------------------------------------------
 Serial1/0/0        2000                          1         no-pat
 ----------------------------------------------------------------
  Total : 1
```


HUAWEI

- **display nat address-group** *group-index*命令用来查看NAT地址池配置信息。
- 命令**display nat outbound**用来查看动态NAT配置信息。
- 可以用这两条命令验证动态NAT的详细配置。在本示例中，指定接口Serial1/0/0与ACL关联在一起，并定义了用于地址转换的地址池1。参数**no-pat**说明没有进行端口地址转换。

Easy IP配置




```
[RTA]acl 2000
[RTA-acl-basic-2000]rule 5 permit source 192.168.1.0 0.0.0.255
[RTA-acl-basic-2000]quit
[RTA]interface serial1/0/0
[RTA-Serial1/0/0]nat outbound 2000
```


- **nat outbound** *acl-number*命令用来配置Easy-IP地址转换。Easy IP的配置与动态NAT的配置类似，需要定义ACL和**nat outbound**命令，主要区别是Easy IP不需要配置地址池，所以**nat outbound**命令中不需要配置参数**address-group**。
- 在本示例中，命令**nat outbound 2000**表示对ACL 2000定义的地址段进行地址转换，并且直接使用Serial1/0/0接口的IP地址作为NAT转换后的地址。

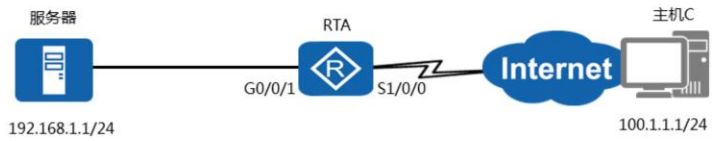
配置验证

```
[RTA]display nat outbound
 NAT Outbound Information:
 ---------------------------------------------------------------------
 Interface          Acl     Address-group/IP/Interface      Type
 ---------------------------------------------------------------------
 Serial1/0/0        2000      200.10.10.1                  easyip
 ---------------------------------------------------------------------
  Total : 1
```

  HUAWEI

- 命令**display nat outbound**用于查看命令**nat outbound**的配置结果。
- Address-group/IP/Interface表项表明接口和ACL已经关联成功，type表项表明Easy IP已经配置成功。

NAT服务器配置




```
[RTA]interface GigabitEthernet0/0/1
[RTA-GigabitEthernet0/0/1]ip address 192.168.1.254 24
[RTA-GigabitEthernet0/0/1]interface Serial1/0/0
[RTA-Serial1/0/0]ip address 200.10.10.2 24
[RTA-Serial1/0/0]nat server protocol tcp global 202.10.10.1 www
inside 192.168.1.1 8080
```

  HUAWEI

- **nat server** [**protocol** {*protocol-number* | icmp | tcp | udp} **global** { *global-address* | current-interface *global-port*} **inside** {*host-address host-port* } **vpn-instance** *vpn-instance-name* **acl** *acl-number* **description** *description*]命令用来定义一个内部服务器的映射表，外部用户可以通过公网地址和端口来访问内部服务器。
- 参数protocol指定一个需要地址转换的协议；
- 参数*global-address*指定需要转换的公网地址；
- 参数inside指定内网服务器的地址。

配置验证

```
[RTA]display nat server
 Nat Server Information:
 Interface  : Serial1/0/0
   Global IP/Port     : 202.10.10.1/80(www)
   Inside IP/Port     : 192.168.1.1/8080
   Protocol : 6(tcp)
   VPN instance-name  : ----
   Acl number         : ----
   Description : ----

 Total :    1
```


- **display nat server**命令用于查看详细的NAT服务器配置结果。
- 可以通过此命令验证地址转换的接口、全局和内部IP地址以及关联的端口号。在本示例中，全局地址202.10.10.1和关联的端口号80（www）分别被转换成内部服务器地址192.168.1.1和端口号8080。

本章总结

- 哪种NAT转换允许服务器既能被内部访问又能被外部访问？
- NAPT有什么功能和特点？

- 通过NAT内部服务器配置，将公网地址与一个私网服务器地址绑定，在地址转换后，外网主机便可以通过公有地址访问内网服务器。同时，私网地址用户可以通过服务器的私网地址访问内网服务器。
- NAPT是基于端口的转换，而不是基于IP地址的转换。NAPT允许多个内部地址映射到同一个公有地址的不同端口。


谢 谢
www.huawei.com

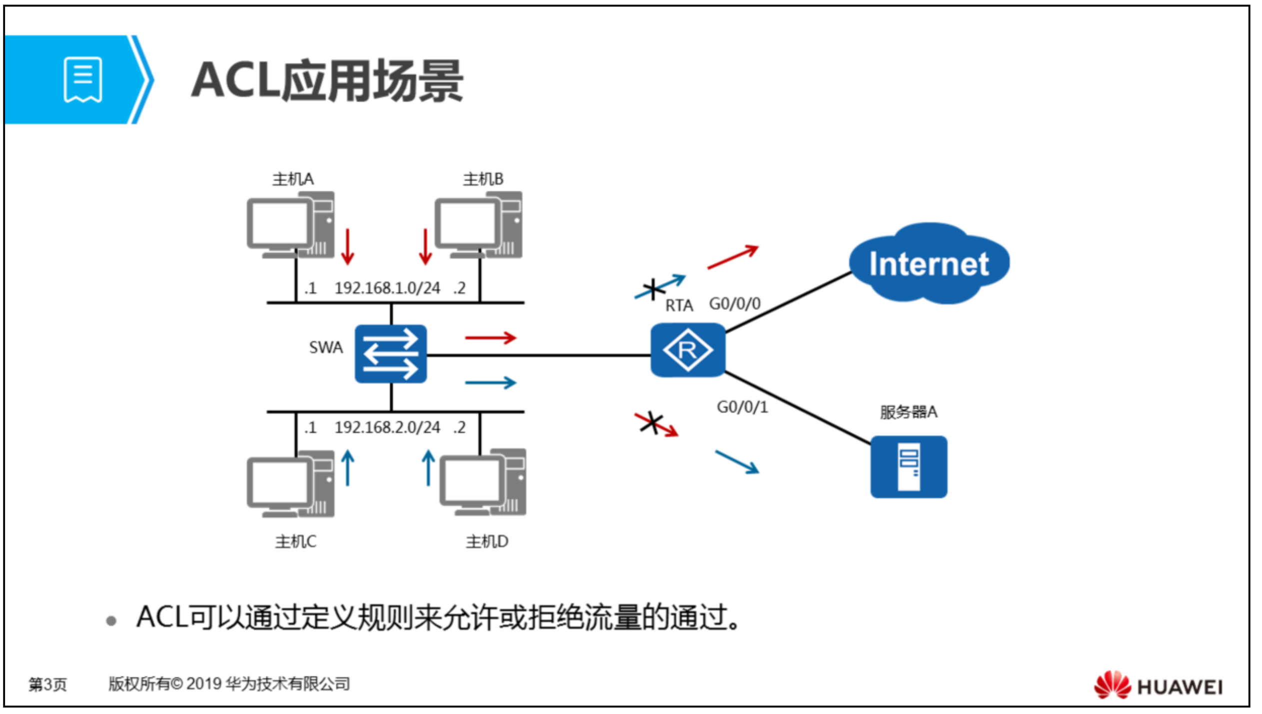
访问控制列表

前言

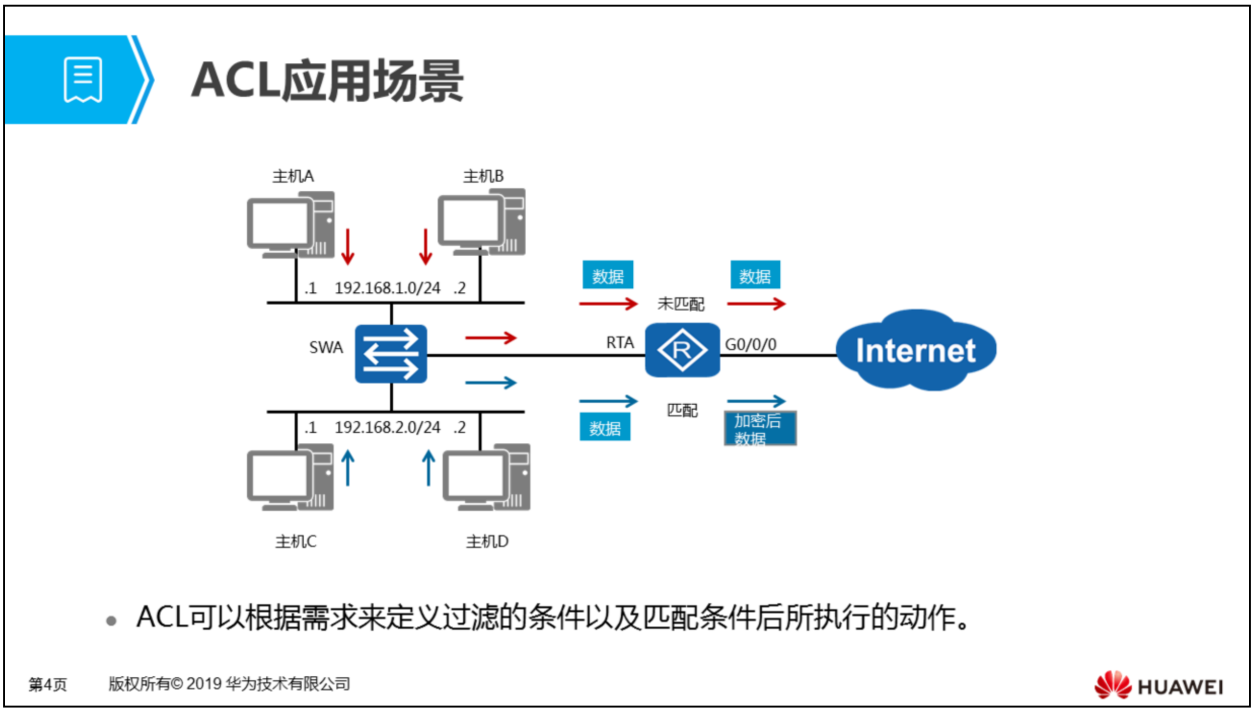
- 企业网络中的设备进行通信时，需要保障数据传输的安全可靠和网络的性能稳定。
- 访问控制列表ACL（Access Control List）可以定义一系列不同的规则，设备根据这些规则对数据包进行分类，并针对不同类型的报文进行不同的处理，从而可以实现对网络访问行为的控制、限制网络流量、提高网络性能、防止网络攻击等等。

目标

- 学完本课程后，您将能够：
  - 掌握ACL在企业网络中的应用
  - 掌握ACL的工作原理
  - 掌握ACL的配置



- ACL是由一系列规则组成的集合。设备可以通过这些规则对数据包进行分类，并对不同类型的报文进行不同的处理。
- 本示例中，网关RTA允许192.168.1.0/24中的主机可以访问外网，也就是Internet；而192.168.2.0/24中的主机则被禁止访问Internet。对于服务器A而言，情况则相反。网关允许192.168.2.0/24中的主机访问服务器A，但却禁止192.168.1.0/24中的主机访问服务器A。



- 设备可以依据ACL中定义的条件（例如源IP地址）来匹配入方向的数据，并对匹配了条件的数据执行相应的动作。在本示例所述场景中，RTA依据所定义的ACL而匹配到的感兴趣流量来自192.168.2.0/24网络，RTA会对这些感兴趣流量进行加密（虚拟局域网VPN中会进行介绍）之后再转发。

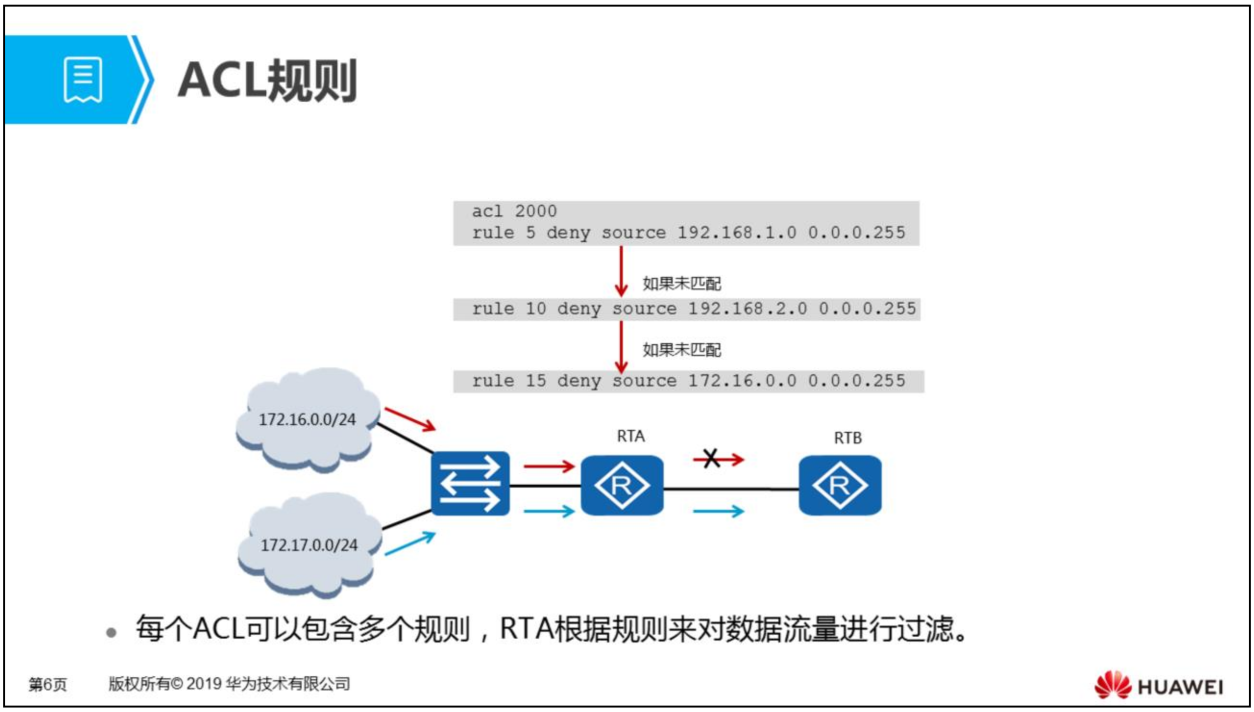
ACL分类

分类	编号范围	参数
基本ACL	2000-2999	源IP地址等
高级ACL	3000-3999	源IP地址、目的IP地址、 源端口、目的端口等
二层ACL	4000-4999	源MAC地址、目的MAC地址、以太帧协议类型等

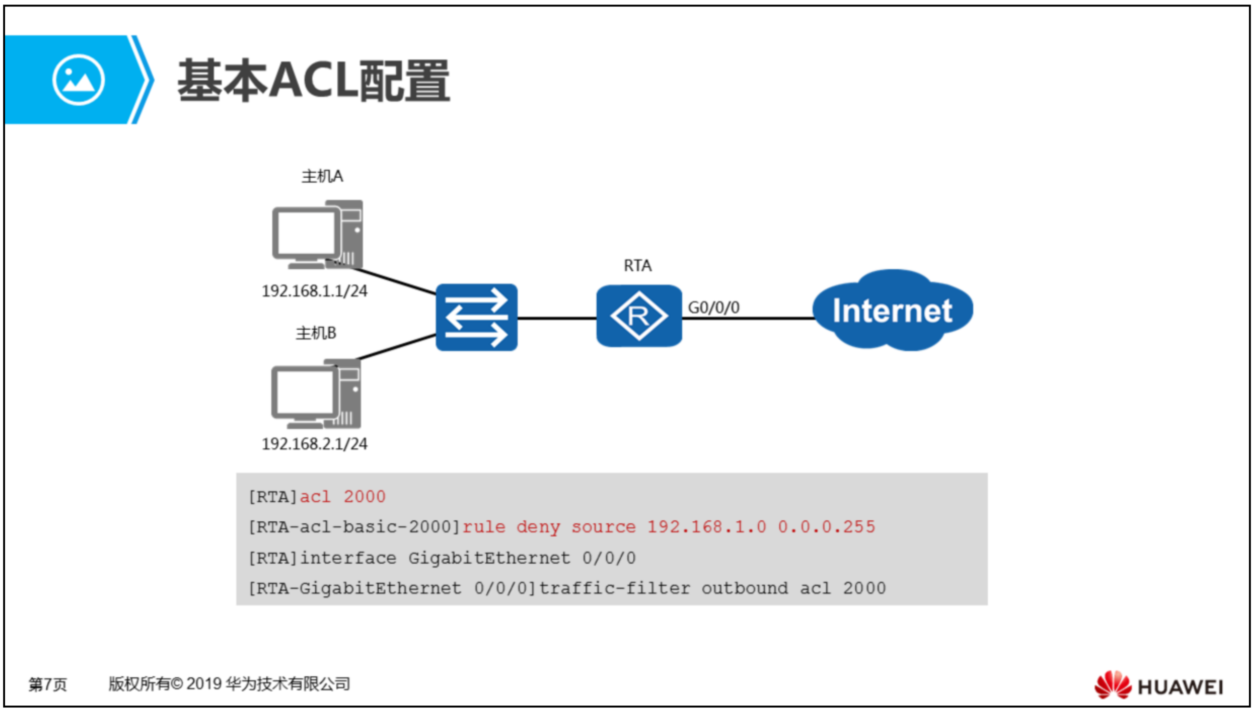
  HUAWEI

- 根据不同的划分规则，ACL可以有不同的分类。最常见的三种分类是基本ACL、高级ACL和二层ACL。

1. 基本ACL可以使用报文的源IP地址、分片标记和时间段信息来匹配报文，其编号取值范围是2000-2999。
2. 高级ACL可以使用报文的源/目的IP地址、源/目的端口号以及协议类型等信息来匹配报文。高级ACL可以定义比基本ACL更准确、更丰富、更灵活的规则，其编号取值范围是3000-3999。
3. 二层ACL可以使用源/目的MAC地址以及二层协议类型等二层信息来匹配报文，其编号取值范围是4000-4999。



- 一个ACL可以由多条“deny | permit”语句组成，每一条语句描述了一条规则。设备收到数据流量后，会逐条匹配ACL规则，看其是否匹配。如果不匹配，则匹配下一条。一旦找到一条匹配的规则，则执行规则中定义的动作，并不再继续与后续规则进行匹配。如果找不到匹配的规则，则设备不对报文进行任何处理。需要注意的是，ACL中定义的这些规则可能存在重复或矛盾的地方。规则的匹配顺序决定了规则的优先级，ACL通过设置规则的优先级来处理规则之间重复或矛盾的情形。
- ARG3系列路由器支持两种匹配顺序：配置顺序和自动排序。
- 配置顺序按ACL规则编号（rule-id）从小到大的顺序进行匹配。设备会在创建ACL的过程中自动为每一条规则分配一个编号，规则编号决定了规则被匹配的顺序。例如，如果将步长设定为5，则规则编号将按照5、10、15…这样的规律自动分配。如果步长设定为2，则规则编号将按照2、4、6、8…这样的规律自动分配。通过设置步长，使规则之间留有一定的空间，用户可以在已存在的两个规则之间插入新的规则。路由器匹配规则时默认采用配置顺序。另外，ARG3系列路由器默认规则编号的步长是5。
- 自动排序使用“深度优先”的原则进行匹配，即根据规则的精确度排序。
- 本示例中，RTA收到了来自两个网络的报文。默认情况下，RTA会依据ACL的配置顺序来匹配这些报文。网络172.16.0.0/24发送的数据流量将被RTA上配置的ACL2000的规则15匹配，因此会被拒绝。而来自网络172.17.0.0/24的报文不能匹配访问控制列表中的任何规则，因此RTA对报文不做任何处理，而是正常转发。



- **acl** [**number**] 命令用来创建一个ACL，并进入ACL视图。
- **rule** [*rule-id*] { **deny** | **permit** } **source** { *source-address source-wildcard* | **any** } 命令用来增加或修改ACL的规则。**deny**用来指定拒绝符合条件的数据包，**permit**用来指定允许符合条件的数据包，**source**用来指定ACL规则匹配报文的源地址信息，**any**表示任意源地址。
- **traffic-filter** { **inbound** | **outbound** }**acl**{ *acl-number* }命令用来在接口上配置基于ACL对报文进行过滤。
- 本示例中，主机A发送的流量到达RTA后，会匹配ACL2000中创建的规则rule deny source 192.168.1.0 0.0.0.255
- ，因而将被拒绝继续转发到Internet。主机B发送的流量不匹配任何规则，所以会被RTA正常转发到Internet。

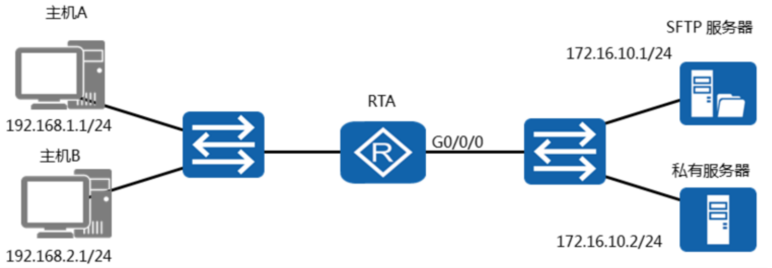
配置确认

```
[RTA]display acl 2000
Basic ACL 2000, 1 rule
Acl's step is 5
 rule 5 deny source 192.168.1.0 0.0.0.255
```

```
[RTA]display  traffic-filter applied-record
-----------------------------------------------------------
Interface                   Direction  AppliedRecord
-----------------------------------------------------------
GigabitEthernet0/0/0        outbound   acl 2000
-----------------------------------------------------------
```


- 执行**display acl** *<acl-number>*命令可以验证配置的基本ACL。
- 本例中，所配置的ACL只有一条规则，即拒绝源IP地址在192.168.1.0/24范围的所有IP报文。
- 执行**display traffic-filter applied-record**命令可以查看设备上所有基于ACL进行报文过滤的应用信息，这些信息可以帮助用户了解报文过滤的配置情况并核对其是否正确，同时也有助于进行相关的故障诊断与排查。

高级ACL配置




```
[RTA]acl 3000
[RTA-acl-adv-3000]rule deny tcp source 192.168.1.0 0.0.0.255
destination 172.16.10.1 0.0.0.0 destination-port eq 21
[RTA-acl-adv-3000]rule deny tcp source 192.168.2.0 0.0.0.255
destination 172.16.10.2 0.0.0.0
[RTA-acl-adv-3000]rule permit ip
[RTA-GigabitEthernet 0/0/0]traffic-filter outbound acl 3000
```


- 基本ACL可以依据源IP地址进行报文过滤，而高级ACL能够依据源/目的IP地址、源/目的端口号、网络层及传输层协议以及IP流量分类和TCP标记值等各种参数（SYN|ACK|FIN等）进行报文过滤。
- 本示例中，RTA上定义了高级ACL3000，其中第一条规则“rule deny tcp source 192.168.1.0 0.0.0.255 destination 172.16.10.1 0.0.0.0 destination-port eq 21”用于限制源地址范围是192.168.1.0/24，目的IP地址为172.16.10.1，目的端口号为21的所有TCP报文；第二条规则“rule deny tcp source 192.168.2.0 0.0.0.255 destination 172.16.10.2 0.0.0.0 ”用于限制源地址范围是192.168.2.0/24，目的地址是172.16.10.2的所有TCP报文；第三条规则“rule permit ip”用于匹配所有IP报文，并对报文执行允许动作。

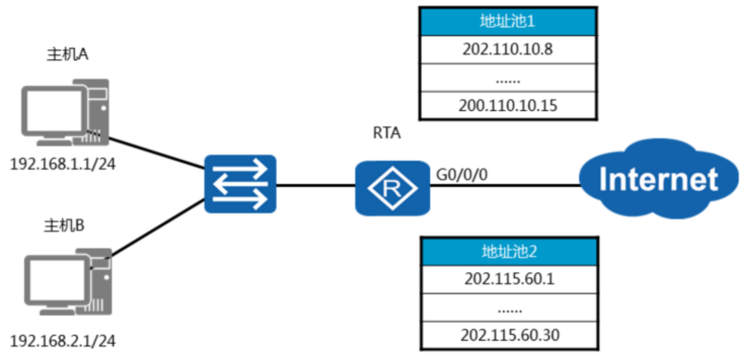
配置验证

```
[RTA]display acl 3000
Advanced ACL 3000, 3 rules
Acl's step is 5
rule 5 deny tcp source 192.168.1.0 0.0.0.255 destination 172.16.10.1 0
destination-port eq sftp
rule 10 deny tcp source 192.168.2.0 0.0.0.255 destination 172.16.10.2
0
rule 15 permit ip
```

```
[RTA]display  traffic-filter applied-record
-----------------------------------------------------------
Interface                    Direction  AppliedRecord
-----------------------------------------------------------
GigabitEthernet0/0/0         outbound   acl 3000
-----------------------------------------------------------
```


- 执行**display acl** *<acl-number>*命令可以验证配置的高级ACL。
- 显示信息表明：RTA上一共配置了3条高级ACL规则。第一条规则用于拒绝来自源IP地址192.168.1.0/24，目的IP地址为172.16.10.1，目的端口为21（SFTP）的TCP报文；第二条规则用于拒绝来自源IP地址192.168.2.0/24，目的IP地址为172.16.10.2的所有TCP报文；第三条规则允许所有IP报文通过。

ACL应用-NAT



- 本例要求通过ACL来实现主机A和主机B分别使用不同的公网地址池来进行NAT转换。

- ACL还可用于网络地址转换操作，以便在存在多个地址池的情况下，确定哪些内网地址是通过哪些特定外网地址池进行地址转换的。例如，某企业网络作为客户连接到多个服务供应商网络，企业网络内的用户位于不同的网段/子网，他们期望分别通过某个特定的地址组进行地址转换来实现报文转发。这种情况极有可能发生在连接不同服务供应商网络的路由器上行端口。
- 本示例中，要求192.168.1.0/24中的主机使用地址池1中的公网地址进行地址转换，而192.168.2.0/24中的主机使用地址池2中的公网地址进行地址转换。

ACL应用-NAT

```
[RTA]nat address-group 1 202.110.10.8 202.110.10.15
[RTA]nat address-group 2 202.115.60.1 202.115.60.30
[RTA]acl 2000
[RTA-acl-basic-2000]rule permit source 192.168.1.0 0.0.0.255
[RTA-acl-basic-2000]acl 2001
[RTA-acl-basic-2001]rule permit source 192.168.2.0 0.0.0.255
[RTA-acl-basic-2001]interface GigabitEthernet0/0/0
[RTA-GigabitEthernet0/0/0]nat outbound 2000 address-group 1
[RTA-GigabitEthernet0/0/0]nat outbound 2001 address-group 2
```


- 执行**nat outbound** *<acl-number>* **address-group** *<address-group number>*命令，可以将NAT与ACL绑定。
- 本示例中，私网192.168.1.0/24将使用地址池220.110.10.8-220.110.10.15进行地址转换，私网192.168.2.0/24将使用地址池202.115.60.1-202.115.60.30进行地址转换。

本章总结

- 高级ACL可以基于哪些条件来定义规则？

- 1.高级ACL可以基于源/目的IP地址，源/目的端口号，协议类型以及IP流量分类和TCP标记值（SYN|ACK|FIN等）等参数来定义规则。


谢 谢
www.huawei.com

AAA原理与配置

HUAWEI

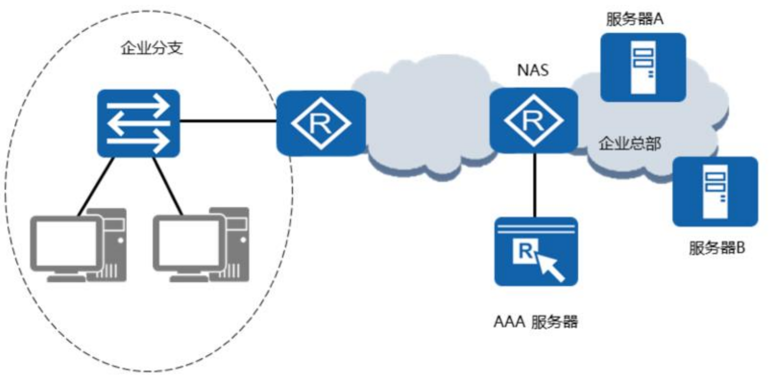
前言

- AAA是Authentication（认证）、Authorization（授权）和Accounting（计费）的简称，它提供了认证、授权、计费三种安全功能。AAA可以通过多种协议来实现，目前华为设备支持基于RADIUS（Remote Authentication Dial-In User Service）协议或HWTACACS（Huawei Terminal Access Controller Access Control System）协议来实现AAA。

目标

- 学完本课程后，您将能够：
  - 掌握AAA的基本概念
  - 掌握AAA认证和授权的配置

AAA 应用场景

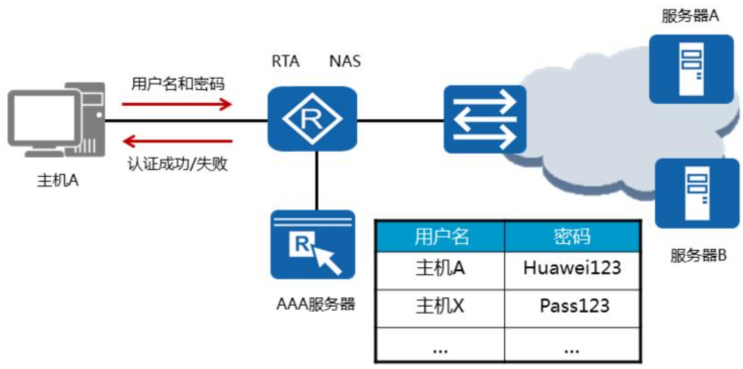


- AAA提供对用户进行认证、授权和计费三种安全功能。

  HUAWEI

- AAA是一种提供认证、授权和计费的安全技术。该技术可以用于验证用户帐户是否合法，授权用户可以访问的服务，并记录用户使用网络资源的情况。
- 例如，企业总部需要对服务器的资源访问进行控制，只有通过认证的用户才能访问特定的资源，并对用户使用资源的情况进行记录。在这种场景下，可以按照如图所示的方案进行AAA部署，NAS为网络接入服务器，负责集中收集和管理用户的访问请求。
- AAA服务器表示远端的Radius或HWTACACS服务器，负责制定认证、授权和计费方案。如果企业分支的员工希望访问总部的服务器，远端的Radius或HWTACACS服务器会要求员工发送正确的用户名和密码，之后会进行验证，通过后则执行相关的授权策略，接下来，该员工就可以访问特定的服务器了。如果还需要记录员工访问网络资源的行为，网络管理员还可以在Radius或HWTACACS服务器上配置计费方案。
- 目前，ARG3系列路由器只支持配置认证和授权。

认证



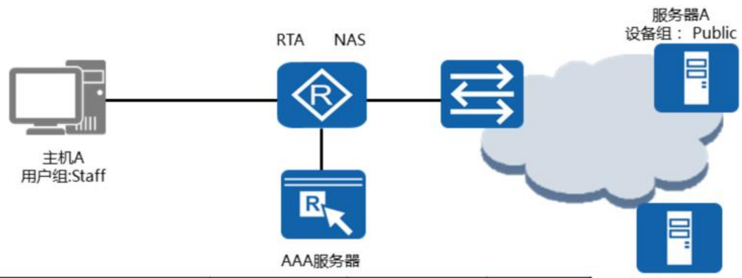
- 认证：验证用户是否可以获得网络访问的权限。
- AAA支持的认证方式有：不认证，本地认证，远端认证。

- AAA支持三种认证方式:

1. 不认证：完全信任用户，不对用户身份进行合法性检查。鉴于安全考虑，这种认证方式很少被采用。
2. 本地认证：将本地用户信息（包括用户名、密码和各种属性）配置在NAS上。本地认证的优点是处理速度快、运营成本低；缺点是存储信息量受设备硬件条件限制。
3. 远端认证：将用户信息（包括用户名、密码和各种属性）配置在认证服务器上。AAA支持通过RADIUS协议或HWTACACS协议进行远端认证。NAS作为客户端，与RADIUS服务器或HWTACACS服务器进行通信。

- 如果一个认证方案采用多种认证方式，这些认证方式按配置顺序生效。比如，先配置了远端认证，随后配置了本地认证，那么在远端认证服务器无响应时，会转入本地认证方式。如果只在本地设备上配置了登录账号，没有在远端服务器上配置，AR2200认为账号没有通过远端认证，不再进行本地认证。

授权



设备组	用户组	时间	权限级别
Private	admin	09:00-12:00	15
Public	admin	09:00-18:00	15
Public	Staff	09:00-18:00	2

- 授权：授权用户可以访问或使用网络上的哪些服务。
- AAA支持的授权方式有：不授权，本地授权，远端授权。

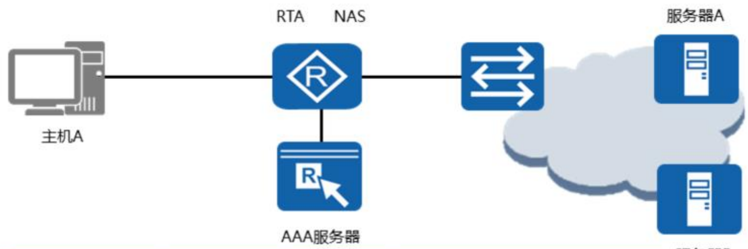
  HUAWEI

- AAA授权功能赋予用户访问的特定网络或设备的权限。AAA支持以下授权方式：

1. 不授权：不对用户进行授权处理。
2. 本地授权：根据NAS上配置的本地用户账号的相关属性进行授权。
3. 远端授权：HWTACACS授权，使用TACACS服务器对用户授权。RADIUS授权，对通过RADIUS服务器认证的用户授权。RADIUS协议的认证和授权是绑定在一起的，不能单独使用RADIUS进行授权。

- 如果在一个授权方案中使用多种授权方式，这些授权方式按照配置顺序生效。不授权方式最后生效。

计费



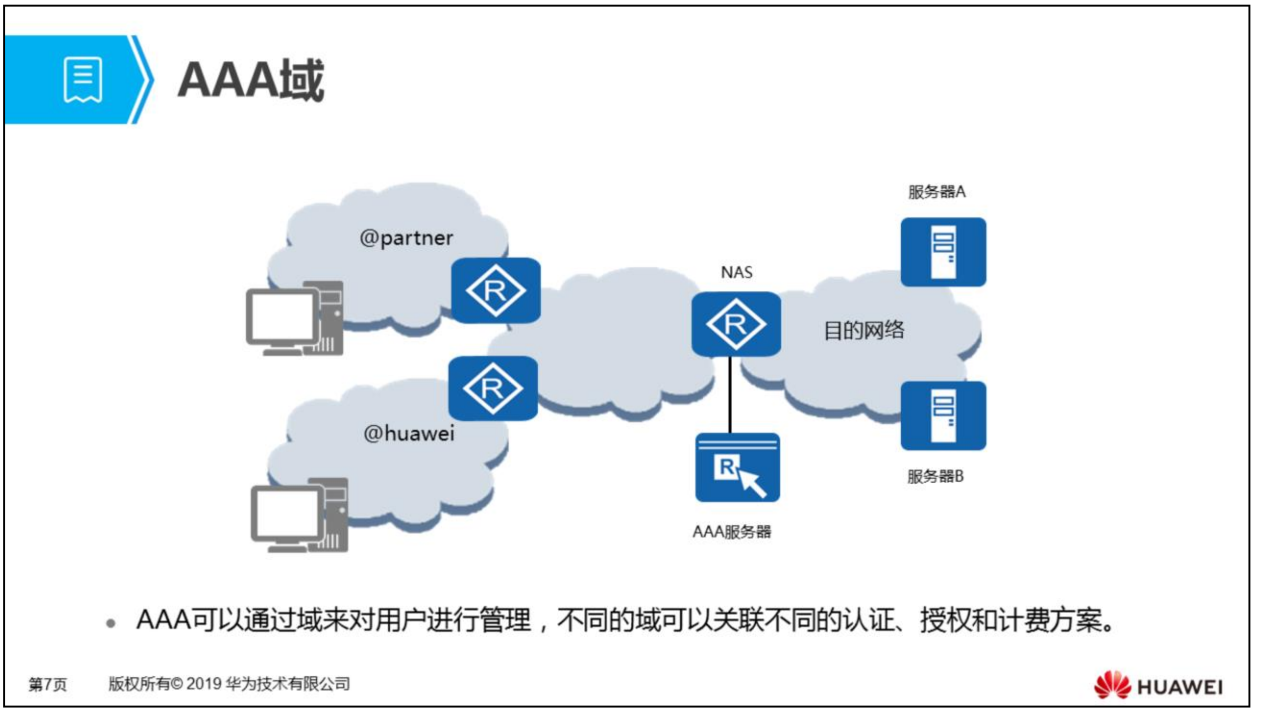
登录时间	用户名	上线时长	流量 上行/下行
2013.5.1 03:20:55	主机A	01:22:15	496.2KB / 21MB
2013.4.16 12:40:51	主机X	00:30:12	123KB / 1MB

- 计费：记录用户使用网络资源的情况。
- AAA支持的计费方式有：不计费，远端计费。

- 计费功能用于监控授权用户的网络行为和网络资源的使用情况。AAA支持以下两种计费方式：

1. 不计费：为用户提供免费上网服务，不产生相关活动日志。
2. 远端计费：通过RADIUS服务器或HWTACACS服务器进行远端计费。RADIUS服务器或HWTACACS服务器具备充足的储存空间，可以储存各授权用户的网络访问活动日志，支持计费功能。

- 本示例中展示了用户计费日志中记录的典型信息。

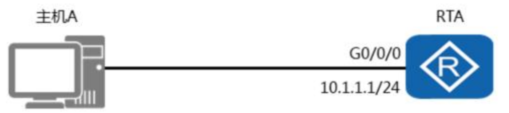


- 设备基于域来对用户进行管理，每个域都可以配置不同的认证、授权和计费方案，用于对该域下的用户进行认证、授权和计费。每个用户都属于某一个域。用户属于哪个域是由用户名中的域名分隔符@后的字符串决定。例如，如果用户名是user@huawei，则用户属于huawei域。如果用户名后不带有@，则用户属于系统缺省域default。
- ARG3系列路由设备支持两种缺省域：

1. default域为普通用户的缺省域。
2. default_admin域为管理用户的缺省域。

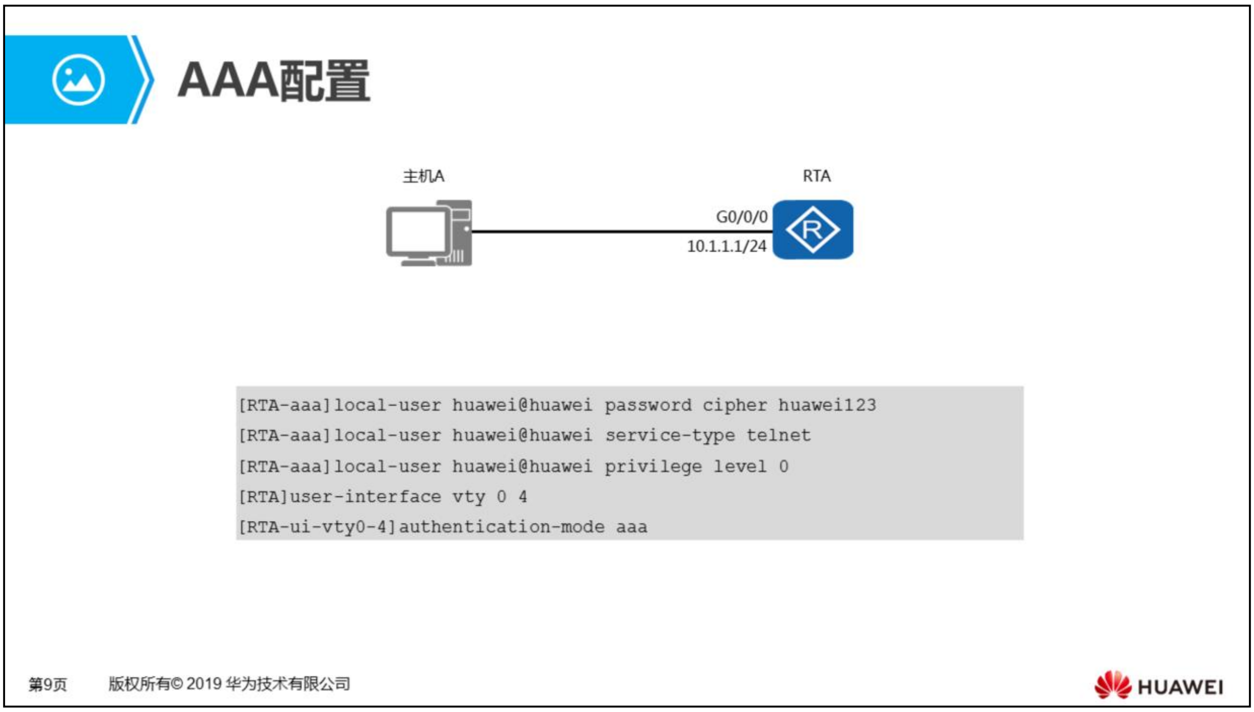
- 用户可以修改但不能删除这两个缺省域。默认情况下，设备最多支持32个域，包括两个缺省域。

AAA配置




```
[RTA]aaa
[RTA-aaa]authentication-scheme auth1
[RTA-aaa-authen-auth1]authentication-mode local
[RTA-aaa-authen-auth1]quit
[RTA-aaa]authorization-scheme auth2
[RTA-aaa-author-auth2]authorization-mode local
[RTA-aaa-author-auth2]quit
[RTA-aaa]domain huawei
[RTA-aaa-domain-huawei]authentication-scheme auth1
[RTA-aaa-domain-huawei]authorization-scheme auth2
[RTA-aaa-domain-huawei]quit
```


- **authentication-scheme** *authentication-scheme-name*命令用来配置域的认证方案。缺省情况下，域使用名为“default”的认证方案。
- **authentication-mode** { **hwtacacs** | **radius** | **local** }命令用来配置认证方式，**local**指定认证模方式为本地认证。缺省情况下，认证方式为本地认证。
- **authorization-scheme** *authorization-scheme-name*命令用来配置域的授权方案。缺省情况下，域下没有绑定授权方案。
- **authorization-mode** { [**hwtacacs** | **if-authenticated** | **local**] * [**none**] }命令用来配置当前授权方案使用的授权方式。缺省情况下，授权模式为本地授权方式。
- **domain** *domain-name*命令用来创建域，并进入AAA域视图。



- **local-user** *user-name* **password cipher** *password*命令用来创建本地用户，并配置本地用户的密码。如果用户名中带域名分隔符，如@，则认为@前面的部分是用户名，后面部分是域名。如果没有@，则整个字符串为用户名，域为默认域。
- **local-user** *user-name* **privilege level** *level*命令用来指定本地用户的优先级。

- 免责申明：
- 设备支持通过Telnet协议和Stelnet协议登录。使用Telnet、Stelnet v1协议存在安全风险，建议您使用STelnet v2登录设备。

配置验证

```
[RTA]display domain name huawei
  Domain-name                        : huawei
  Domain-state                       : Active
  Authentication-scheme-name         : auth1
  Accounting-scheme-name             : default
  Authorization-scheme-name          : auth2
  Service-scheme-name                : -
  RADIUS-server-template             : -
  HWTACACS-server-template           : -
  User-group                         : -
```

- AAA中，每个域都会与相应的认证授权和计费方案相关联。

- **display domain** [**name** *domain-name*]命令用来查看域的配置信息。
- Domain-state为Active表示激活状态。Authentication-scheme-name表示域使用的认证方案为auth1。缺省情况下，域使用系统自带的default认证方案。Authorization-scheme-name表示域使用的授权方案为auth2。

本章总结

- ARG3系列路由器上支持配置哪些AAA方案?
- 如果在ARG3系列路由器上创建用户时，没有关联自定义的域，则该用户属于哪个域?

- ARG3系列路由器上支持配置认证方案和授权方案，计费方案需要配置在HWTACACS或RADIUS服务器上。
- 如果创建用户时未指定用户所属的域，用户会自动关联缺省域default。


谢 谢
www.huawei.com

IPSec VPN原理与配置

前言

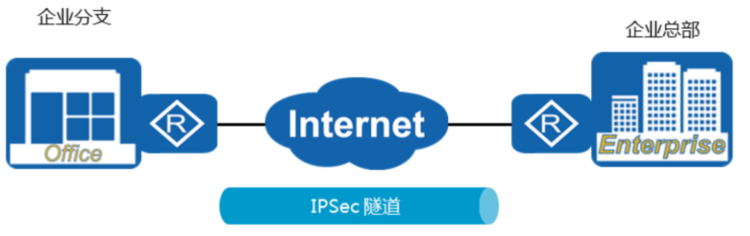
- 企业对网络安全性的需求日益提升，而传统的TCP/IP协议缺乏有效的安全认证和保密机制。IPSec（Internet Protocol Security）作为一种开放标准的安全框架结构，可以用来保证IP数据报文在网络上传输的机密性、完整性和防重放。

HUAWEI

目标

- 学完本课程后，您将能够：
  - 掌握IPSec VPN的基本概念
  - 掌握IPSec VPN的基本配置

IPSec VPN应用场景



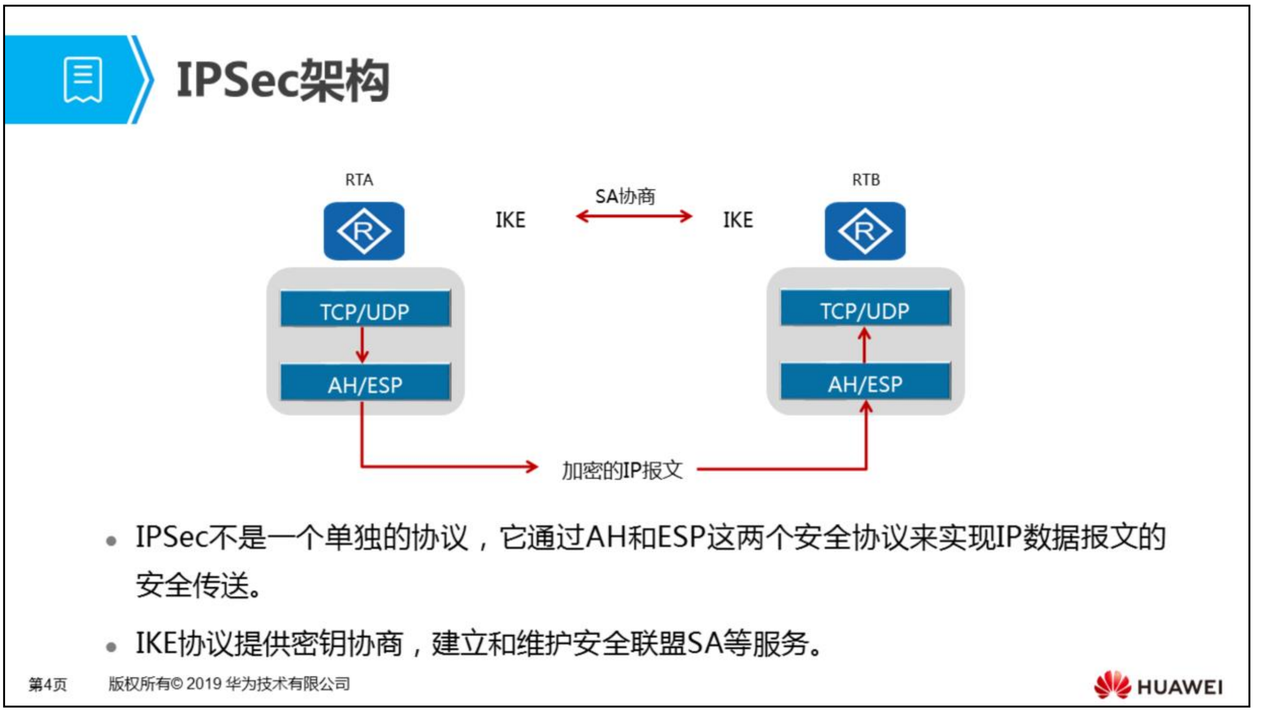
- 企业分支可以通过IPSec VPN接入到企业总部网络。

  HUAWEI

- IPSec是IETF定义的一个协议组。通信双方在IP层通过加密、完整性校验、数据源认证等方式，保证了IP数据报文在网络上传输的机密性、完整性和防重放。

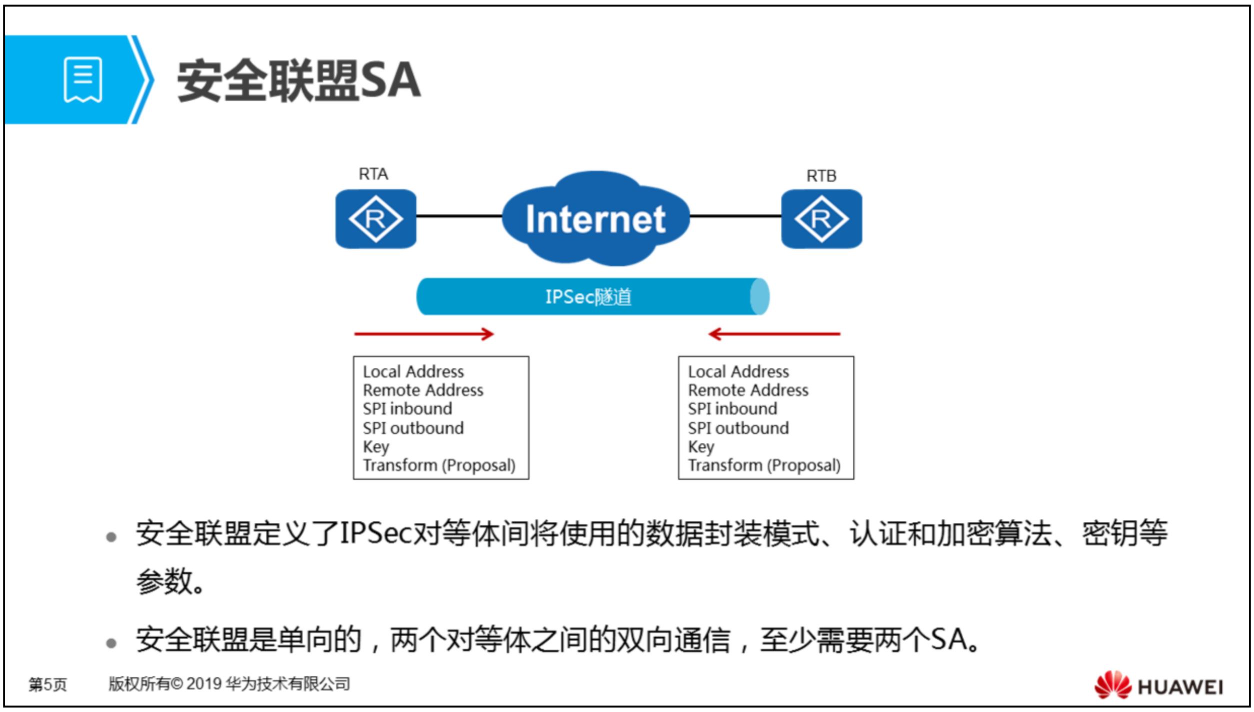
1. 机密性（Confidentiality）指对数据进行加密保护，用密文的形式传送数据。
2. 完整性（Data integrity）指对接收的数据进行认证，以判定报文是否被篡改。
3. 防重放（Anti-replay）指防止恶意用户通过重复发送捕获到的数据包所进行的攻击，即接收方会拒绝旧的或重复的数据包。

- 企业远程分支机构可以通过使用IPSec VPN建立安全传输通道，接入到企业总部网络。



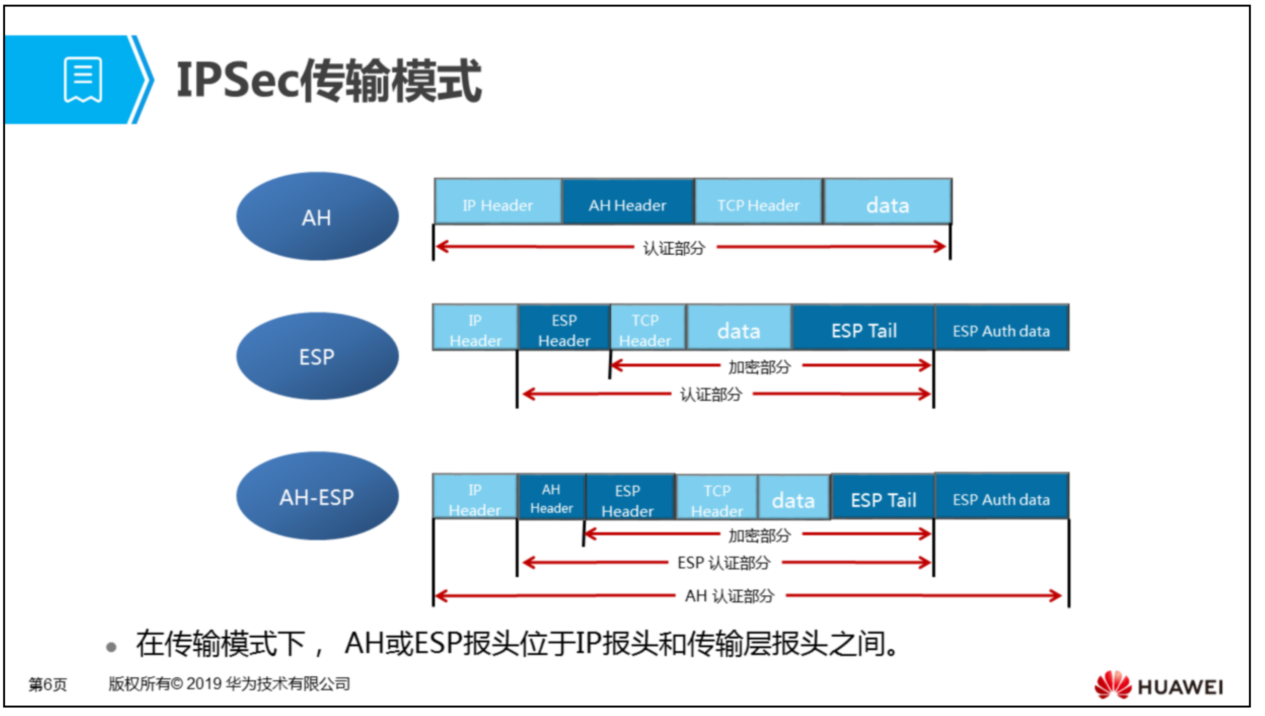
- IPSec VPN体系结构主要由AH（Authentication Header）、ESP（Encapsulating Security Payload）和IKE（Internet Key Exchange）协议套件组成。

1. AH协议：主要提供的功能有数据源验证、数据完整性校验和防报文重放功能。然而，AH并不加密所保护的数据报。
2. ESP协议：提供AH协议的所有功能外（但其数据完整性校验不包括IP头），还可提供对IP报文的加密功能。
3. IKE协议：用于自动协商AH和ESP所使用的密码算法。



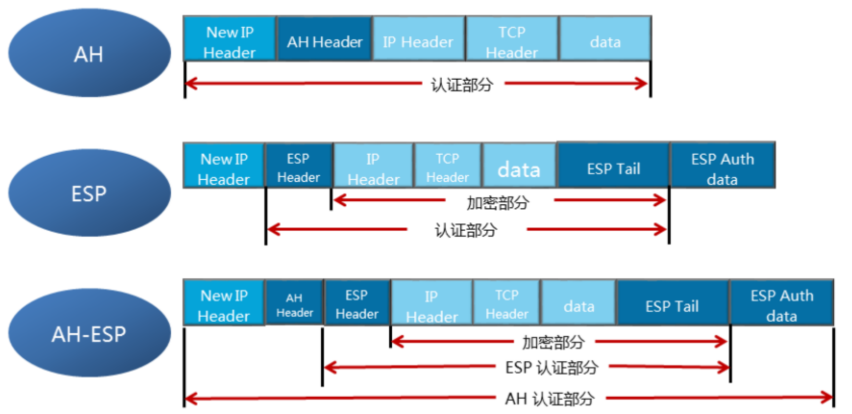
- SA（Security Association）安全联盟定义了IPSec通信对等体间将使用的数据封装模式、认证和加密算法、秘钥等参数。SA是单向的，两个对等体之间的双向通信，至少需要两个SA。如果两个对等体希望同时使用AH和ESP安全协议来进行通信，则对等体针对每一种安全协议都需要协商一对SA。
- SA由一个三元组来唯一标识，这个三元组包括安全参数索引SPI（Security Parameter Index）、目的IP地址、安全协议（AH或ESP）。
- 建立SA的方式有以下两种：

1. 手工方式：安全联盟所需的全部信息都必须手工配置。手工方式建立安全联盟比较复杂，但优点是可以不依赖IKE而单独实现IPSec功能。当对等体设备数量较少时，或是在小型静态环境中，手工配置SA是可行的。
2. IKE动态协商方式：只需要通信对等体间配置好IKE协商参数，由IKE自动协商来创建和维护SA。动态协商方式建立安全联盟相对简单些。对于中、大型的动态网络环境中，推荐使用IKE协商建立SA。



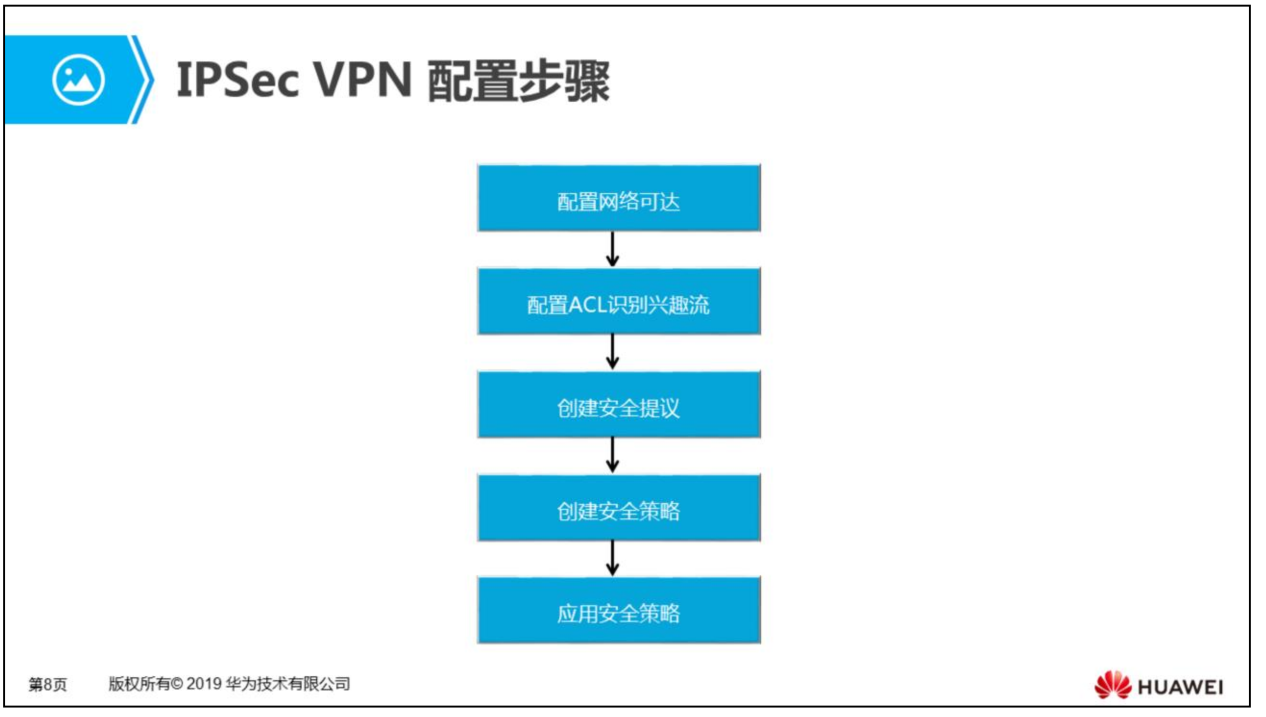
- IPSec协议有两种封装模式：传输模式和隧道模式。
- 传输模式中，在IP报文头和高层协议之间插入AH或ESP头。传输模式中的AH或ESP主要对上层协议数据提供保护。
- 传输模式中的AH：在IP头部之后插入AH头，对整个IP数据包进行完整性校验。
- 传输模式中的ESP：在IP头部之后插入ESP头，在数据字段后插入尾部以及认证字段。对高层数据和ESP尾部进行加密，对IP数据包中的ESP报文头，高层数据和ESP尾部进行完整性校验。
- 传输模式中的AH+ESP：在IP头部之后插入AH和ESP头，在数据字段后插入尾部以及认证字段。

IPSec隧道模式



- 在隧道模式下，IPSec会另外生成一个新的IP报头，并封装在AH或ESP之前。

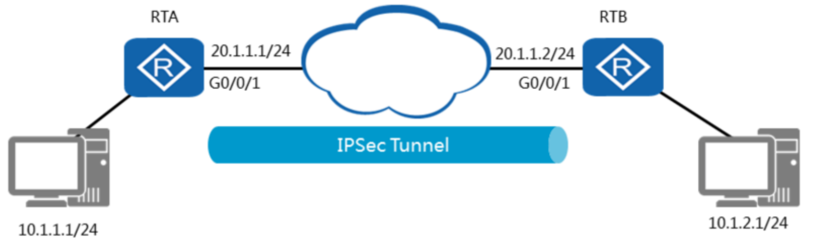
- 隧道模式中，AH或ESP头封装在原始IP报文头之前，并另外生成一个新的IP头封装到AH或ESP之前。隧道模式可以完全地对原始IP数据报进行认证和加密，而且，可以使用IPSec对等体的IP地址来隐藏客户机的IP地址。
- 隧道模式中的AH：对整个原始IP报文提供完整性检查和认证，认证功能优于ESP。但AH不提供加密功能，所以通常和ESP联合使用。
- 隧道模式中的ESP：对整个原始IP报文和ESP尾部进行加密，对ESP报文头、原始IP报文和ESP尾部进行完整性校验。
- 隧道模式中的AH+ESP：对整个原始IP报文和ESP尾部进行加密，AH、ESP分别会对不同部分进行完整性校验。



- 配置IPSec VPN的步骤如下：

1. 首先需要检查报文发送方和接收方之间的网络层可达性，确保双方只有建立IPSec VPN隧道才能进行IPSec通信。
2. 第二步是定义数据流。因为部分流量无需满足完整性和机密性要求，所以需要对流量进行过滤，选择出需要进行IPSec处理的兴趣流。可以通过配置ACL来定义和区分不同的数据流。
3. 第三步是配置IPSec安全提议。IPSec提议定义了保护数据流所用的安全协议、认证算法、加密算法和封装模式。安全协议包括AH和ESP，两者可以单独使用或一起使用。AH支持MD5和SHA-1认证算法；ESP支持两种认证算法（MD5和SHA-1）和三种加密算法（DES、3DES和AES）。为了能够正常传输数据流，安全隧道两端的对等体必须使用相同的安全协议、认证算法、加密算法和封装模式。如果要在两个安全网关之间建立IPSec隧道，建议将IPSec封装模式设置为隧道模式，以便隐藏通信使用的实际源IP地址和目的IP地址。
4. 第四步是配置IPSec安全策略。IPSec策略中会应用IPSec提议中定义的安全协议、认证算法、加密算法和封装模式。每一个IPSec安全策略都使用唯一的名称和序号来标识。IPSec策略可分成两类：手工建立SA的策略和IKE协商建立SA的策略。
5. 第五步是在一个接口上应用IPSec安全策略。

IPSec VPN 配置




```
[RTA]ip route-static 10.1.2.0 24 20.1.1.2
[RTA]acl number 3001
[RTA-acl-adv-3001]rule 5 permit ip source 10.1.1.0 0.0.0.255
destination 10.1.2.0 0.0.0.255
[RTA]ipsec proposal tran1
[RTA-ipsec-proposal-tran1]esp authentication-algorithm sha1
```


- 本示例中的IPSec VPN连接是通过配置静态路由建立的，下一跳指向RTB。需要配置两个方向的静态路由确保双向通信可达。建立一条高级ACL，用于确定哪些感兴趣流需要通过IPSec VPN隧道。高级ACL能够依据特定参数过滤流量，继而对流量执行丢弃、通过或保护操作。
- 执行**ipsec proposal**命令，可以创建IPSec提议并进入IPSec提议视图。配置IPSec策略时，必须引用IPSec提议来指定IPSec隧道两端使用的安全协议、加密算法、认证算法和封装模式。缺省情况下，使用**ipsec proposal**命令创建的IPSec提议采用ESP协议、MD5认证算法和隧道封装模式。在IPSec提议视图下执行下列命令可以修改这些参数。
- 执行**transform [ah | ah-esp | esp]**命令，可以重新配置隧道采用的安全协议。
- 执行**encapsulation-mode {transport | tunnel }**命令，可以配置报文的封装模式。
- 执行**esp authentication-algorithm [md5 | sha1 | sha2-256 | sha2-384 | sha2-512]**命令，可以配置ESP协议使用的认证算法。
- 执行**esp encryption-algorithm [des | 3des | aes-128 | aes-192 | aes-256]**命令，可以配置ESP加密算法。
- 执行**ah authentication-algorithm [md5 | sha1 | sha2-256 | sha2-384 | sha2-512]**命令，可以配置AH协议使用的认证算法。

配置验证

```
[RTA]display ipsec proposal
 Number of proposals: 1
 IPSec proposal name: tran1
 Encapsulation mode: Tunnel
 Transform          : esp-new
 ESP protocol       : Authentication SHA1-HMAC-96
                      Encryption     DES
```

- IPSec VPN对等体配置的安全提议参数必须一致。

- 执行**display ipsec proposal [name <proposal-name>]**命令，可以查看IPSec提议中配置的参数。
- Number of proposals字段显示的是已创建的IPSec提议的个数。
- IPSec proposal name字段显示的是已创建IPSec提议的名称。
- Encapsulation mode字段显示的指定提议当前使用的封装模式，其值可以为传输模式或隧道模式。
- Transform字段显示的是IPSec所采用的安全协议，其值可以是AH、ESP或AH-ESP。
- ESP protocol字段显示的是安全协议所使用的认证和加密算法。

IPSec VPN 配置

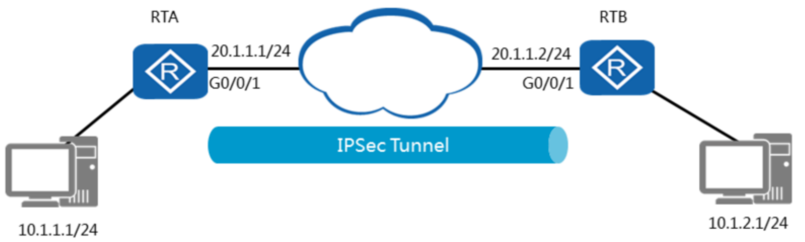
```
[RTA]ipsec  policy P1 10 manual
[RTA-ipsec-policy-manual-P1-10]security acl 3001
[RTA-ipsec-policy-manual-P1-10]proposal tran1
[RTA-ipsec-policy-manual-P1-10]tunnel remote 20.1.1.2
[RTA-ipsec-policy-manual-P1-10]tunnel local 20.1.1.1
[RTA-ipsec-policy-manual-P1-10]sa spi outbound esp 54321
[RTA-ipsec-policy-manual-P1-10]sa spi inbound esp 12345
[RTA-ipsec-policy-manual-P1-10]sa string-key outbound esp simple
huawei
[RTA-ipsec-policy-manual-P1-10]sa string-key inbound esp simple huawei
```

- 安全策略将要保护的数据流和安全提议进行绑定。

  HUAWEI

- ipsec policy policy-name seq-number命令用来创建一条IPSec策略，并进入IPSec策略视图。安全策略是由policy-name和seq-number共同来确定的，多个具有相同policy-name的安全策略组成一个安全策略组。在一个安全策略组中最多可以设置16条安全策略，而seq-number越小的安全策略，优先级越高。在一个接口上应用了一个安全策略组，实际上是同时应用了安全策略组中所有的安全策略，这样能够对不同的数据流采用不同的安全策略进行保护。
- IPSec策略除了指定策略的名称和序号外，还需要指定SA的建立方式。如果使用的是IKE协商，需要执行ipsec-policy-template命令配置指定参数。如果使用的是手工建立方式，所有参数都需要手工配置。本示例采用的是手工建立方式。
- security acl acl-number命令用来指定IPSec策略所引用的访问控制列表。
- proposal proposal-name命令用来指定IPSec策略所引用的提议。
- tunnel local { ip-address | binding-interface }命令用来配置安全隧道的本端地址。
- tunnel remote ip-address命令用来设置安全隧道的对端地址。
- sa spi { inbound | outbound } { ah | esp } spi-number命令用来设置安全联盟的安全参数索引SPI。在配置安全联盟时，入方向和出方向安全联盟的安全参数索引都必须设置，并且本端的入方向安全联盟的SPI值必须和对端的出方向安全联盟的SPI值相同，而本端的出方向安全联盟的SPI值必须和对端的入方向安全联盟的SPI值相同。
- sa string-key { inbound | outbound } { ah | esp } { simple | cipher } string-key命令用来设置安全联盟的认证密钥。入方向和出方向安全联盟的认证密钥都必须设置，并且本端的入方向安全联盟的密钥必须和对端的出方向安全联盟的密钥相同；同时，本端的出方向安全联盟密钥必须和对端的入方向安全联盟的密钥相同。

IPSec VPN配置




```
[RTA]interface GigabitEthernet 0/0/1
[RTA-GigabitEthernet0/0/1]ipsec policy P1
[RTA-GigabitEthernet0/0/1]quit
```


- **ipsec policy** *policy-name*命令用来在接口上应用指定的安全策略组。手工方式配置的安全策略只能应用到一个接口。

配置验证

```
[RTA]display ipsec policy
===========================================
IPSec policy group: "P1"
Using interface: GigabitEthernet 0/0/1
===========================================
    Sequence number: 10
    Security data flow: 3001
    Tunnel local  address: 20.1.1.1
    Tunnel remote address: 20.1.1.2
    Qos pre-classify: Disable
    Proposal name:tran1
...
```

  HUAWEI

- 执行**display ipsec policy [brief | name** *policy-name* [*seq-number*]**]**命令，可以查看指定IPSec策略或所有IPSec策略。命令的显示信息中包括：策略名称、策略序号、提议名称、ACL、隧道的本端地址和隧道的远端地址等。

配置验证

```
……
Inbound ESP setting:
      ESP SPI: 12345 (0x3039)
      ESP string-key: huawei
      ESP encryption hex key:
      ESP authentication hex key:
   Outbound ESP setting:
      ESP SPI: 54321 (0xd431)
      ESP string-key: huawei
      ESP encryption hex key:
      ESP authentication hex key:
……
```


- 执行**display ipsec policy**命令，还可以查看出方向和入方向SA相关的参数。

本章总结

- 安全联盟的作用是什么？
- IPSec VPN将会对过滤后的感兴趣数据流如何操作？

- SA（Security Association）安全联盟定义了IPSec通信对等体间将使用的数据封装模式、认证和加密算法、密钥等参数。
- 经过IPSec过滤后的感兴趣数据流将会通过SA协商的各种参数进行处理并封装，之后通过IPSec隧道转发。


谢 谢
www.huawei.com

GRE原理与配置

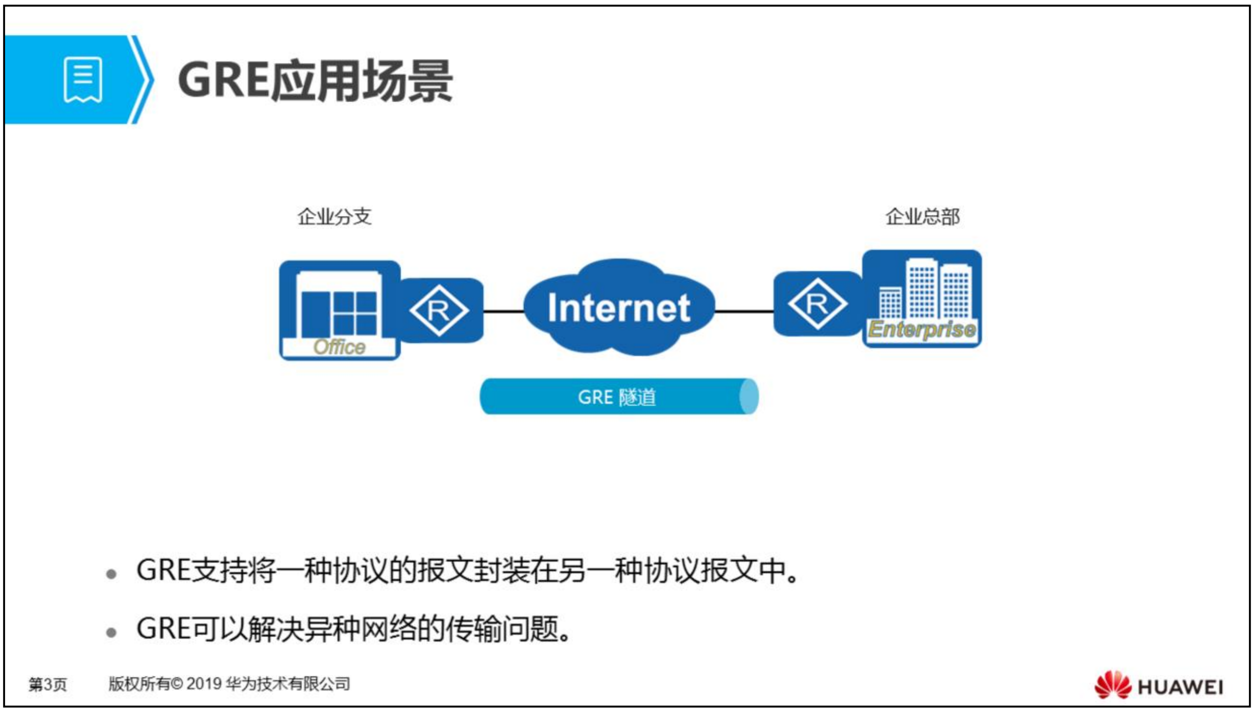
前言

- IPSec VPN用于在两个端点之间提供安全的IP通信，但只能加密并传播单播数据，无法加密和传输语音、视频、动态路由协议信息等组播数据流量。
- 通用路由封装协议GRE（Generic Routing Encapsulation）提供了将一种协议的报文封装在另一种协议报文中的机制，是一种隧道封装技术。GRE可以封装组播数据，并可以和IPSec结合使用，从而保证语音、视频等组播业务的安全。

  HUAWEI

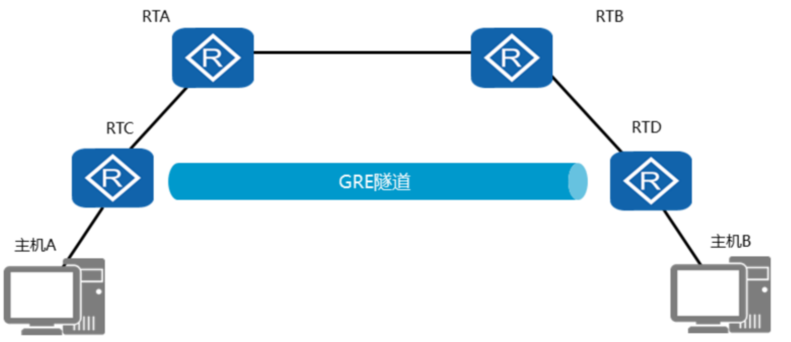
目标

- 学完本课程后，您将能够：
  - 掌握GRE的应用场景
  - 掌握GRE的工作原理
  - 掌握GRE over IPSec的配置



- GRE用来对某些网络层协议如IPX（Internet Packet Exchange）的报文进行封装，使这些被封装的报文能够在另一网络层协议（如IP）中传输。GRE可以解决异种网络的传输问题。
- IPSec VPN技术可以创建一条跨越共享公网的隧道，从而实现私网互联。IPSec VPN能够安全传输IP报文，但是无法在隧道的两个端点之间运行RIP和OSPF等路由协议。GRE可以将路由协议信息封装在另一种协议报文（例如IP）中进行传输。

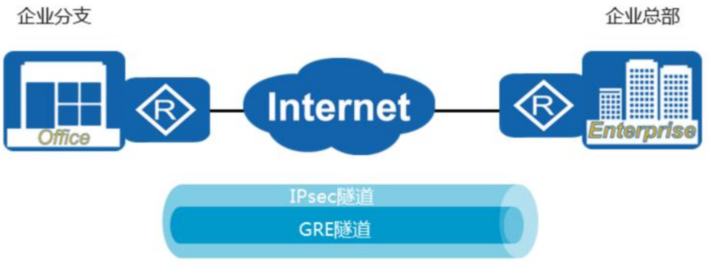
GRE应用场景



- GRE隧道扩展了受跳数限制的路由协议的工作范围，支持企业灵活设计网络拓扑。

- 使用GRE可以克服IGP协议的一些局限性。例如，RIP路由协议是一种距离矢量路由协议，最大跳数为15。如果网络直径超过15，设备将无法通信。这种情况下，可以使用GRE技术在两个网络节点之间搭建隧道，隐藏它们之间的跳数，扩大网络的工作范围。

GRE应用场景

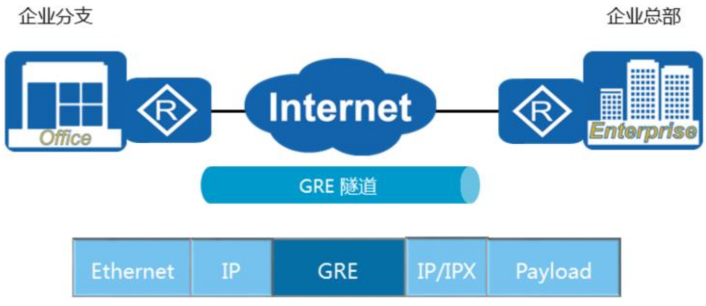


- 首先通过GRE对报文进行封装，然后再由IPSec对封装后的报文进行加密和传输。

  HUAWEI

- GRE本身并不支持加密，因而通过GRE隧道传输的流量是不加密的。将IPSec技术与GRE相结合，可以先建立GRE隧道对报文进行GRE封装，然后再建立IPSec隧道对报文进行加密，以保证报文传输的完整性和私密性。

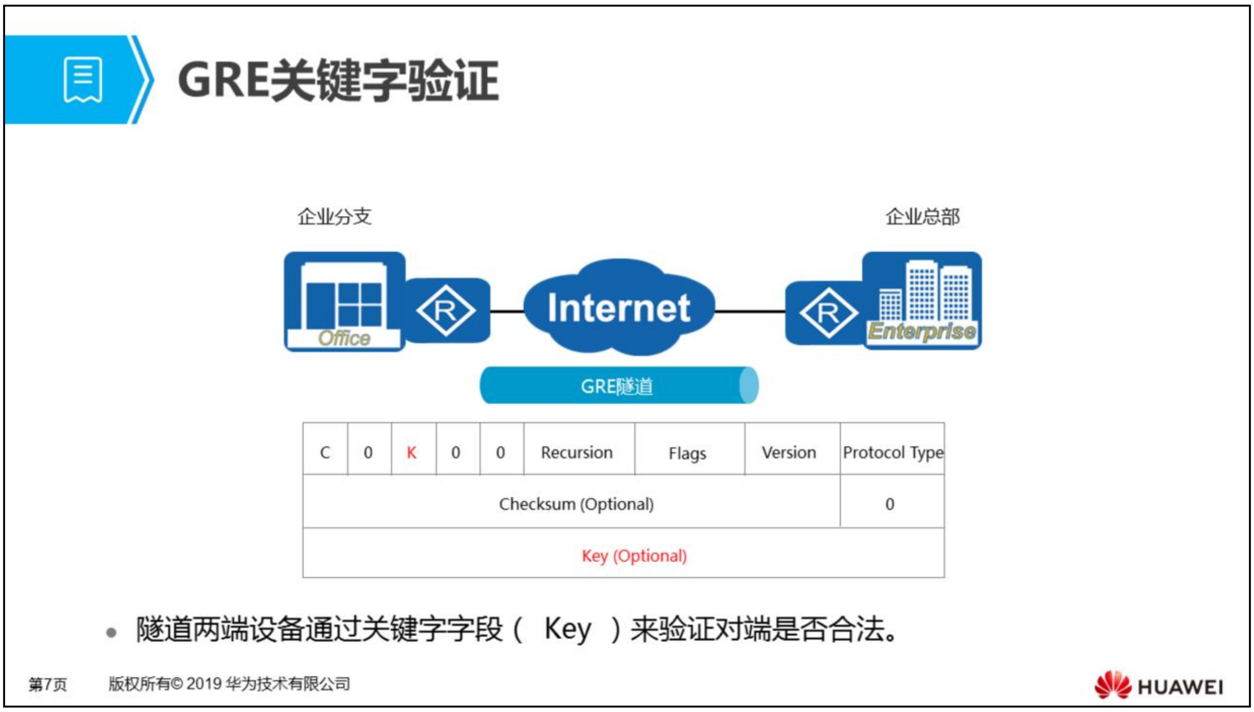
GRE报文结构



- GRE在封装数据时，会添加GRE头部信息，还会添加新的传输协议头部信息。

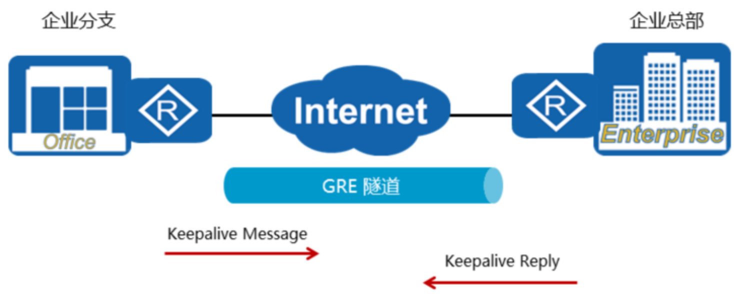
- GRE封装报文时，封装前的报文称为净荷，封装前的报文协议称为乘客协议，然后GRE会封装GRE头部，GRE成为封装协议，也叫运载协议，最后负责对封装后的报文进行转发的协议称为传输协议。
- GRE封装和解封装报文的过程如下：

1. 设备从连接私网的接口接收到报文后，检查报文头中的目的IP地址字段，在路由表查找出接口，如果发现出接口是隧道接口，则将报文发送给隧道模块进行处理。
2. 隧道模块接收到报文后首先根据乘客协议的类型和当前GRE隧道配置的校验和参数，对报文进行GRE封装，即添加GRE报文头。
3. 然后，设备给报文添加传输协议报文头，即IP报文头。该IP报文头的源地址就是隧道源地址，目的地址就是隧道目的地址。
4. 最后，设备根据新添加的IP报文头目的地址，在路由表中查找相应的出接口，并发送报文。之后，封装后的报文将在公网中传输。
5. 接收端设备从连接公网的接口收到报文后，首先分析IP报文头，如果发现协议类型字段的值为47，表示协议为GRE，于是出接口将报文交给GRE模块处理。GRE模块去掉IP报文头和GRE报文头，并根据GRE报文头的协议类型字段，发现此报文的乘客协议为私网中运行的协议，于是将报文交给该协议处理。



- 关键字（Key）验证是指对隧道接口进行校验，这种安全机制可以防止错误接收到来自其他设备的报文。关键字字段是一个四字节长的数值，若GRE报文头中的K位为1，则在GRE报文头中会插入关键字字段。只有隧道两端设置的关键字完全一致时才能通过验证，否则报文将被丢弃。

Keepalive检测



- Keepalive检测功能用于检测隧道对端是否可达。

- Keepalive检测功能用于在任意时刻检测隧道链路是否处于Keepalive状态，即检测隧道对端是否可达。如果对端不可达，隧道连接就会及时关闭，避免形成数据空洞。使能Keepalive检测功能后，GRE隧道本端会定期向对端发送Keepalive探测报文。若对端可达，则本端会收到对端的回应报文；若对端不可达，则收不到对端的回应报文。如果在隧道一端配置了Keepalive功能，无论对端是否配置Keepalive，配置的Keepalive功能在该端都生效。隧道对端收到Keepalive探测报文，无论是否配置Keepalive，都会给源端发送一个回应报文。
- 使能Keepalive检测功能后，GRE隧道的源端会创建一个计数器，并周期性地发送Keepalive探测报文，同时进行不可达计数。每发送一个探测报文，不可达计数加1。
- 如果源端在计数器值达到预先设置的值之前收到回应报文，则表明对端可达。如果计数器值达到预先设置的重试次数，源端还是没有收到回应报文，则认为对端不可达。此时，源端将关闭隧道连接。

GRE配置


```
[RTA]interface Tunnel 0/0/1
[RTA-Tunnel0/0/1]ip address 40.1.1.1 24
[RTA-Tunnel0/0/1]tunnel-protocol gre
[RTA-Tunnel0/0/1]source 20.1.1.1
[RTA-Tunnel0/0/1]destination 20.1.1.2
[RTA-Tunnel0/0/1]quit
[RTA]ip route-static 10.1.2.0 24 Tunnel 0/0/1
```


- **interface tunnel** *interface-number*命令用来创建Tunnel接口。创建Tunnel接口后，需要配置Tunnel接口的IP地址和Tunnel接口的封装协议。
- **tunnel-protocol**命令用来配置Tunnel接口的隧道协议。
- **source** { *source-ip-address* | *interface-type interface-number* }命令用来配置Tunnel源地址或源接口。
- **destination** *dest-ip-address*命令用来指定Tunnel接口的目的IP地址。
- 在本端设备和远端设备上还必须存在经过Tunnel转发的路由，这样，需要进行GRE封装的报文才能正确转发。经过Tunnel接口转发的路由可以是静态路由，也可以是动态路由。配置静态路由时，路由的目的地址是GRE封装前原始报文的目的地址，出接口是本端Tunnel接口。

配置验证

```
[RTA]display interface Tunnel 0/0/1
Tunnel0/0/1 current state : UP
Line protocol current state : UP
Last line protocol up time : 2016-03-21 13:37:38
Description:HUAWEI, AR Series, Tunnel0/0/1 Interface
Route Port, The Maximum Transmit Unit is 1476
Internet Address is 40.1.1.1/24
Encapsulation is TUNNEL, loopback not set
Tunnel source 20.1.1.1 (GigabitEthernet0/0/1), destination 20.1.1.2
Tunnel protocol/transport GRE/IP, key disabled
keepalive disabled
Checksumming of packets disabled
……
```


- 执行**display interface Tunnel 0/0/1**命令，可以查看接口的运行状态和路由信息。如果接口的当前状态和链路层协议的状态均显示为UP，则接口处于正常转发状态。隧道的源地址和目的地址分别为建立GRE隧道使用的物理接口的IP地址。

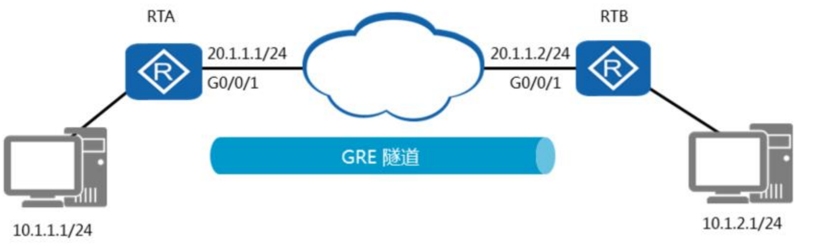
配置验证

```
[RTA]display ip routing-table
Route Flags: R - relay, D - download to fib
------------------------------------------------------------
Routing Tables: Public  Destinations : 13       Routes : 14
Destination/Mask Proto  Pre Cost Flags  NextHop   Interface
……
10.1.2.0/24      Static 60   0    RD     40.1.1.2  Tunnel 0/0/1
……
```


HUAWEI

- 执行**display ip routing table**命令，可以查看IP路由表，判断GRE隧道连接的两个网络的可达信息。在本示例中，可以看出目的地址为通过GRE隧道可达的网络地址，下一跳地址为GRE隧道远端接口的IP地址。

配置Keepalive检测




```
[RTA]interface Tunnel 0/0/1
[RTA-Tunnel0/0/1]keepalive period 3
[RTA-Tunnel0/0/1]quit
```

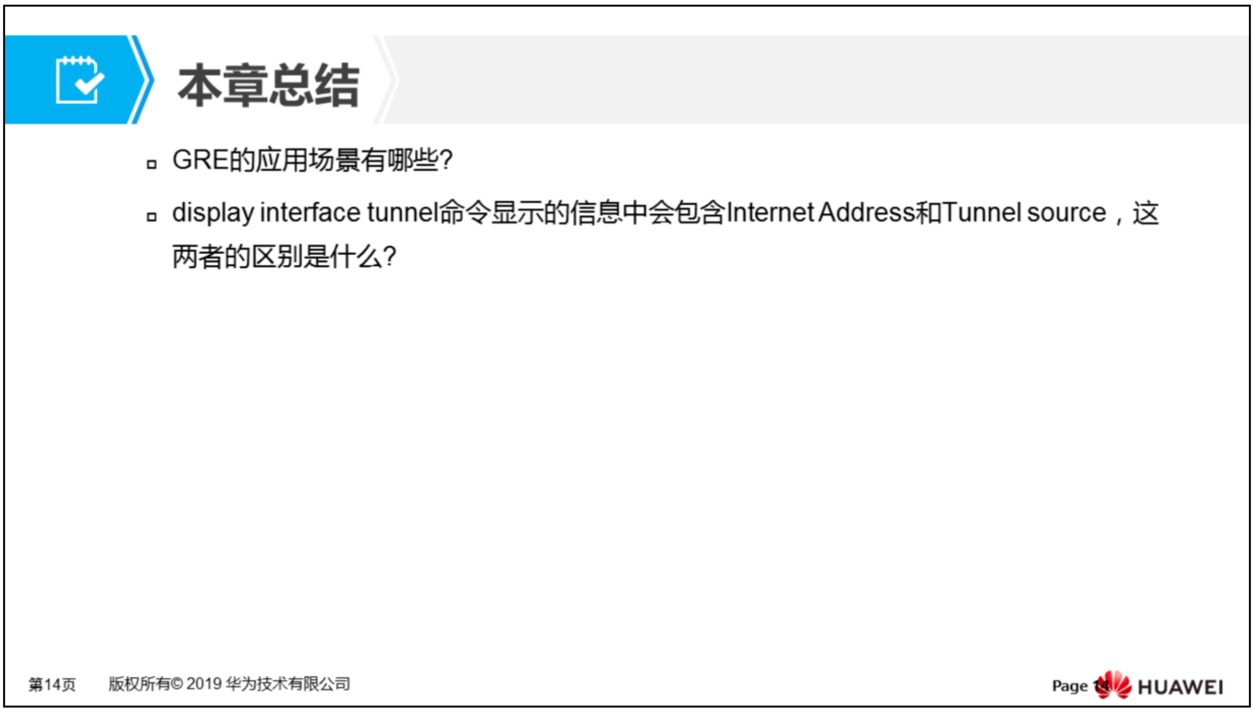

- 执行**keepalive** [**period** *period* [**retry-times** *retry-times*]] 命令，可以在GRE隧道接口启用Keepalive检测功能。其中，**period**参数指定Keepalive检测报文的发送周期，默认值为5秒；**retry-times**参数指定Keepalive检测报文的重传次数，默认值为3。如果在指定的重传次数内未收到对端的回应报文，则认为隧道两端通信失败，GRE隧道将被拆除。

配置验证

```
[RTA]display interface Tunnel 0/0/1
Tunnel0/0/1 current state : UP
Line protocol current state : DOWN
Description:HUAWEI, AR Series, Tunnel0/0/1 Interface
Route Port, The Maximum Transmit Unit is 1476
Internet Address is 40.1.1.1/24
Encapsulation is TUNNEL, loopback not set
Tunnel source 20.1.1.1 (GigabitEthernet0/0/1), destination 20.1.1.2
Tunnel protocol/transport GRE/IP, key disabled
keepalive enable period 3 retry-times 3
Checksumming of packets disabled
……
```

  HUAWEI

- 执行display interface tunnel命令，还可以查看GRE的Keepalive功能是否使能。本示例中，Keepalive检测功能的当前状态显示为启用，且报文的发送周期为3秒，重传次数为3次。



- GRE可以解决异种网络的传输问题；GRE隧道扩展了受跳数限制的路由协议的工作范围，支持企业灵活设计网络拓扑；GRE可以与IPSec结合来实现加密传输组播数据。
- Internet Address代表建立GRE隧道所用的虚拟隧道地址，Tunnel source表示隧道的起点，是设备的出接口物理地址。


谢 谢
www.huawei.com

SNMP原理与配置

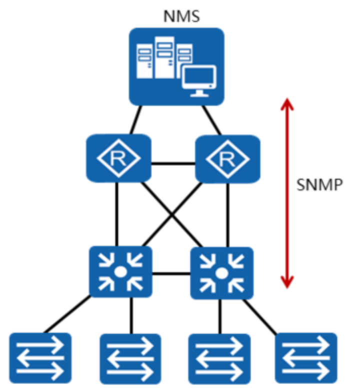
前言

- 随着网络技术的飞速发展，企业中网络设备的数量成几何级数增长，网络设备的种类也越来越多，这使得企业网络的管理变得十分复杂。
- 简单网络管理协议SNMP（ Simple Network Management Protocol ）可以实现对不同种类和不同厂商的网络设备进行统一管理，大大提升了网络管理的效率。

目标

- 学完本课程后，您将能够：
  - 掌握SNMP的基本概念
  - 掌握SNMP的基本配置

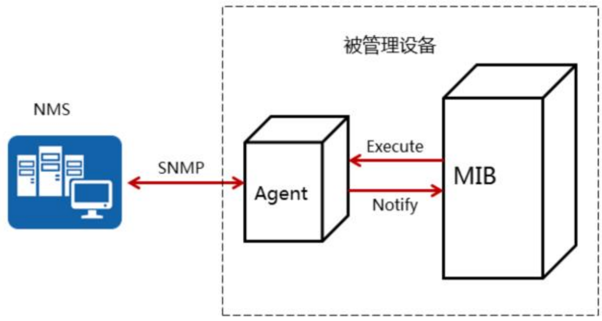
SNMP应用场景



- SNMP用来在网络管理系统NMS和被管理设备之间传输管理信息。

- SNMP是广泛应用于TCP/IP网络的一种网络管理协议。SNMP提供了一种通过运行网络管理软件NMS（Network Management System）的网络管理工作站来管理网络设备的方法。
- SNMP支持以下几种操作：
- NMS通过SNMP协议给网络设备发送配置信息。
- NMS通过SNMP来查询和获取网络中的资源信息。
- 网络设备主动向NMS上报告警消息，使得网络管理员能够及时处理各种网络问题。

SNMP架构



- SNMP包括NMS，Agent和MIB等。
- Agent是被管理设备中的一个代理进程。
- MIB是一个数据库，它包含了被管理设备所维护的变量。

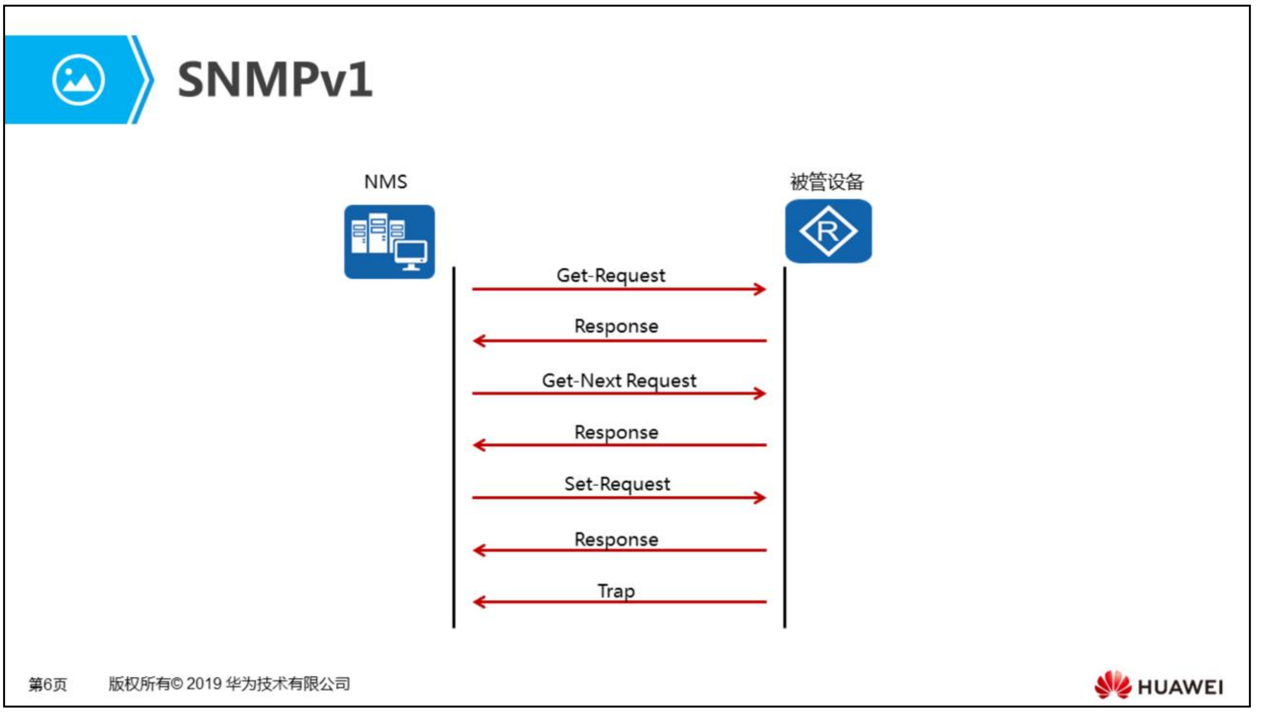
  HUAWEI

- NMS是运行在网管主机上的网络管理软件。网络管理员通过操作NMS，向被管理设备发出请求，从而可以监控和配置网络设备。
- Agent是运行在被管理设备上的代理进程。被管理设备在接收到NMS发出的请求后，由Agent作出响应操作。Agent的主要功能包括：收集设备状态信息、实现NMS对设备的远程操作、向NMS发送告警消息。
- 管理信息库MIB（Management Information Base）是一个虚拟的数据库，是在被管理设备端维护的设备状态信息集。Agent通过查找MIB来收集设备状态信息。

SNMP版本

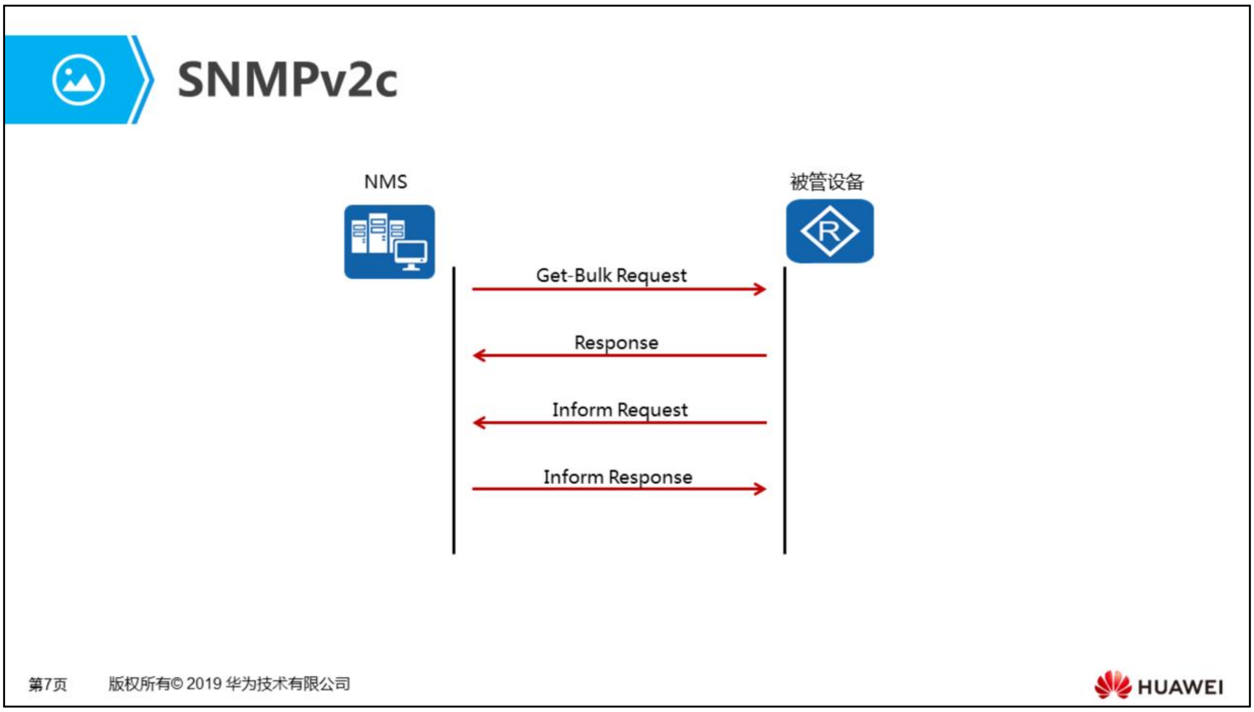
版本	描述
SNMPv1	实现方便，安全性弱。
SNMPv2c	有一定的安全性。 现在应用最为广泛。
SNMPv3	定义了一种管理框架，为用户提供了安全的访问机制。

- SNMPv1：网管端工作站上的NMS与被管理设备上的Agent之间，通过交互SNMPv1报文，可以实现网管端对被管理设备的管理。SNMPv1基本上没有什么安全性可言。
- SNMPv2c在继承SNMPv1的基础上，其性能、安全性、机密性等方面都有了大的改进。
- SNMPv3是在SNMPv2基础之上增加、完善了安全和管理机制。SNMPv3体系结构体现了模块化的设计思想，使管理者可以方便灵活地实现功能的增加和修改。SNMPv3的主要特点在于适应性强，可适用于多种操作环境，它不仅可以管理最简单的网络，实现基本的管理功能，也可以提供强大的网络管理功能，满足复杂网络的管理需求。



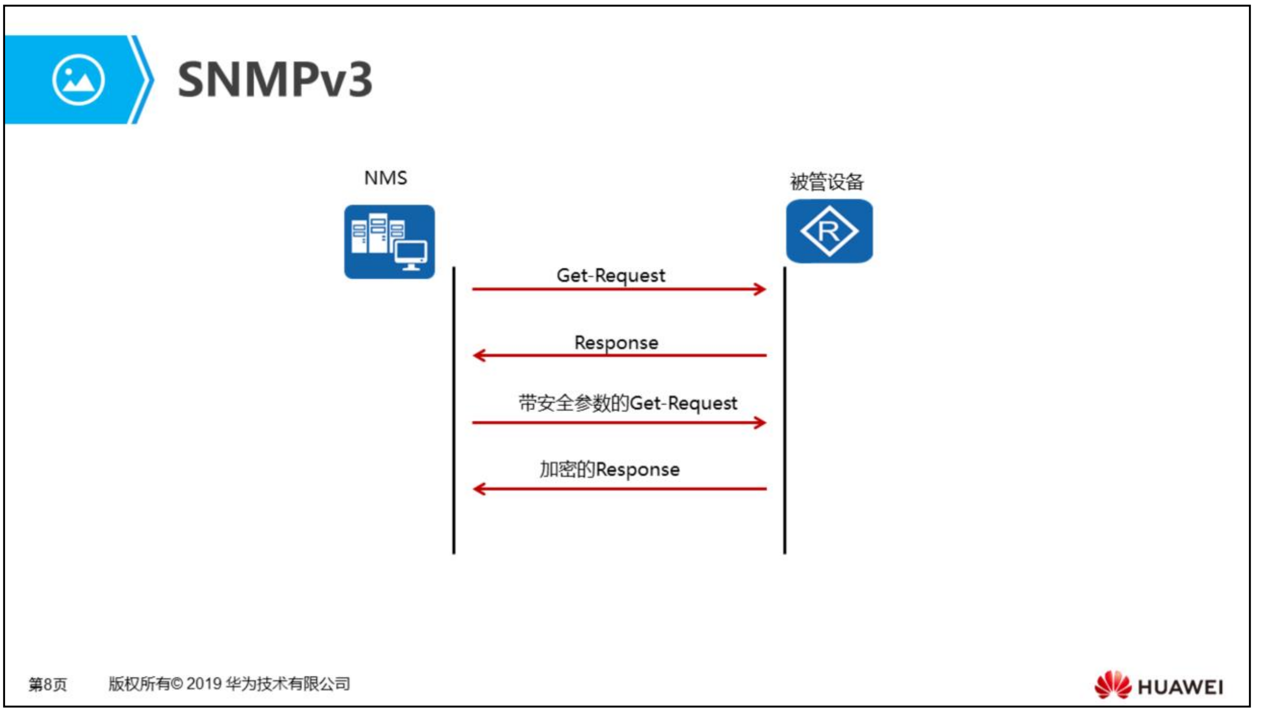
- SNMPv1定义了5种协议操作：

1. Get-Request：NMS从代理进程的MIB中提取一个或多个参数值。
2. Get-Next-Request：NMS从代理进程的MIB中按照字典式排序提取下一个参数值。
3. Set-Request：NMS设置代理进程MIB中的一个或多个参数值。
4. Response：代理进程返回一个或多个参数值。它是前三种操作的响应操作。
5. Trap：代理进程主动向NMS发送报文，告知设备上发生的紧急或重要事件。



- SNMPv2c新增了2种协议操作：

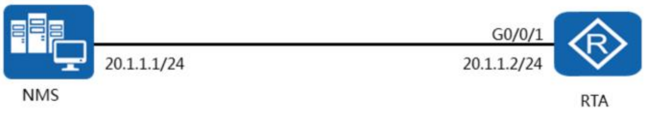
1. GetBulk：相当于连续执行多次GetNext操作。在NMS上可以设置被管理设备在一次GetBulk报文交互时，执行GetNext操作的次数。
2. Inform：被管理设备向NMS主动发送告警。与trap告警不同的是，被管理设备发送Inform告警后，需要NMS进行接收确认。如果被管设备没有收到确认信息则会将告警暂时保存在Inform缓存中，并且会重复发送该告警，直到NMS确认收到了该告警或者发送次数已经达到了最大重传次数。



- SNMPv3的实现原理和SNMPv1/SNMPv2c基本一致，主要的区别是SNMPv3增加了身份验证和加密处理。

1. NMS向Agent发送不带安全参数的Get请求报文，向Agent获取安全参数等信息。
2. Agent响应NMS的请求，向NMS反馈所请求的参数。
3. NMS向Agent发送带安全参数的Get请求报文。
4. Agent对NMS发送的请求消息进行认证，认证通过后对消息进行解密，解密成功后，向NMS发送加密的响应。

SNMP配置




```
[RTA]snmp-agent
[RTA]snmp-agent sys-info version v2c
[RTA]snmp-agent trap enable
[RTA]snmp-agent trap source GigabitEthernet0/0/1
```

  HUAWEI

- **snmp-agent**命令用来使能SNMP代理。
- 执行**snmp-agent sys-info version [[v1 | v2c | v3] * | all]**命令可以配置SNMP系统信息，其中**version [[v1 | v2c | v3] * | all]**指定设备运行的SNMP版本。缺省情况下，ARG3系列路由器支持SNMPv1，SNMPv2c，SNMPv3版本。
- 执行**snmp-agent trap enable**命令，可以激活代理向NMS发送告警消息的功能，这一功能激活后，设备将向NMS上报任何异常事件。另外，还需要指定发送告警通告的接口，本示例中指定的是与NMS相连的GigabitEthernet 0/0/1接口。

配置验证

```
[RTA]display snmp-agent sys-info
  The contact person for this managed node:
          R&D Shenzhen, Huawei Technologies Co., Ltd.

  The physical location of this node:
          Shenzhen China

  SNMP version running in the system:
          SNMPv2c
```

  HUAWEI

- 执行display snmp-agent sys-info命令，可以查看系统维护的相关信息，包括设备的物理位置和SNMP版本。

本章总结

- 配置SNMP时，默认的版本号是多少?
- 代理进程Agent发送trap信息给NMS时，目的端口号是多少?

HUAWEI

- 华为ARG3系列路由器默认使能SNMP的所有版本（SNMPv1、SNMPv2c和SNMPv3）。
- 代理进程使用UDP协议向NMS发送告警消息，目的端口号为162。


谢 谢
www.huawei.com

IPv6基础介绍

前言

- 随着Internet规模的扩大，IPv4地址空间已经消耗殆尽。针对IPv4的地址短缺问题，曾先后出现过CIDR和NAT等临时性解决方案，但是CIDR和NAT都有各自的弊端，并不能作为IPv4地址短缺问题的彻底解决方案。另外，安全性、QoS(服务质量)、简便配置等要求也表明需要一个新的协议来根本解决目前IPv4面临的问题。
- IETF在20世纪90年代提出了下一代互联网协议 - IPv6，IPv6支持几乎无限的地址空间。IPv6使用了全新的地址配置方式，使得配置更加简单。IPv6还采用了全新的报文格式，提高了报文处理的效率、安全性，也能更好的支持QoS。

目标

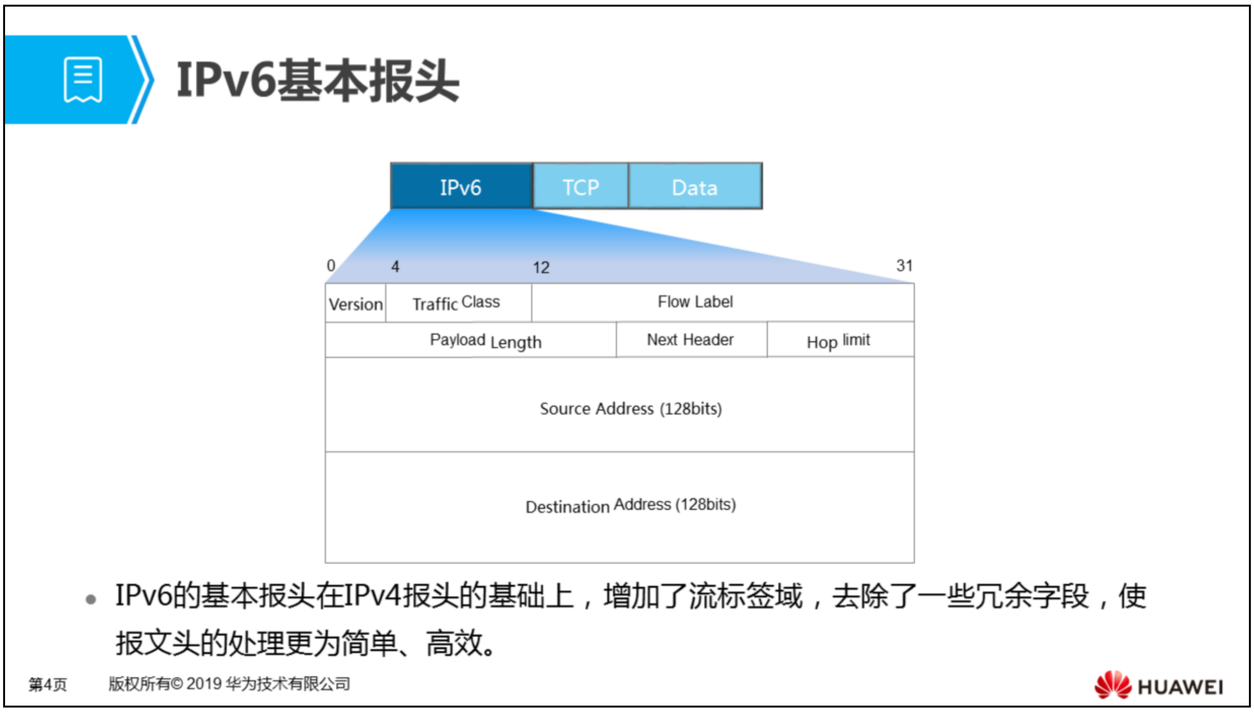
- 学完本课程后，您将能够：
  - 掌握IPv6的基本概念
  - 掌握IPv6地址格式和地址类型
  - 掌握IPv6无状态地址配置的过程

IPv6地址

版本	长度	地址数量
IPv4	32 bit	4,294,967,296
IPv6	128 bit	340,282,366,920,938,463,463,374,607,431,768,211,456

- IPv4地址空间已经消耗殆尽，近乎无限的地址空间是IPv6的最大优势。

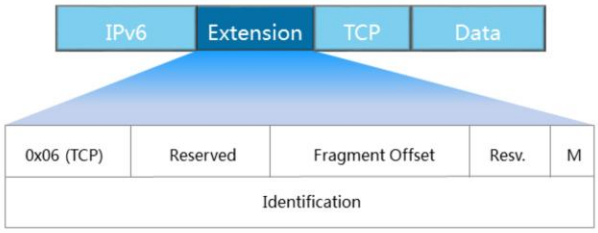
- 在因特网发展初期，IPv4以其协议简单、易于实现、互操作性好的优势而得到快速发展。然而，随着因特网的迅猛发展，IPv4地址不足等设计缺陷也日益明显。IPv4理论上仅仅能够提供的地址数量是43亿，但是由于地址分配机制等原因，实际可使用的数量还远远达不到43亿。因特网的迅猛发展令人始料未及，同时也带来了地址短缺的问题。针对这一问题，曾先后出现过几种解决方案，比如CIDR和NAT。但是CIDR和NAT都有各自的弊端和不能解决的问题，在这样的情况下，IPv6的应用和推广便显得越来越急迫。
- IPv6是Internet工程任务组（IETF）设计的一套规范，它是网络层协议的第二代标准协议，也是IPv4（Internet Protocol Version 4）的升级版本。IPv6与IPv4的最显著区别是，IPv4地址采用32比特标识，而IPv6地址采用128比特标识。128比特的IPv6地址可以划分更多地址层级、拥有更广阔的地址分配空间，并支持地址自动配置。



- IPv6报文由IPv6基本报头、IPv6扩展报头以及上层协议数据单元三部分组成。
- 基本报头中的各字段解释如下：
- Version：版本号，长度为4bit。对于IPv6，该值为6。
- Traffic Class：流类别，长度为8bit，它等同于IPv4报头中的TOS字段，表示IPv6数据报文的类或优先级，主要应用于QoS。
- Flow Label：流标签，长度为20bit，它用于区分实时流量。流可以理解为特定应用或进程的来自某一源地址发往一个或多个目的地址的连续单播、组播或任播报文。IPv6中的流标签字段、源地址字段和目的地址字段一起为特定数据流指定了网络中的转发路径。这样，报文在IP网络中传输时会保持原有的顺序，提高了处理效率。随着三网合一的发展趋势，IP网络不仅要求能够传输传统的数据报文，还需要能够传输语音、视频等报文。这种情况下，流标签字段的作用就显得更加重要。
- Payload Length：有效载荷长度，长度为16bit，它是指紧跟IPv6报头的数据报文的其它部分。
- Next Header：下一个报头，长度为8bit。该字段定义了紧跟在IPv6报头后面的第一个扩展报头（如果存在）的类型。
- 跳数限制（Hop Limit），长度为8bit，该字段类似于IPv4报头中的Time to Live字段，它定义了IP数据报文所能经过的最大跳数。每经过一个路由器，该数值减去1；当该字段的值为0时，数据报文将被丢弃。
- Source Address：源地址，长度为128bit，表示发送方的地址。
- Destination Address：目的地址，长度为128bit，表示接收方的地址。
- 与IPv4相比，IPv6报头去除了IHL、Identifier、Flags、Fragment Offset、Header Checksum、 Options、Padding域，只增了流标签域，因此IPv6报文头的处理较IPv4大大简

化，提高了处理效率。另外，IPv6为了更好支持各种选项处理，提出了扩展头的概念。

IPv6扩展报头



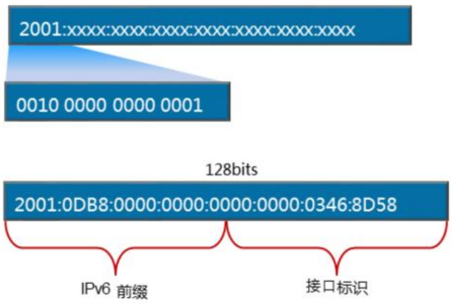
- IPv6扩展报头是跟在IPv6基本报头后面的可选报头，可以有一个或多个。
- 如图所示的扩展报头是分片扩展报头。

- IPv6增加了扩展报头，使得IPv6报头更加简化。一个IPv6报文可以包含0个、1个或多个扩展报头，仅当需要路由器或目的节点做某些特殊处理时，才由发送方添加一个或多个扩展头。IPv6支持多个扩展报头，各扩展报头中都含有一个下一个报头字段，用于指明下一个扩展报头的类型。这些报头必须按照以下顺序出现：

1. IPv6基本报头
2. 逐跳选项扩展报头
3. 目的选项扩展报头
4. 路由扩展报头
5. 分片扩展报头
6. 认证扩展报头
7. 封装安全有效载荷扩展报头
8. 目的选项扩展报头（指那些将被分组报文的最终目的地处理的选项）
9. 上层协议数据报文

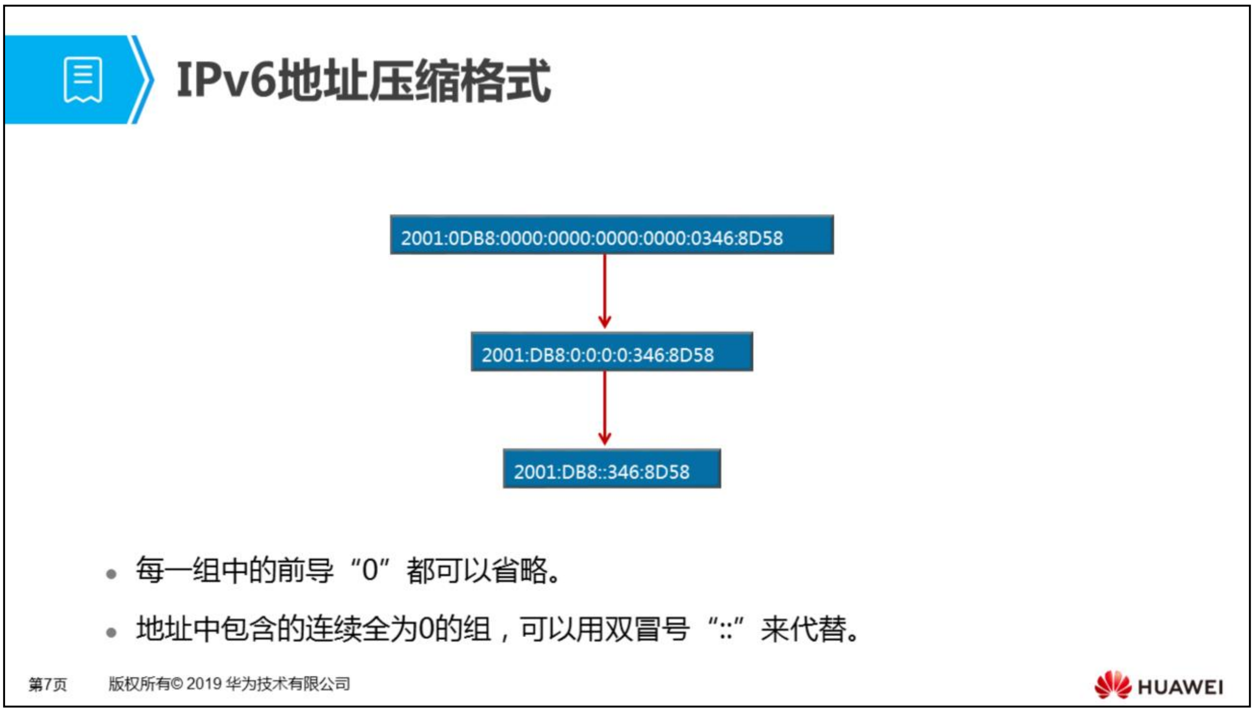
- 除了目的选项扩展报头外，每个扩展报头在一个报文中最多只能出现一次。目的选项扩展报头在一个报文中最多也只能出现两次，一次是在路由扩展报头之前，另一次是在上层协议扩展报头之前。

IPv6地址格式



- IPv6地址长度为128比特，每16比特划分为一段，每段由4个十六进制数表示，并用冒号隔开。
- IPv6地址包括网络前缀和接口标识两部分。

- IPv6地址长度为128比特，用于标识一个或一组接口。IPv6地址通常写作xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx，其中xxxx是4个十六进制数，等同于16个二进制数；八组xxxx共同组成了一个128比特的IPv6地址。一个IPv6地址由IPv6地址前缀和接口ID组成，IPv6地址前缀用来标识IPv6网络，接口ID用来标识接口。



- 由于IPv6地址长度为128比特，书写时会非常不方便。此外，IPv6地址的巨大地址空间使得地址中往往会包含多个0。为了应对这种情况，IPv6提供了压缩方式来简化地址的书写。压缩规则如下：
- 每16比特组中的前导0可以省略。
- 地址中包含的连续两个或多个均为0的组，可以用双冒号“::”来代替。需要注意的是，在一个IPv6地址中只能使用一次双冒号“::”，否则，设备将压缩后的地址恢复成128位时，无法确定每段中0的个数。
- 本示例展示了如何利用压缩规则对IPv6地址进行简化表示。

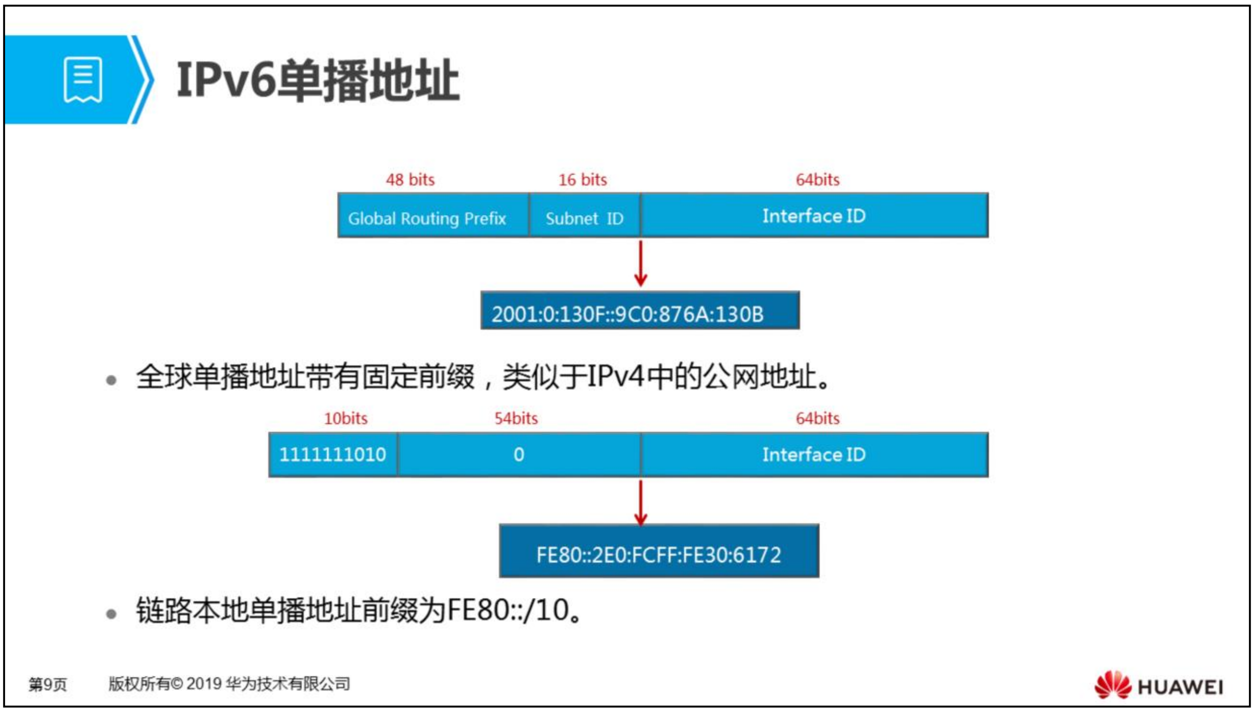
IPv6地址分类

地址范围	描述
2000::/3	全球单播地址
FC00::/7	唯一本地地址
FE80::/10	链路本地地址
FF00::/8	组播地址
::/128	未指定地址
::1/128	环回地址

- IPv6地址分为单播地址、任播地址、组播地址三种类型。

  HUAWEI

- 目前，IPv6地址空间中还有很多地址尚未分配。这一方面是因为IPv6有着巨大的地址空间，足够在未来很多年使用，另一方面是因为寻址方案还有待发展，同时关于地址类型的适用范围也多有值得商榷的地方。
- 目前，有一小部分全球单播地址已经由IANA（互联网名称与数字地址分配机构ICANN的一个分支）分配给了用户。单播地址的格式是2000::/3，代表公共IP网络上任意可及的地址。IANA负责将该段地址范围内的地址分配给多个区域互联网注册管理机构（RIR）。RIR负责全球5个区域的地址分配。以下几个地址范围已经分配：2400::/12 (APNIC)、2600::/12 (ARIN)、2800::/12 (LACNIC)、2A00::/12 (RIPE NCC)和2C00::/12 (AfriNIC)，它们使用单一地址前缀标识特定区域中的所有地址。2000::/3地址范围中还为文档示例预留了地址空间，例如2001:0DB8::/32。
- 链路本地地址只能在连接到同一本地链路的节点之间使用。可以在自动地址分配、邻居发现和链路上没有路由器的情况下使用链路本地地址。以链路本地地址为源地址或目的地址的IPv6报文不会被路由器转发到其他链路。链路本地地址的前缀是FE80::/10。
- 组播地址的前缀是FF00::/8。组播地址范围内的大部分地址都是为特定组播组保留的。跟IPv4一样，IPv6组播地址还支持路由协议。IPv6中没有广播地址。组播地址替代广播地址可以确保报文只发送给特定的组播组而不是IPv6网络中的任意终端。
- IPv6还包括一些特殊地址，比如未指定地址::/128。如果没有给一个接口分配IP地址，该接口的地址则为::/128。需要注意的是，不能将未指定地址跟默认IP地址::/0相混淆。默认IP地址::/0跟IPv4中的默认地址0.0.0.0/0类似。环回地址127.0.0.1在IPv6中被定义为保留地址::1/128。



- 单播地址主要包含全球单播地址和链路本地地址。全球单播地址（例如2000::/3）带有固定的地址前缀，即前三位为固定值001。其地址结构是一个三层结构，依次为全球路由前缀、子网标识和接口标识。全球路由前缀由RIR和互联网服务供应商（ISP）组成，RIR为ISP分配IP地址前缀。子网标识定义了网络的管理子网。
- 链路本地单播地址的前缀为FE80::/10，表示地址最高10位值为1111111010。前缀后面紧跟的64位是接口标识，这64位已足够主机接口使用，因而链路本地单播地址的剩余54位为0。本示例展示了上述两种单播地址类型。

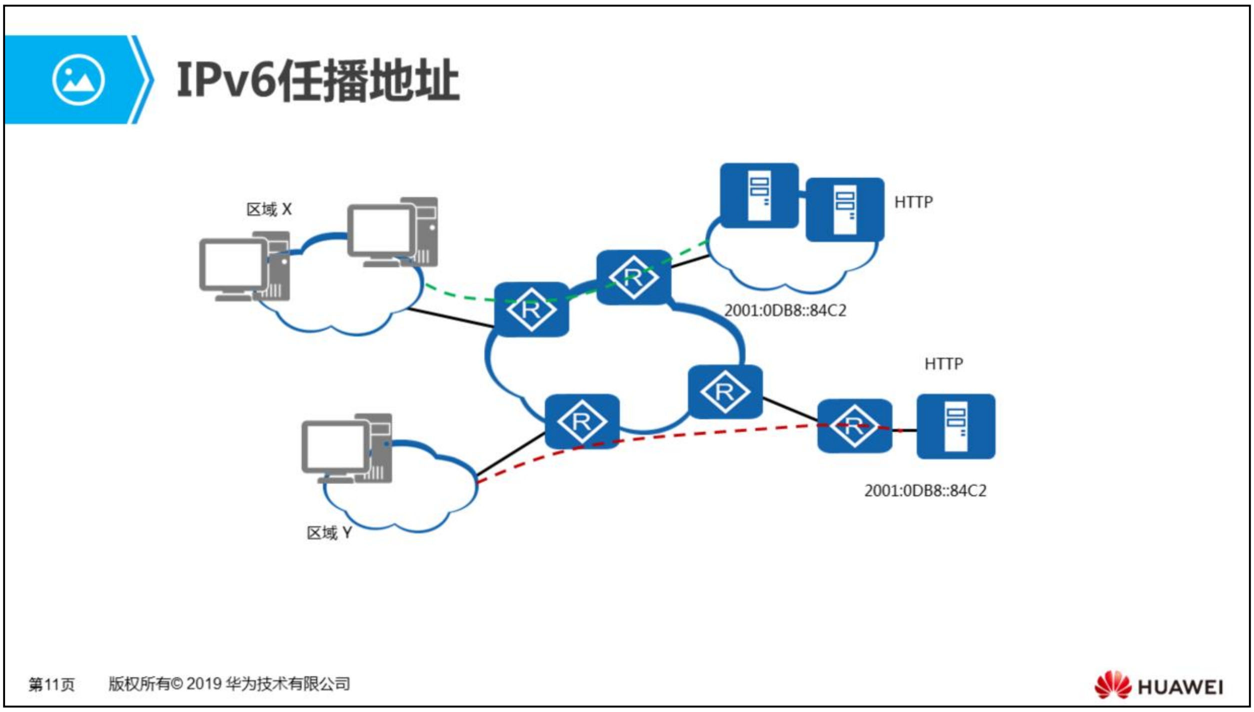
IPv6组播地址

8bits	4bits	4bits	112bits
11111111	Flags	Scope	Group ID

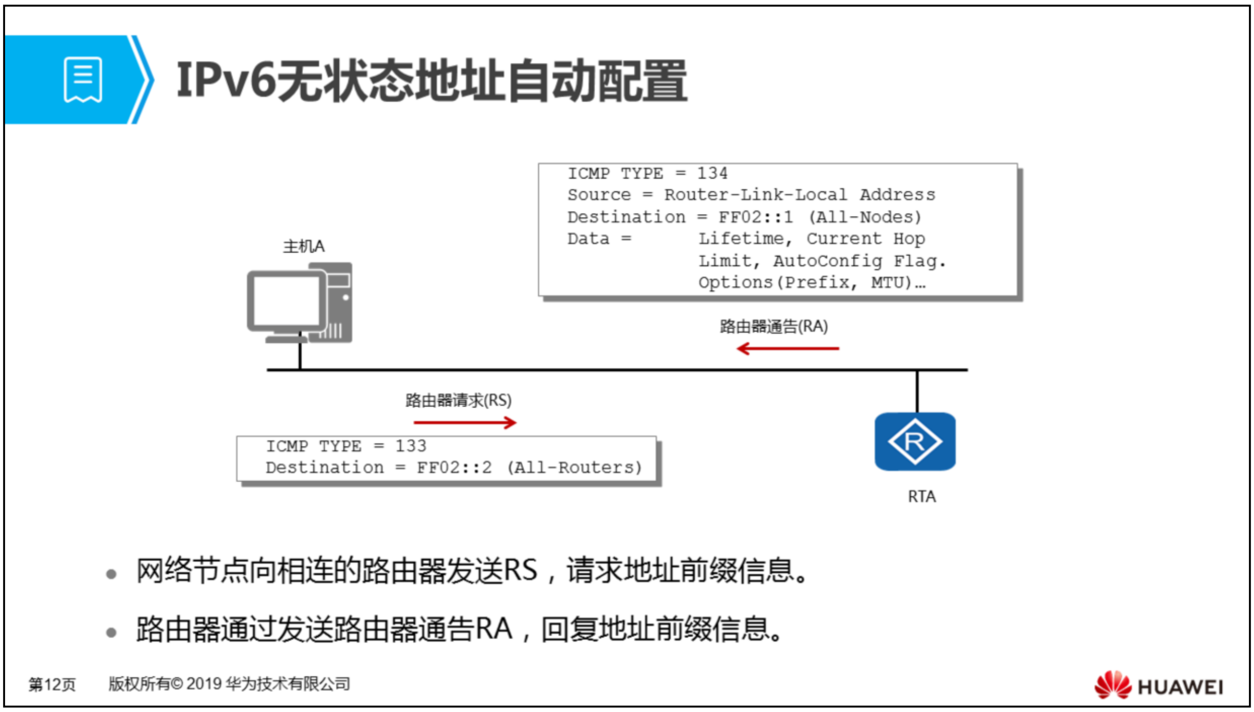
地址范围	描述
FF02::1	链路本地范围所有节点
FF02::2	链路本地范围所有路由器

- 所有IPv6组播地址都以FF开始。
- IPv6为需要使用组播发送数据的协议预留了一些组播组。

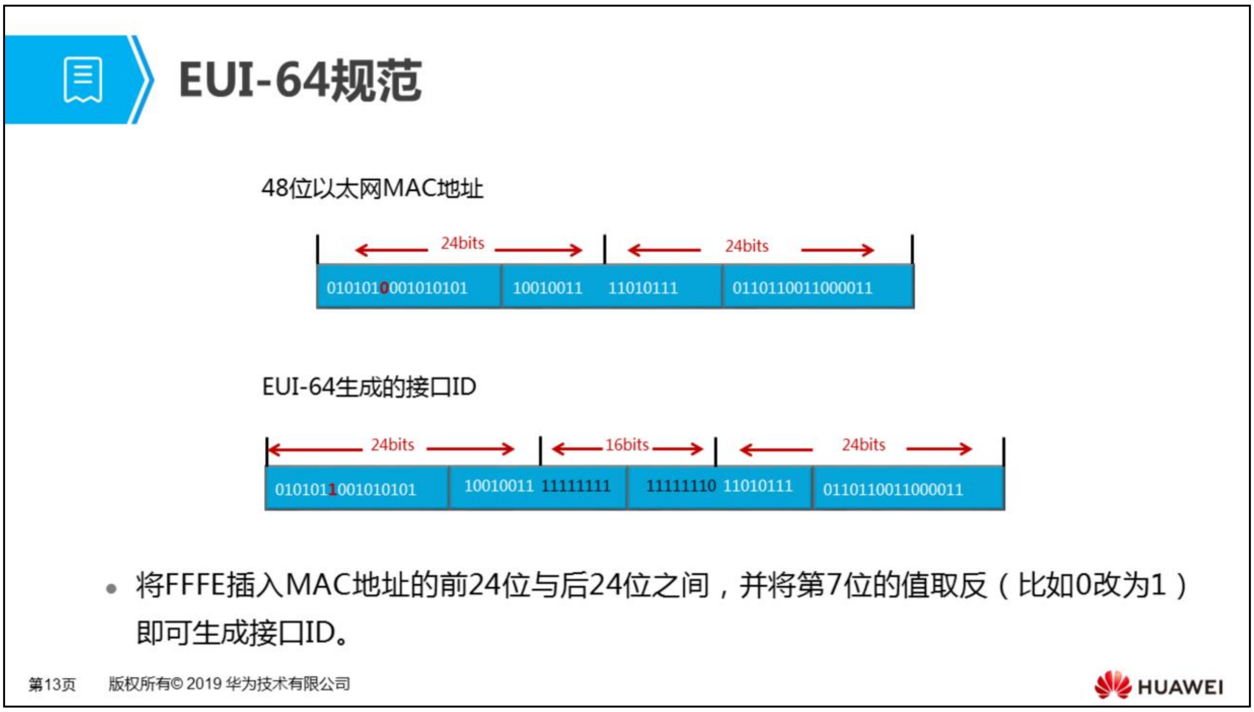
- IPv6的组播与IPv4相同，用来标识一组接口，一般这些接口属于不同的节点。一个节点可能属于0到多个组播组。目的地址为组播地址的报文会被该组播地址标识的所有接口接收。
- 一个IPv6组播地址是由前缀、标志（Flag）字段、范围（Scope）字段以及组播组ID（Group ID）4个部分组成：
- 前缀：IPv6组播地址的前缀是FF00::/8（1111 1111）。
- 标志字段（Flag）：长度4bit，目前只使用了最后一个比特（前三位必须置0），当该位值为0时，表示当前的组播地址是由IANA所分配的一个永久分配地址；当该值为1时，表示当前的组播地址是一个临时组播地址（非永久分配地址）。
- 范围字段（Scope）：长度4bit，用来限制组播数据流在网络中发送的范围。
- 组播组ID（Group ID）：长度112bit，用以标识组播组。目前，RFC2373并没有将所有的112位都定义成组标识，而是建议仅使用该112位的最低32位作为组播组ID，将剩余的80位都置0，这样，每个组播组ID都可以映射到一个唯一的以太网组播MAC地址（RFC2464）。



- 任播地址标识一组网络接口（通常属于不同的节点）。目标地址是任播地址的数据包将发送给其中路由意义上最近的一个网络接口。任播过程涉及一个任播报文发起方和一个或多个响应方。任播报文的发起方通常为请求某一服务（DNS查找）的主机或请求返还特定数据（例如，HTTP网页信息）的主机。任播地址与单播地址在格式上无任何差异，唯一的区别是一台设备可以给多台具有相同地址的设备发送报文。
- 企业网络中运用任播地址有很多优势。其中一个优势是业务冗余。比如，用户可以通过多台使用相同地址的服务器获取同一个服务（例如，HTTP）。这些服务器都是任播报文的响应方。如果不是采用任播地址通信，当其中一台服务器发生故障时，用户需要获取另一台服务器的地址才能重新建立通信。如果采用的是任播地址，当一台服务器发生故障时，任播报文的发起方能够自动与使用相同地址的另一台服务器通信，从而实现业务冗余。
- 使用多服务器接入还能够提高工作效率。例如，用户（即任播地址的发起方）浏览公司网页时，与相同的单播地址建立一条连接，连接的对端是具有相同任播地址的多个服务器。用户可以从不同的镜像服务器分别下载html文件和图片。用户利用多个服务器的带宽同时下载网页文件，其效率远远高于使用单播地址进行下载。
- 镜像服务器免责声明：
- 根据您的要求，出于检测通信传输中的故障和错误的目的，该特性可能涉及采集或存储个人用户某些通信内容。本公司无法单方采集或存储用户通信内容。建议您只有在所适用法律法规允许的目的和范围内方可启用相应的功能。在采集、存储用户通信内容的过程中，您应采取足够的措施以确保用户的通信内容受到严格保护。

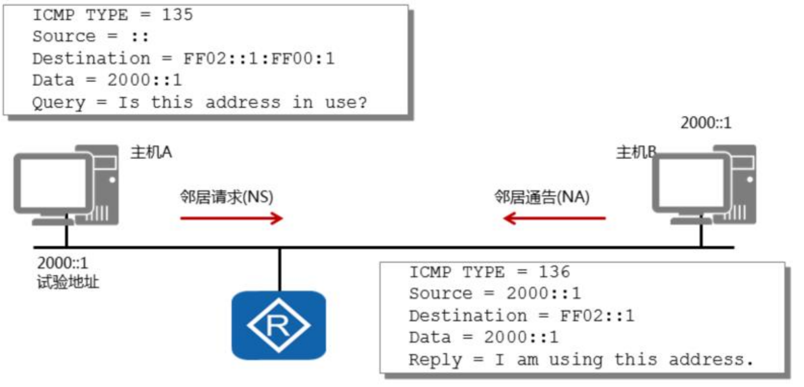


- IPv6支持无状态地址自动配置，无需使用诸如DHCP之类的辅助协议，主机即可获取IPv6前缀并自动生成接口ID。路由器发现功能是IPv6地址自动配置功能的基础，主要通过以下两种报文实现：
- RA报文：每台路由器为了让二层网络上的主机和其它路由器知道自己的存在，定期以组播方式发送携带网络配置参数的RA报文。RA报文的Type字段值为134。
- RS报文：主机接入网络后可以主动发送RS报文。RA报文是由路由器定期发送的，但是如果主机希望能够尽快收到RA报文，它可以立刻主动发送RS报文给路由器。网络上的路由器收到该RS报文后会立即向相应的主机单播回应RA报文，告知主机该网段的默认路由器和相关配置参数。RS报文的Type字段值为133。



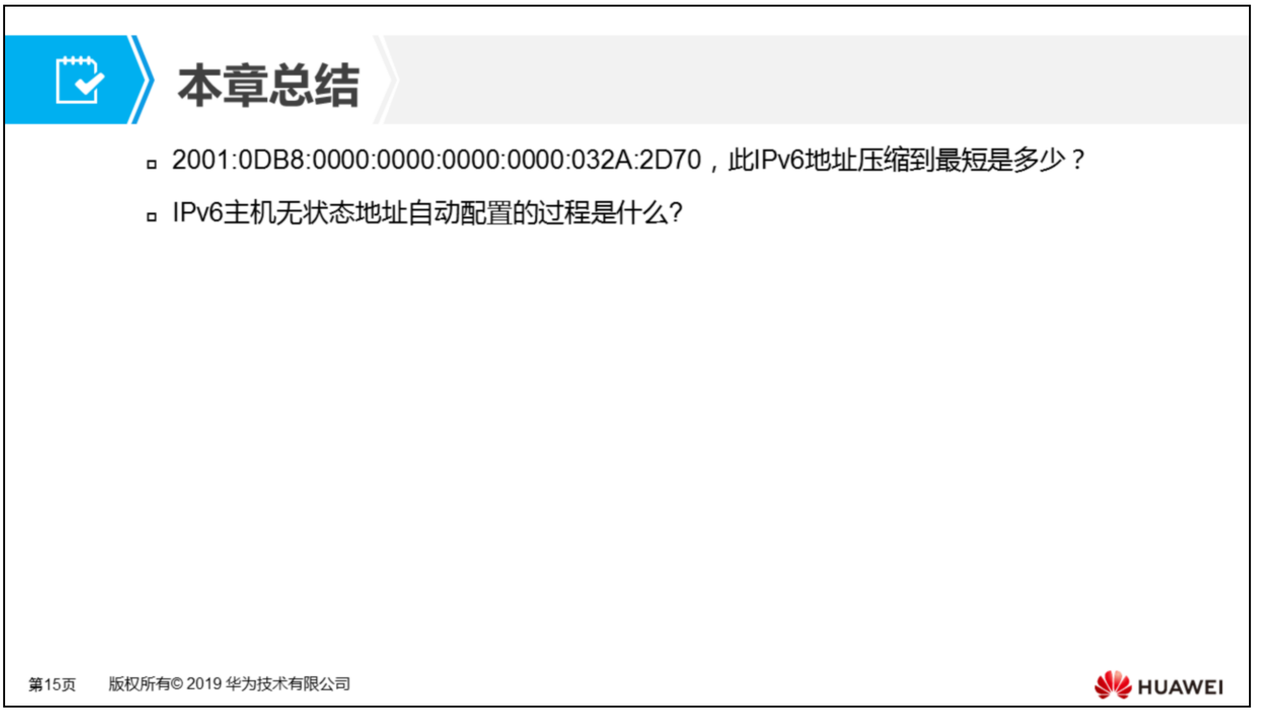
- 为了通过IPv6网络进行通信，各接口必须获取有效的IPv6地址，以下三种方式可以用来配置IPv6地址的接口ID：网络管理员手动配置；通过系统软件生成；采用扩展唯一标识符（EUI-64）格式生成。就实用性而言，EUI-64格式是IPv6生成接口ID的最常用方式。IEEE EUI-64标准采用接口的MAC地址生成IPv6接口ID。MAC地址只有48位，而接口ID却要求64位。MAC地址的前24位代表厂商ID，后24位代表制造商分配的唯一扩展标识。
- MAC地址的第七高位是一个U/L位，值为0时表示MAC地址全局唯一，值为1时表示MAC地址本地唯一。
- 在MAC地址向EUI-64格式的转换过程中，在MAC地址的前24位和后24位之间插入了16比特的FFFE，并将U/L位的值从0变成了1，这样就生成了一个64比特的接口ID，且接口ID的值全局唯一。接口ID和接口前缀一起组成接口地址。

IPv6无状态地址DAD检查



- 当为接口配置IPv6地址时，DAD用来在本地链路范围内检测将要使用的IPv6地址是否唯一。

- 设备在给接口分配IPv6单播地址之前会进行重复地址检测（DAD），确认是否有其它的节点使用了该地址。尤其是在地址自动配置的时候，进行DAD检测是很必要的。一个IPv6单播地址在分配给一个接口之后且通过重复地址检测之前称为试验地址，此时该接口不能使用这个试验地址进行单播通信，但是仍然会加入两个组播组：ALL-nodes组播组和Solicited-node组播组。Solicited-node组播组由单播或任播地址的后24位加上地址前缀FF02:0:0:0:0:1:FF00::/104组成。例如，本示例中配置的试验地址为2000::1，该地址被加入Solicited-node组播组FF02::1:FF00:1。
- IPv6重复地址检测技术和IPv4中的免费ARP类似：用于地址分配或主机连接网络时检测重复的IPv4主机地址。节点向一个自己将使用的试验地址所在的Solicited-node组播组发送一个以该试验地址为请求的目标地址的邻居请求（NS）报文，如果收到某个其它站点回应的邻居通告（NA）报文，就证明该地址已被网络上使用，节点将不能使用该试验地址进行通信。这种情况下，网络管理员需要手动为该节点分配另外一个地址。



- 地址2001:0DB8:0000:0000:0000:0000:032A:2D70可压缩为2001:DB8::32A:2D70。鉴于IPv6近乎无限的地址空间，两个或多个均为0的组可以用双冒号来显示，但是双冒号只能用一次。每组中的前导0都可以省略，但是最后一个0不可以省略。
- IPv6主机首先通过路由器发现功能来获取地址前缀信息，之后通过向接口已有的48比特MAC地址中插入16比特的FFEE生成接口ID，在生成了IPv6地址后会通过重复地址检测来确认地址是否唯一。


谢 谢
www.huawei.com

IPv6路由基础

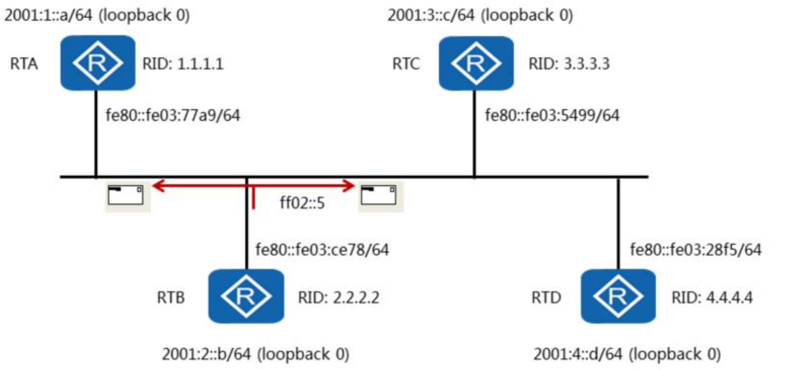
前言

- 在企业网络中，IPv6技术的应用越来越普及。IETF组织针对IPv6网络制定了路由协议OSPFv3。

目标

- 学完本课程后，您将能够：
  - 掌握OSPFv3的工作原理
  - 掌握OSPFv3的配置

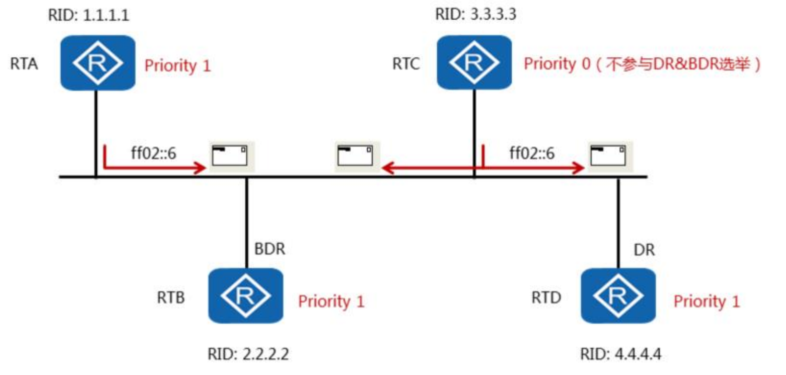
OSPFv3



- ff02::5是为OSPFv3路由协议预留的IPv6组播地址。
- OSPFv3中的路由条目下一跳地址是链路本地地址。

- OSPFv3是运行在IPv6网络的OSPF协议。运行OSPFv3的路由器使用物理接口的链路本地单播地址为源地址来发送OSPF报文。相同链路上的路由器互相学习与之相连的其它路由器的链路本地地址，并在报文转发的过程中将这些地址当成下一跳信息使用，虚链路的场景不在本课程的讨论范围内。
- IPv6中使用组播地址ff02::5来表示AllSPFRouters，而OSPFv2中使用的是组播地址224.0.0.5。需要注意的是，OSPFv3和OSPFv2版本互不兼容。

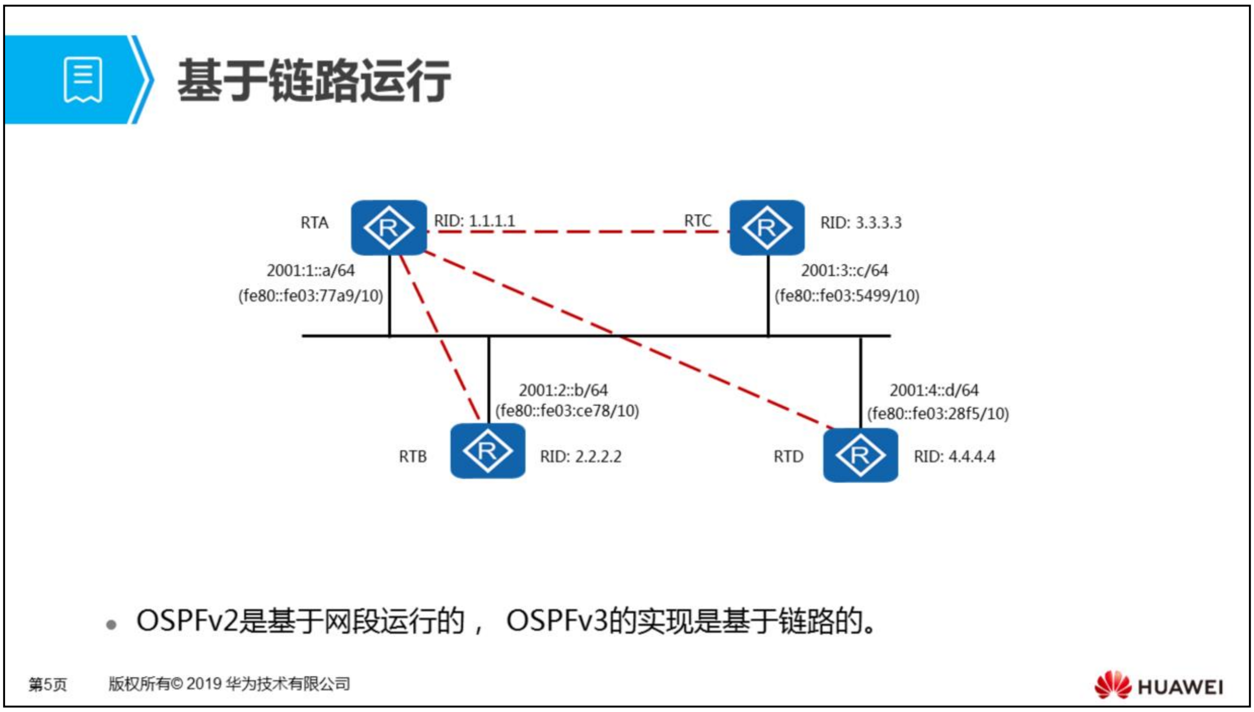
DR&BDR



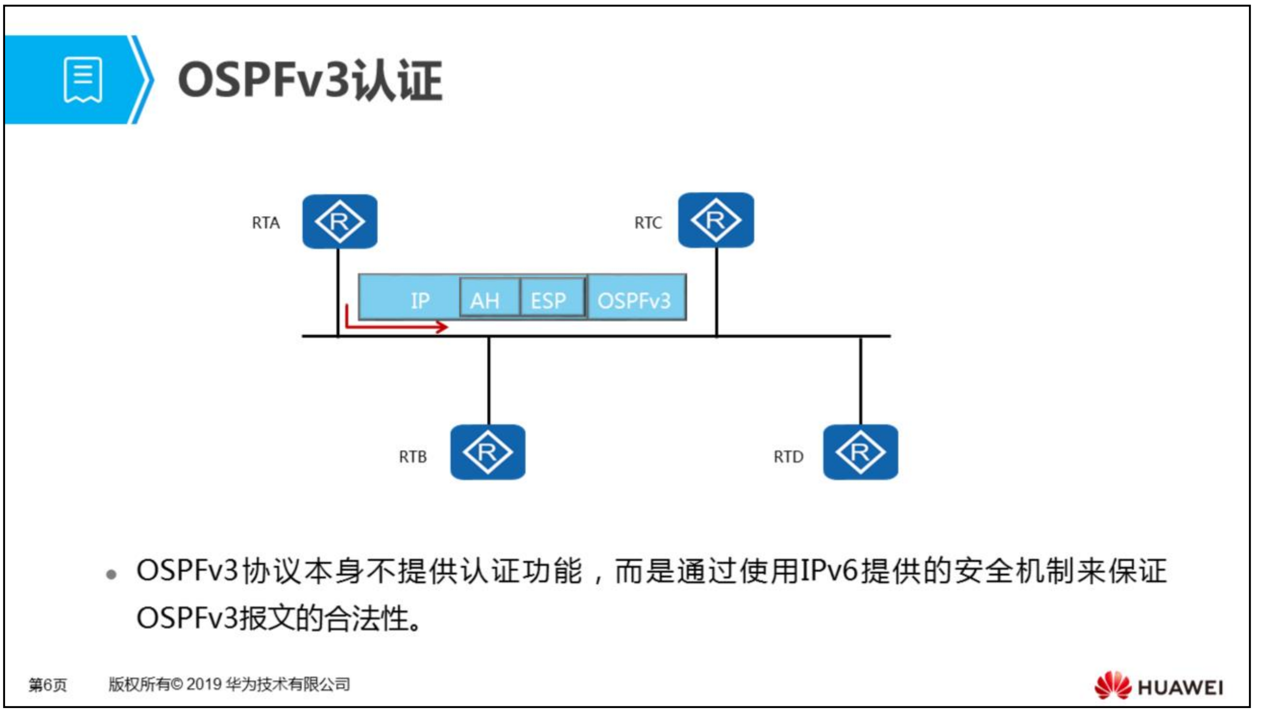
- Router ID在OSPFv3中必须手动配置。
- 在NBMA和广播型网络中OSPFv3选举DR和BDR的过程与OSPFv2相似。

  HUAWEI

- Router ID在OSPFv3中也是用于标识路由器的。与OSPFv2的Router ID不同，OSPFv3的Router ID必须手工配置；如果没有手工配置Router ID，OSPFv3将无法正常运行。OSPFv3在广播型网络和NBMA网络中选举DR和BDR的过程与OSPFv2相似。IPv6使用组播地址FF02::6表示AllDRouters，而OSPFv2中使用的是组播地址224.0.0.6。

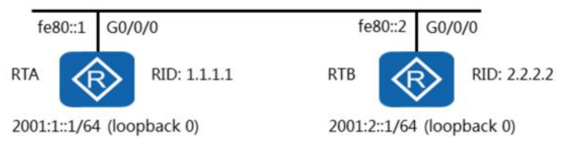


- OSPFv3是基于链路而不是网段的。在配置OSPFv3时，不需要考虑路由器的接口是否配置在同一网段，只要路由器的接口连接在同一链路上，就可以不配置IPv6全局地址而直接建立联系。这一变化影响了OSPFv3协议报文的接收、Hello报文的内容以及网络LSA的内容。



- OSPFv3直接使用IPv6的扩展头部（AH和ESP）来实现认证及安全处理，不再需要OSPFv3自身来完成认证。

OSPFv3配置




```
[RTA]ipv6
[RTA]ospfv3
[RTA-ospfv3-1]router-id 1.1.1.1
[RTA-GigabitEthernet0/0/0]ipv6 enable
[RTA-GigabitEthernet0/0/0]ipv6 address fe80::1 link-local
[RTA-GigabitEthernet0/0/0]ospfv3 1 area 0.0.0.0
[RTA-LoopBack0]ipv6 enable
[RTA-LoopBack0]ipv6 address 2001:1::1/64
[RTA-LoopBack0]ospfv3 1 area 0.0.0.0
```


- **ipv6**命令用来使能路由器的IPv6功能。要在路由器上运行OSPFv3协议，首先必须使能IPv6功能。
- **ospfv3** [*process-id*]命令用来创建并运行OSPFv3进程，*process-id*取值范围是1～65535。如果不指定进程号，缺省使用进程号1。
- **router-id** *router-id*命令用来设置运行OSPFv3协议的路由器ID号。
- **ipv6 enable**命令用来在路由器接口上使能IPv6，使得接口能够接收和转发IPv6报文。接口的IPv6功能默认是去使能的。**ipv6 address** <*link local address*> **link-local**命令用来手动为接口配置链路本地地址。
- **ospfv3** *process-id* **area** *area-id*命令用来在接口上使能OSPFv3的进程，并指定所属区域。
- 本示例中，路由器RTA的loopback接口和GigabitEthernet0/0/0接口都启用OSPFv3进程，并且都属于区域0。

配置验证

```
[RTA]display ospfv3
 Routing Process "OSPFv3 (1)" with ID 1.1.1.1
 Route Tag: 0
 Multi-VPN-Instance is not enabled
 SPF Intelligent Timer[millisecs] Max: 10000, Start: 500, Hold: 2000
 LSA Intelligent Timer[millisecs] Max: 5000, Start: 500, Hold: 1000
 LSA Arrival interval 1000 millisecs
 Default ASE parameters: Metric: 1 Tag: 1 Type: 2
 Number of AS-External LSA 0. AS-External LSA's Checksum Sum 0x0000
 Number of AS-Scoped Unknown LSA 0. AS-Scoped Unknown LSA's Checksum
Sum 0x0000
 Number of FULL neighbors 1
 Number of Exchange and Loading neighbors 0
……
```


- 在邻居路由器上完成OSPFv3配置后，执行display ospfv3命令可以验证OSPFv3配置及相关参数。从显示信息中可以看到正在运行的OSPFv3进程为1，Router ID为1.1.1.1，Number of FULL neighbors值为1。

本章总结

- OSPFv3用来唯一标识一台路由器的参数是什么?

- Router ID用于唯一标识一台运行OSPFv3协议的路由器。


谢 谢
www.huawei.com

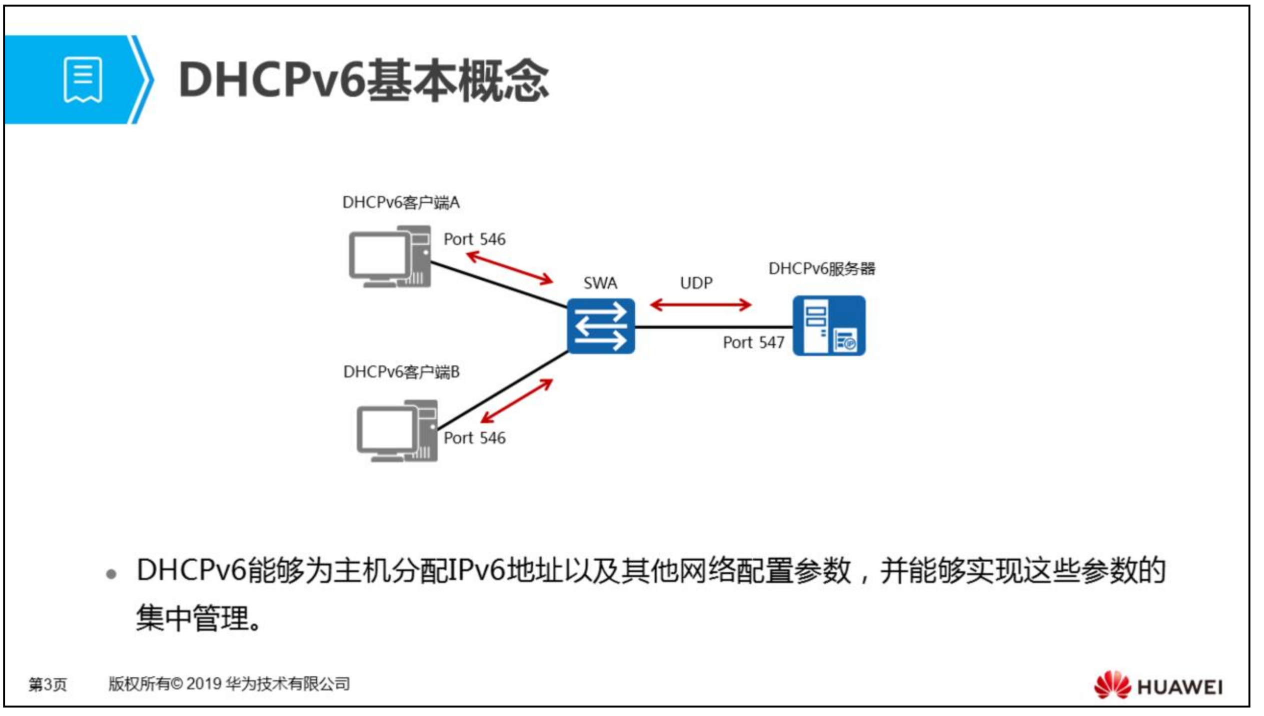
DHCPv6原理与配置

前言

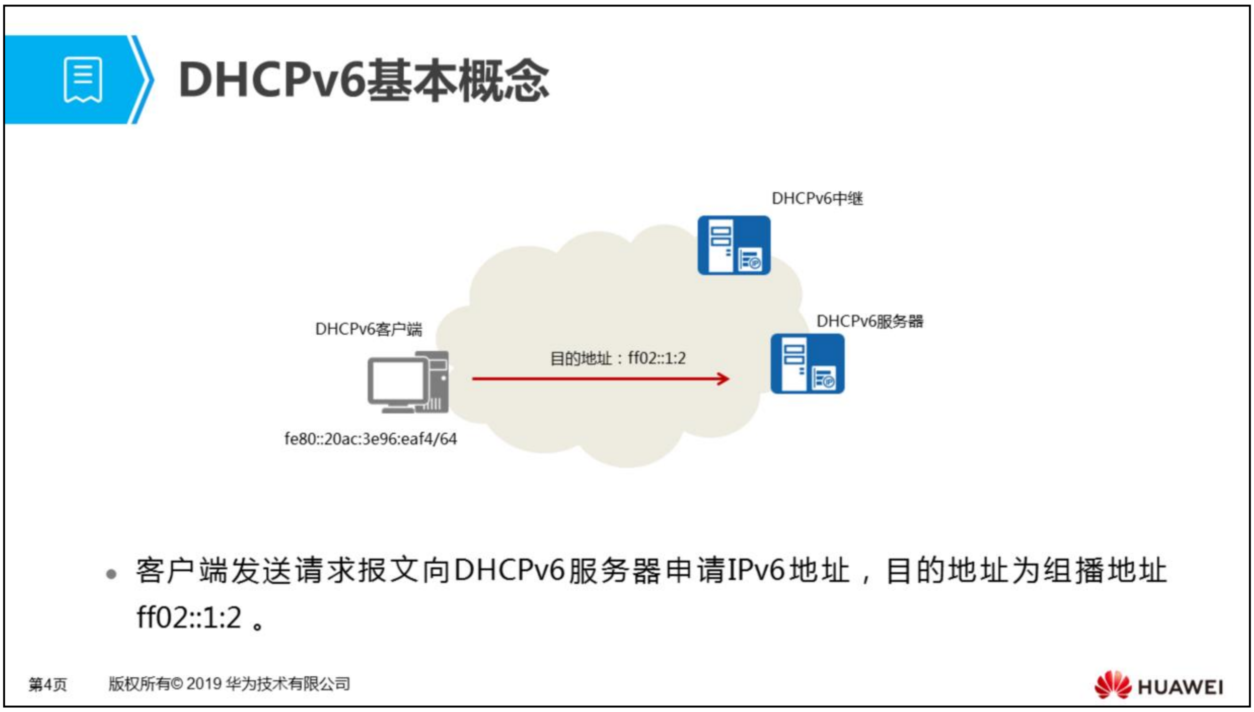
- 主机在运行IPv6时，可以通过使用无状态地址自动配置或DHCPv6协议来获取IPv6地址。IPv6动态主机配置协议DHCPv6(Dynamic Host Configuration Protocol for IPv6)采用了客户端/服务器通信模式，是针对IPv6编址方案设计的、为主机分配IPv6地址和其他网络配置参数的协议。

目标

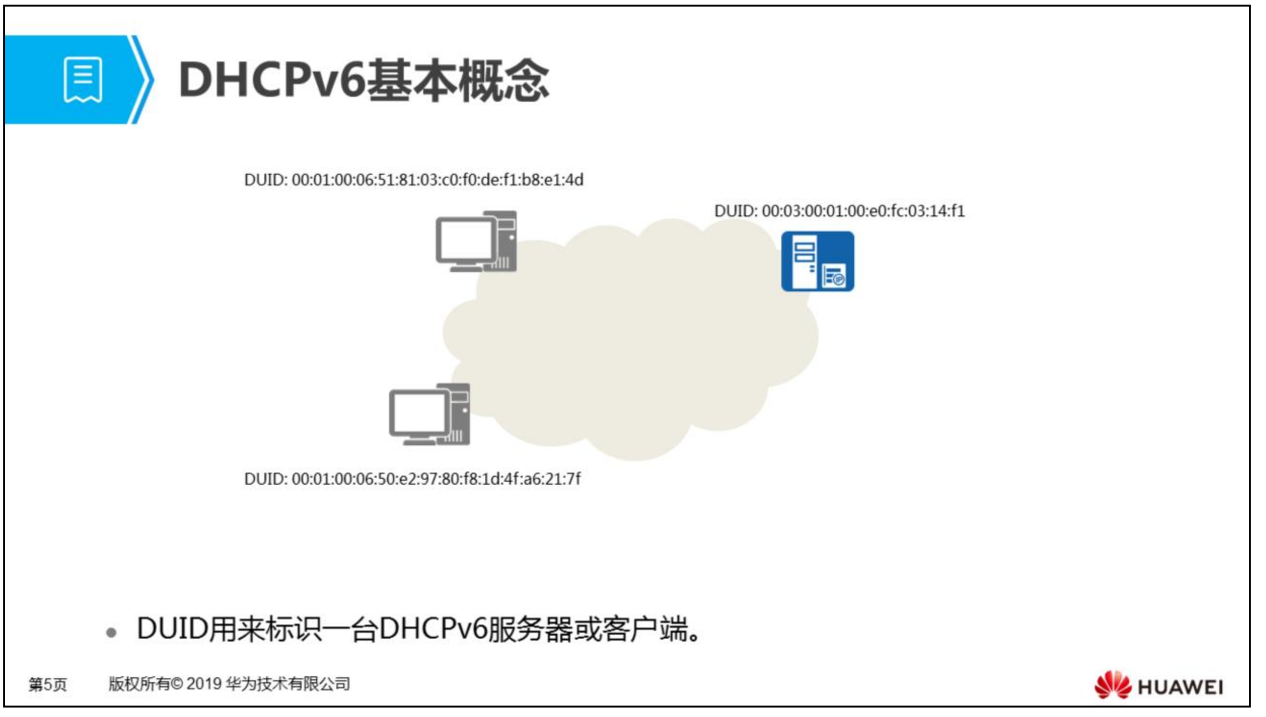
- 学完本课程后，您将能够：
  - 掌握DHCPv6的工作原理
  - 掌握DHCPv6的配置



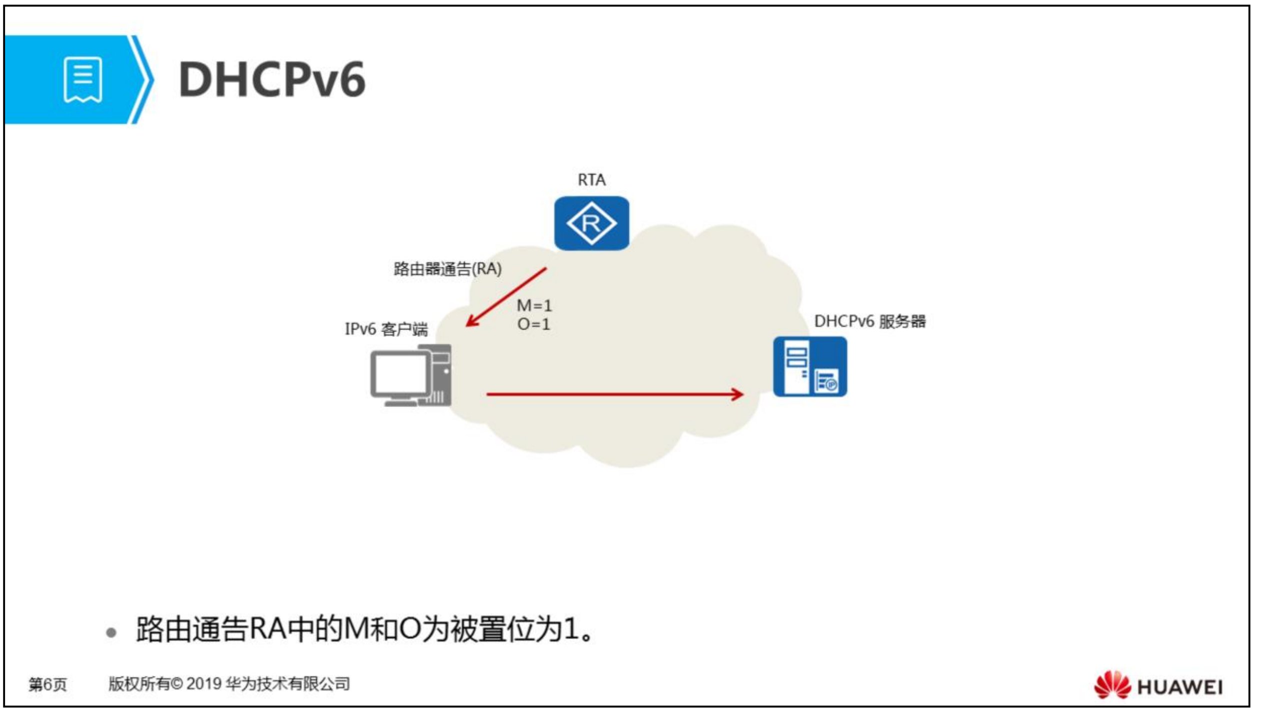
- 主机在运行IPv6时，可以通过使用无状态地址自动配置或DHCPv6协议来获取IPv6地址。
- 主机使用无状态地址自动配置方案来获取IPv6地址时，路由器并不记录主机的IPv6地址信息，可管理性差；另外，IPv6主机无法获取DNS服务器地址等网络配置信息，在可用性上也存在一定的缺陷。
- DHCPv6属于一种有状态地址自动配置协议。在有状态地址配置过程中，DHCPv6服务器为主机分配一个完整的IPv6地址，并提供DNS服务器地址等其他配置信息。此外，DHCPv6服务器还可以对已经分配的IPv6地址和客户端进行集中管理。
- DHCPv6服务器与客户端之间使用UDP协议来交互DHCPv6报文，客户端使用的UDP端口号是546，服务器使用的UDP端口号是547。



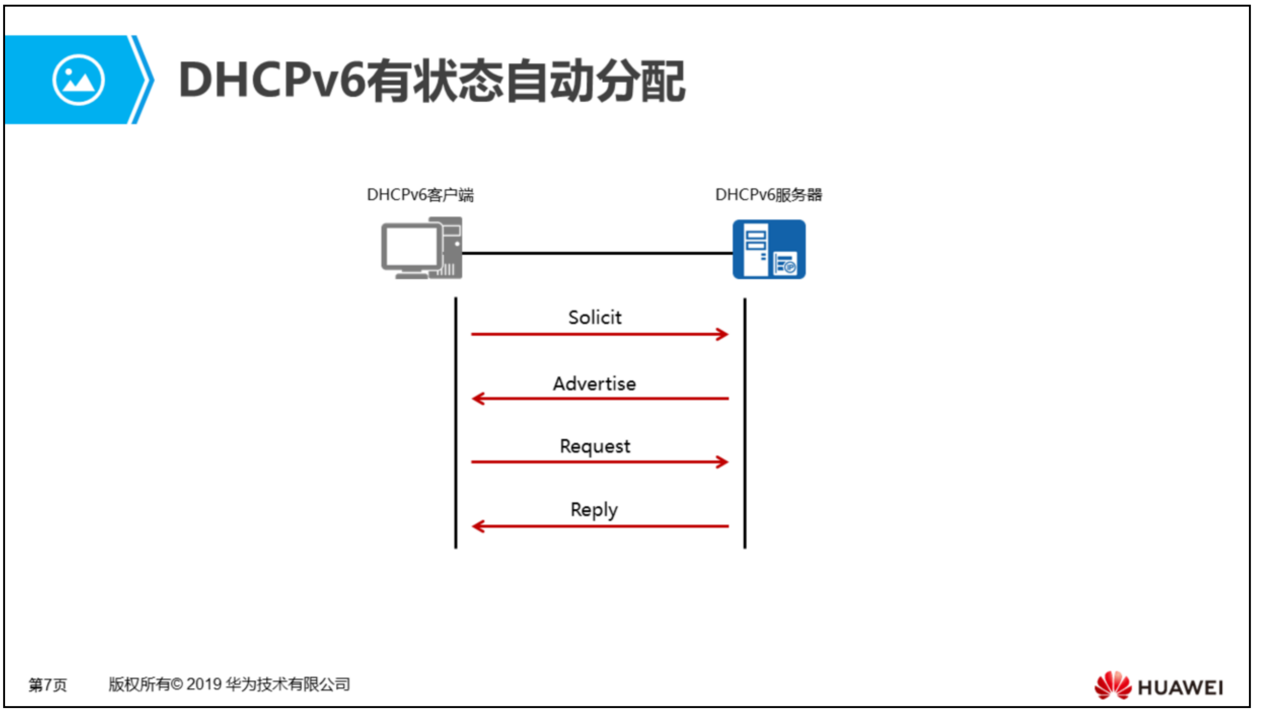
- DHCPv6基本协议架构中，主要包括以下三种角色：
- DHCPv6客户端：通过与DHCPv6服务器进行交互，获取IPv6地址/前缀和网络配置信息，完成自身的地址配置功能。
- DHCPv6中继：负责转发来自客户端方向或服务器方向的DHCPv6报文，协助DHCPv6客户端和DHCPv6服务器完成地址配置功能。只有当DHCPv6客户端和DHCPv6服务器不在同一链路范围内，或者DHCPv6客户端和DHCPv6服务器无法单播交互的情况下，才需要DHCPv6中继的参与。
- DHCPv6服务器：负责处理来自客户端或中继的地址分配、地址续租、地址释放等请求，为客户端分配IPv6地址/前缀和其他网络配置信息。
- 客户端发送DHCPv6请求报文来获取IPv6地址等网络配置参数，使用的源地址为客户端接口的链路本地地址，目的地址为ff02::1:2。ff02::1:2表示的是所有DHCPv6服务器和中继，这个地址是链路范围的。



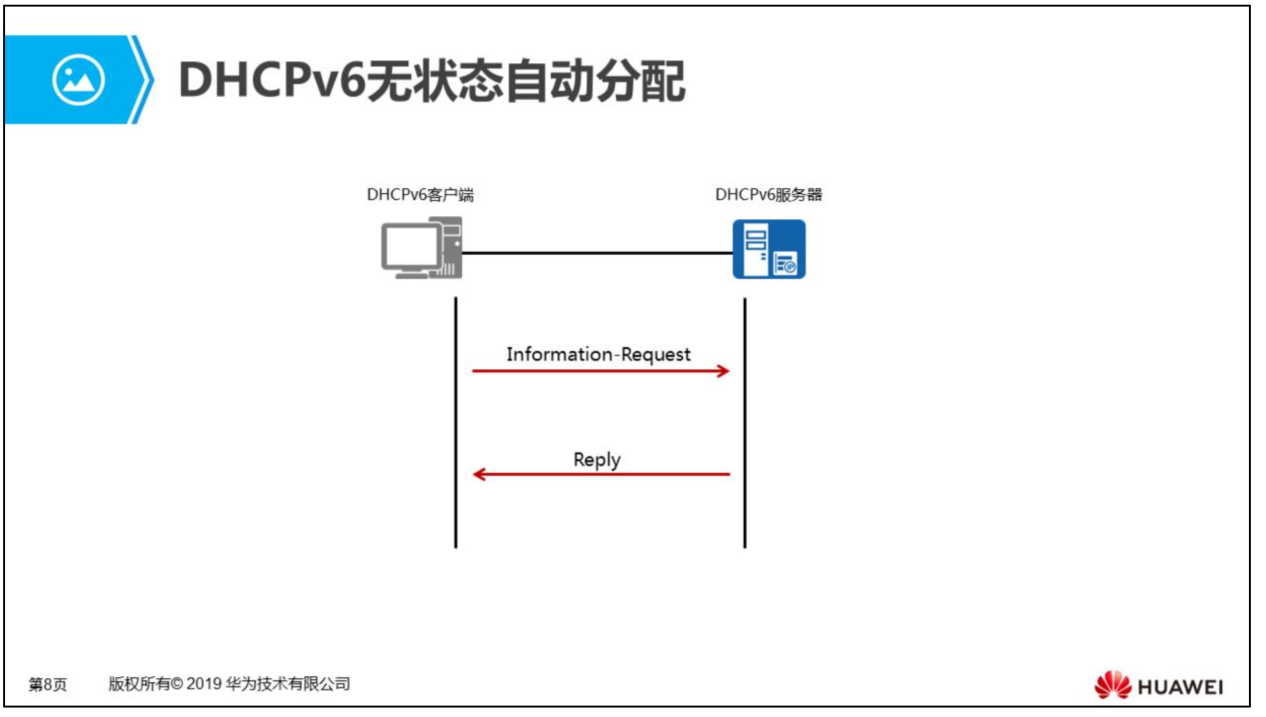
- DHCP设备唯一标识符DUID（DHCPv6 Unique Identifier）用来标识一台DHCPv6服务器或客户端。每个DHCPv6服务器或客户端有且只有一个DUID。
- DUID采用以下两种方式生成：
- 基于链路层地址（LL）：即采用链路层地址方式来生成DUID。
- 基于链路层地址与时间组合（LLT）：即采用链路层地址和时间组合方式来生成DUID。



- DHCPv6分配地址时又分为:
- DHCPv6有状态自动分配：DHCPv6服务器为客户端分配IPv6地址及其他网络配置参数（如DNS、NIS、SNTP服务器地址等）。
- DHCPv6无状态自动分配：主机的IPv6地址仍然通过路由通告方式自动生成，DHCPv6服务器只分配除IPv6地址以外的配置参数（如DNS、NIS、SNTP服务器等）。
- DHCPv6客户端在向DHCPv6服务器发送请求报文之前，会发送RS报文，在同一链路范围的路由器接收到此报文后会回复RA报文。在RA报文中包含管理地址配置标记（M）和有状态配置标记（O）。当M取值为1时，启用DHCPv6有状态地址配置，即DHCPv6客户端需要从DHCPv6服务器获取IPv6地址，取值为0则启用IPv6无状态地址自动分配方案。当O取值为1时，用来定义客户端需要通过有状态的DHCPv6来获取其它网络配置参数，如DNS、NIS、SNTP服务器地址等，取值为0则启用IPv6无状态地址自动分配方案。

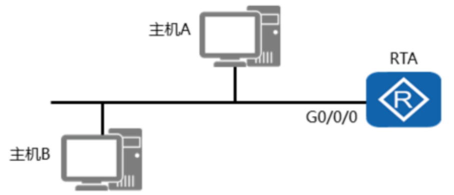


- DHCPv6四步交互地址分配过程如下：
- DHCPv6客户端发送Solicit报文，请求DHCPv6服务器为其分配IPv6地址和网络配置参数。
- DHCPv6服务器回复Advertise报文，该报文中携带了为客户端分配的IPv6地址以及其它网络配置参数。
- DHCPv6客户端如果接收到了多个服务器回复的Advertise报文，则会根据Advertise报文中的服务器优先级等参数来选择优先级最高的一台服务器，并向所有的服务器发送Request组播报文。
- 被选定的DHCPv6服务器回复Reply报文，确认将IPv6地址和网络配置参数分配给客户端使用。



- DHCPv6无状态工作过程如下：
- DHCPv6客户端以组播方式向DHCPv6服务器发送Information-Request报文，该报文中携带Option Request选项，用来指定DHCPv6客户端需要从DHCPv6服务器获取的配置参数。
- DHCPv6服务器收到Information-Request报文后，为DHCPv6客户端分配网络配置参数，并单播发送Reply报文，将网络配置参数返回给DHCPv6客户端。
- DHCPv6客户端根据收到的Reply报文中提供的参数完成DHCPv6客户端无状态配置。

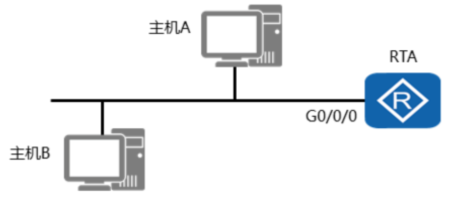
DHCPv6配置




```
[RTA]dhcpv6 duid ll
Warning: The DHCP unique identifier should be globally-unique and
stable. Are you sure to change it? [Y/N]y
```


- dhcpv6 duid { ll | llt }命令可以用来指定DUID格式为DUID-LL或DUID-LLT。缺省情况下，ARG3系列路由器采用的DUID格式是DUID-LL。当使用DUID-LLT格式时，时间戳值引用的是从执行dhcpv6 duid llt命令的时间点开始计算的时间。
- 可以使用display dhcpv6 duid命令来验证当前使用的DUID格式以及DUID值。

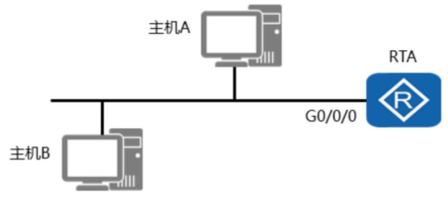
DHCPv6配置




```
[RTA]dhcpv6 pool pool1
[RTA-dhcpv6-pool-pool1]address prefix 3000::/64
[RTA-dhcpv6-pool-pool1]excluded-address 3000::1
[RTA-dhcpv6-pool-pool1]dns-server 4000::1
[RTA-dhcpv6-pool-pool1]dns-domain name huawei.com
```


- dhcpv6 pool pool-name命令用来创建IPv6地址池或进入到IPv6地址池视图。
- address prefix ipv6-prefix/ipv6-prefix-length命令用来在IPv6地址池视图下绑定IPv6地址前缀。ipv6-prefix/ipv6-prefix-length用来指定IPv6地址池绑定的网络前缀和前缀长度。
- excluded-address start-ipv6-address [to end-ipv6-address]命令用来配置IPv6地址池中不参与自动分配的IPv6地址范围。
- dns-server ipv6-address命令用来配置DNS服务器的IPv6地址。
dns-domain-name dns-domain-name命令用来配置为DHCPv6客户端分配的域名后缀。

DHCPv6配置




```
[RTA]ipv6
[RTA]interface GigabitEthernet 0/0/0
[RTA-GigabitEthernet0/0/0]ipv6 enable
[RTA-GigabitEthernet0/0/0]ipv6 address 3000::1/64
[RTA-GigabitEthernet0/0/0]dhcpv6 server pool1
```


- dhcpv6 server pool-name命令用来在接口下配置DHCPv6服务器功能，pool-name用来指定接口下配置的DHCPv6地址池名称。

配置验证

```
[RTA]display dhcpv6 pool
DHCPv6 pool: p
  Address prefix: 3000::/64
    Lifetime valid 172800 seconds, preferred 86400 seconds
    2 in use, 0 conflicts
  Excluded-address 3000::1
  1 excluded addresses
  Information refresh time: 86400
  DNS server address: 4000::1
  Conflict-address expire-time: 172800
  Active normal clients: 2
```

  HUAWEI

- display dhcpv6 pool命令用来查看DHCPv6服务器上配置的地址池信息。
- 本例中，RTA上有一个DHCPv6地址池，该地址池关联的地址前缀为3000::1/64，生存周期为172800秒，即两天（缺省情况下，生存周期是86,400秒或1天）。在必要的情况下，可以在IPv6地址池视图下使用information-refresh命令重新配置其它配置信息。对于处在活跃状态的客户端从DHCPv6服务器租用的IPv6地址，可以查看相关的信息统计。

本章总结

- ARG3系列路由器生成DUID的方式有哪些?
- 如果主机收到的路由器通告信息中M和O位被置1，主机将如何操作？

- ARG3系列路由器支持DUID-LL和DUID-LLT格式生成DUID。
- 当接收到携带M和O（比特值均为1）的RA报文时，主机将主动发现DHCPv6服务器用于有状态地址配置。配置信息包括IPv6地址和其它配置参数，例如地址前缀和DNS服务器地址等。


谢 谢
www.huawei.com

MPLS基本原理

前言

- MPLS（Multiprotocol Label Switching）位于TCP/IP 协议栈中的链路层和网络层之间，用于向IP层提供连接服务，同时又从链路层得到服务。MPLS以标签交换替代IP转发。
- MPLS 并不是一种业务或者应用，它实际上是一种隧道技术。这种技术不仅支持多种高层协议与业务，而且在一定程度上可以保证信息传输的安全性。

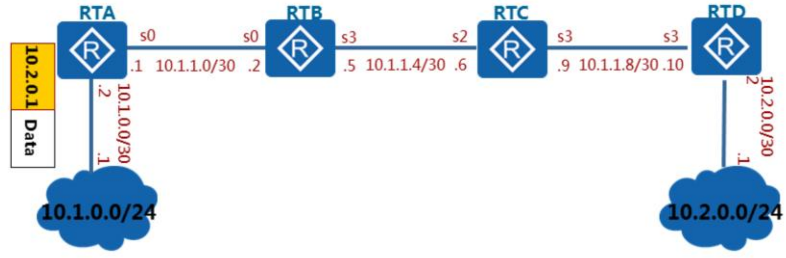
目标

- 学完本课程后，您将能够：
  - 描述MPLS产生的背景及典型应用
  - 描述MPLS的基本概念及工作原理

目录

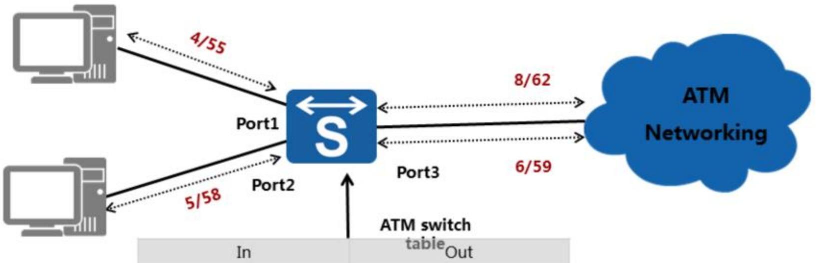
传统IP转发

Network	Nexthop	Network	Nexthop	Network	Nexthop	Network	Nexthop
10.1.0.0/24	10.1.0.2	10.1.0.0/24	10.1.1.1	10.1.0.0/24	10.1.1.5	10.1.0.0/24	10.1.1.9
10.1.0.1/32	10.1.0.1	10.1.1.0/30	10.1.1.2	10.1.1.0/30	10.1.1.5	10.1.1.0/30	10.1.1.9
10.1.1.0/30	10.1.1.1	10.1.1.1/32	10.1.1.1	10.1.1.4/30	10.1.1.6	10.1.1.4/30	10.1.1.9
10.1.1.2/32	10.1.1.2	10.1.1.4/30	10.1.1.5	10.1.1.5/32	10.1.1.5	10.1.1.8/30	10.1.1.10
10.1.1.4/30	10.1.1.2	10.1.1.6/32	10.1.1.6	10.1.1.8/30	10.1.1.9	10.1.1.9/32	10.1.1.9
10.1.1.8/30	10.1.1.2	10.1.1.8/30	10.1.1.6	10.1.1.10/32	10.1.1.10	**10.2.0.0/24**	**10.2.0.2**
10.2.0.0/24	**10.1.1.2**	**10.2.0.0/24**	**10.1.1.6**	**10.2.0.0/24**	**10.1.1.10**	10.2.0.1/32	10.2.0.1



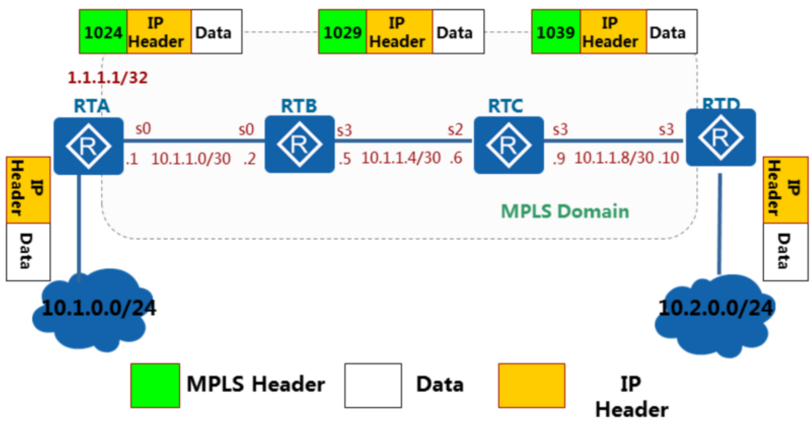
HUAWEI

ATM信元转发

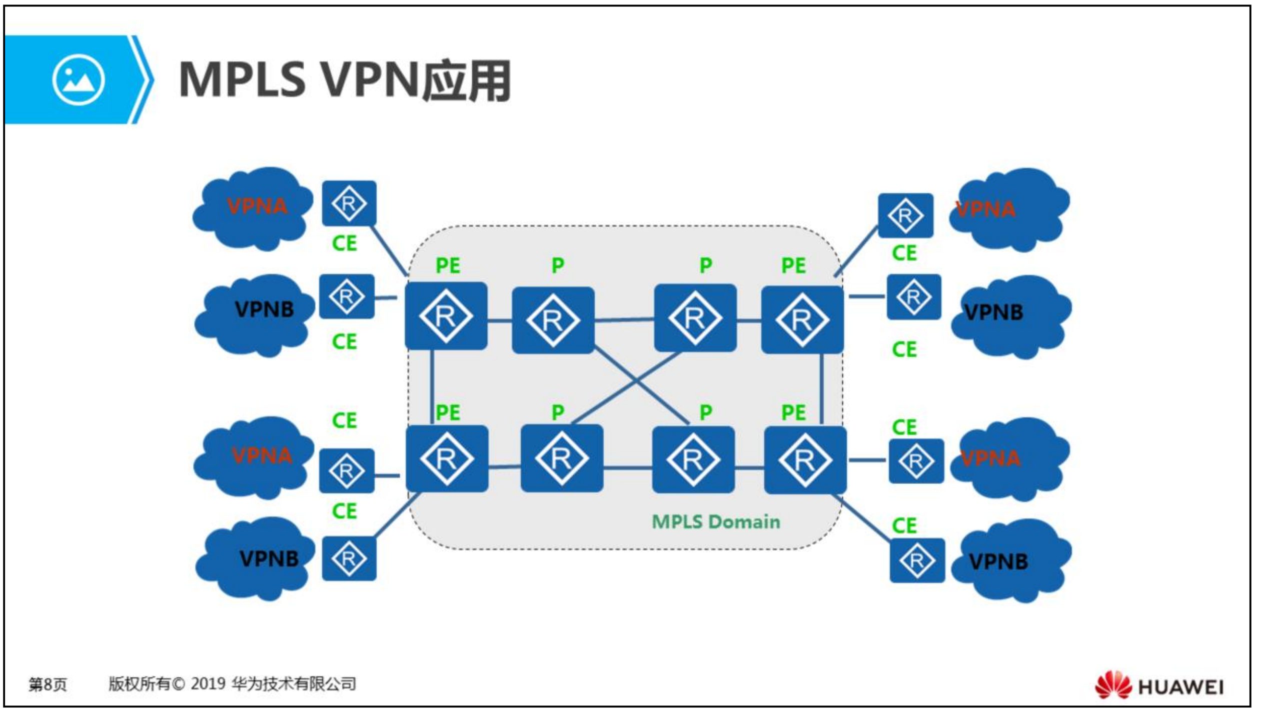


In		Out	
port	VPI/VCI	port	VPI/VCI
1	4/55	3	8/62
2	5/58	3	6/59
3	8/62	1	4/55
3	6/59	2	5/58

MPLS标签转发

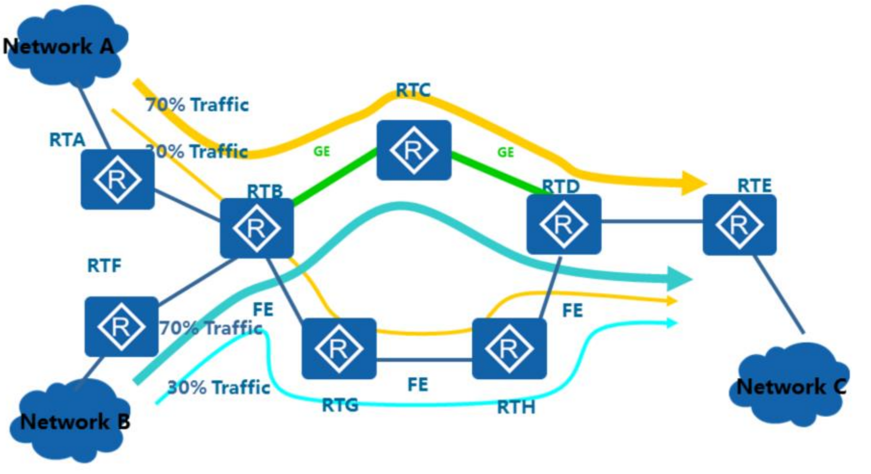


- 传统的IP技术简单，且部署成本低，但由于硬件技术存在限制，基于最长匹配算法的IP技术必须使用软件查找路由，转发性能低下。为了适应网络的发展，ATM（Asynchronous Transfer Mode）技术应运而生。ATM采用定长标签（即信元），并且只需要维护比路由表规模小得多的标签表，能够提供比IP路由方式高得多的转发性能。如何结合IP与ATM的优点成为当时热门话题。多协议标签交换技术MPLS（Multiprotocol Label Switching）就是在这种背景下产生的。



- 然而，随着ASIC(Application Specific Integrated Circuit)技术的发展，路由查找速度已经不是阻碍网络发展的瓶颈。这使得MPLS在提高转发速度方面不再具备明显的优势。但另一方面结合其它技术的应用上MPLS成为了一大亮点。

MPLS TE应用

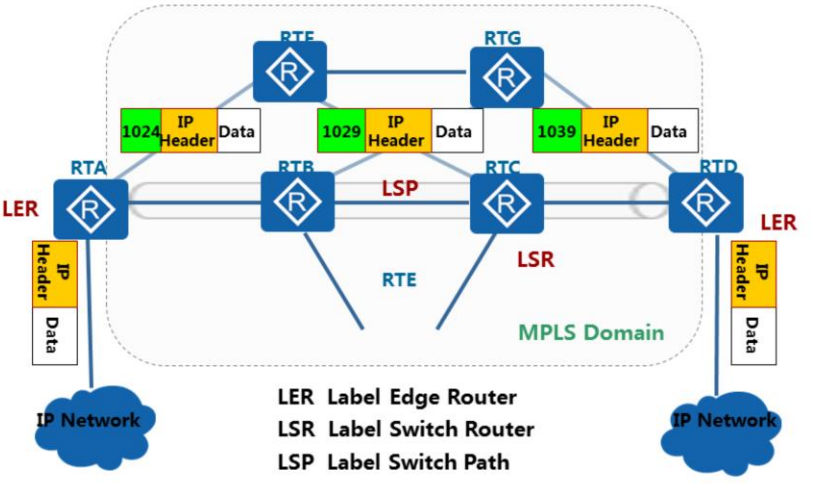


  HUAWEI

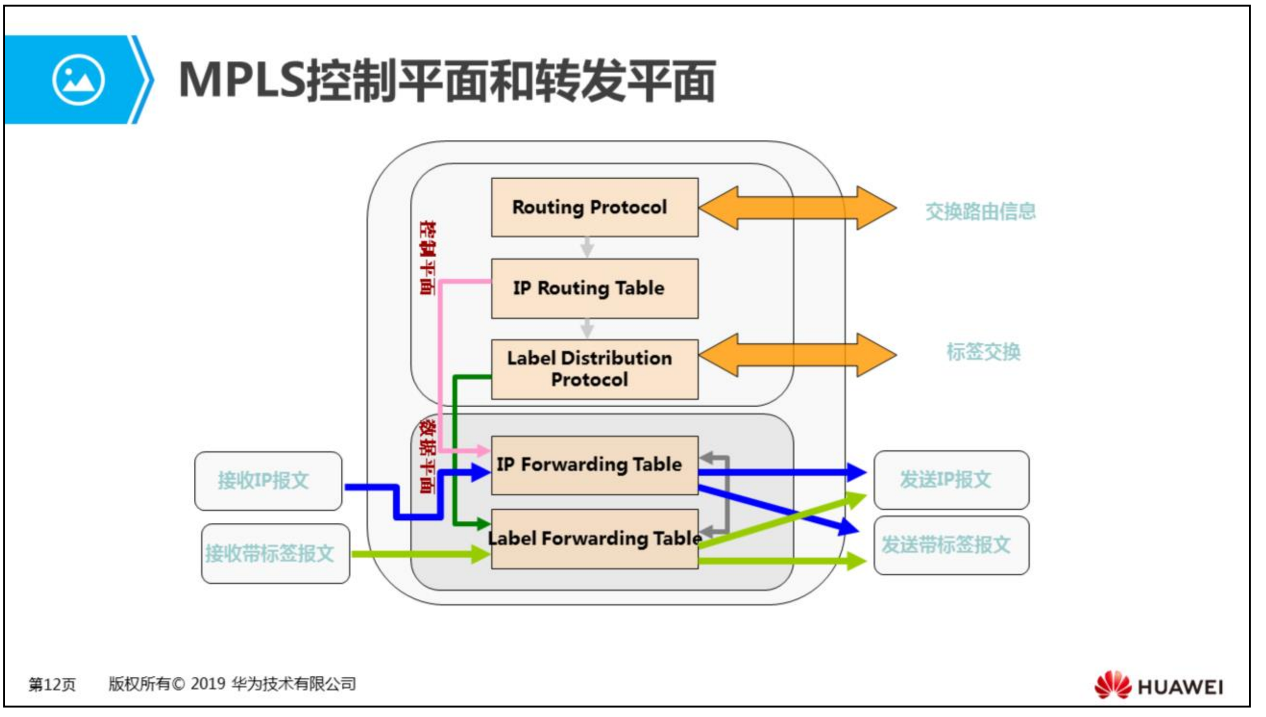
目录

HUAWEI

MPLS网络模型



LER Label Edge Router
LSR Label Switch Router
LSP Label Switch Path



- 为了更好的理解MPLS技术必须要了解它的体系结构，MPLS的体系结构由控制平面（Control Plane）和转发平面（Forwarding Plane）组成。

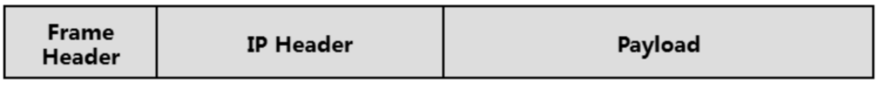
目录

HUAWEI

范例

帧模式MPLS

- MPLS有两种封装模式：帧模式和信元模式（ATM采用MPLS信元模式封装，本课不涉及）。帧模式封装是直接在报文的二层头部和三层头部之间增加一个MPLS标签头。以太网 、PPP采用这种封装模式。



二层帧格式

MPLS帧模式封装

  HUAWEI

- MPLS标签是MPLS信息传递的载体，路由器之间通过标签的交互，完成在建立的标签转发路径上传送数据。

范例

MPLS报文头部

- MPLS 头部长度为32bits。
- LABEL：该标签用于报文转发，长度为20bits；
- EXP：通常用来承载IP报文中的优先级，长度为3bits；
- S：标识栈底用来表明是否是最后一个标签（MPLS标签可以多层嵌套），长度为1bit；
- TTL：类似IP头部的TTL，用来防止报文环路等，长度为8bits；

  HUAWEI

范例

MPLS标签嵌套

- PID 标识二层头部后面的报文类型
  - Ethernet 0x0800 IPv4 0x8847 MPLS单播报文0x8848 MPLS多播报文
  - PPP 0x8021 IPv4 0x8281 MPLS单播报文0x8283 MPLS多播报文
- S 标识是否是栈底标签
- 标签嵌套应用
  - MPLS VPN
  - MPLS TE

- 这里的Label1，Label2，Label3都指的是前一个胶片中的4个Bytes的MPLS头部，其中包含有20bits的标签信息。

目录

范例

FEC与NHLFE

- FEC（Forwarding Equivalence Class）转发等价类，是一组具有某些共性的数据流的集合。这些数据流在转发过程中被LSR 以相同方式处理。
- FEC 可以根据地址、业务类型、QoS 等要素进行划分。例如，在传统的采用最长匹配算法的IP 转发中，到同一条路由的所有报文就是一个转发等价类。
- NHLFE（Next Hop Label Forwarding Entry）：进行标签转发时用到，NHLFE包含这样一些基本信息：
  - 报文的下一跳
  - 如何进行标签操作（包括压入新的标签，弹出标签，用新的标签替换原有的标签等操作）。
  - NHLFE还可能包含一些其他信息，如发送报文使用的链路层封装等。

HUAWEI

范例

FEC与NHLFE举例

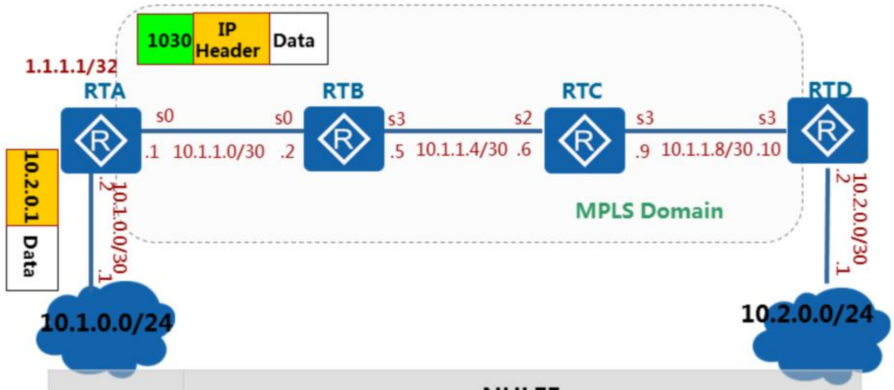
- *FEC：Forwarding Equivalence Classes（转发等价类）*
- NHLFE：Next Hop Label Forwarding Entry （下一跳标签转发表项）

```
<RTA>display  mpls lsp include 10.2.0.0 24 verbose
--------------------------------------------------
                 LSP Information: LDP LSP
--------------------------------------------------
  No                     :  1
  VrfIndex               :
  Fec                    :  10.2.0.0/24
  Nexthop                :  10.1.1.2
  In-Label               :  NULL
  Out-Label              :  1030
  In-Interface           :  ----------
  Out-Interface          :  Serial0
  LspIndex               :  10249
  Token                  :  0x22005
  LsrType                :  Ingress
  Outgoing token         :  0x0
  Label Operation        :  PUSH
  Mpls-Mtu               :  1500
  TimeStamp              :  822sec
```

范例

MPLS转发过程-Ingress LER(RTA)

- 在Ingress，通过查询FIB 表和NHLFE 表指导报文的转发。



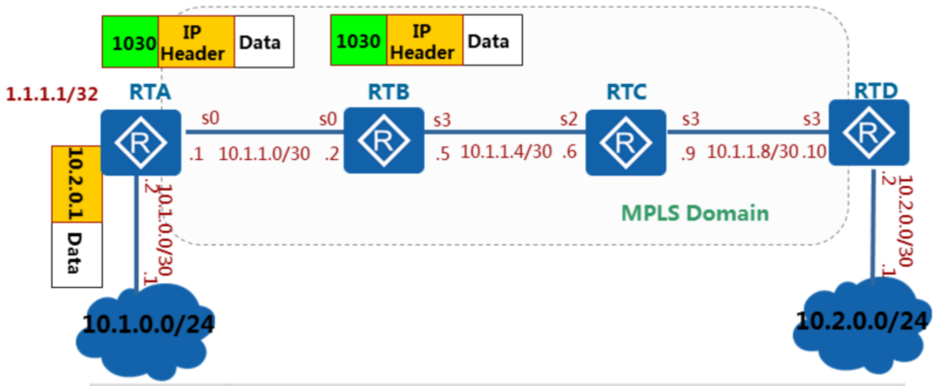
FEC	NHLFE			
	NextHop	Out Interface	Label Operation	Others
10.2.0.0	10.1.1.2	Serial0	Push	......

  HUAWEI

范例

MPLS转发过程-LSR(RTB)

- 在Transit，通过查询ILM(Incoming Label Map)表和NHLFE 表指导MPLS 报文的转发。

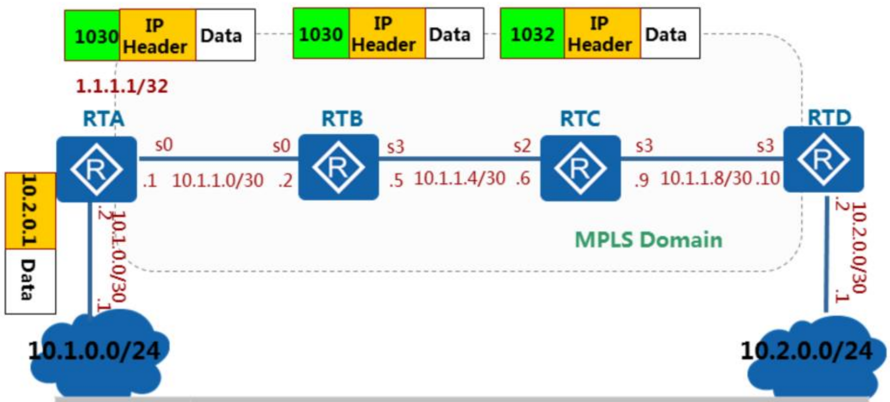


InLabel	NHLFE			
	NextHop	Out Interface	Label Operation	Others
1030	10.1.1.6	Serial3	SWAP	……

- 入标签到一组下一跳标签转发表项的映射称为入标签映射ILM（Incoming Label Map）。ILM在Transit节点的作用是将标签和NHLFE绑定。通过标签索引ILM表，就相当于使用目的IP地址查询FIB，能够得到所有的标签转发信息。

范例

MPLS转发过程-LSR(RTC)



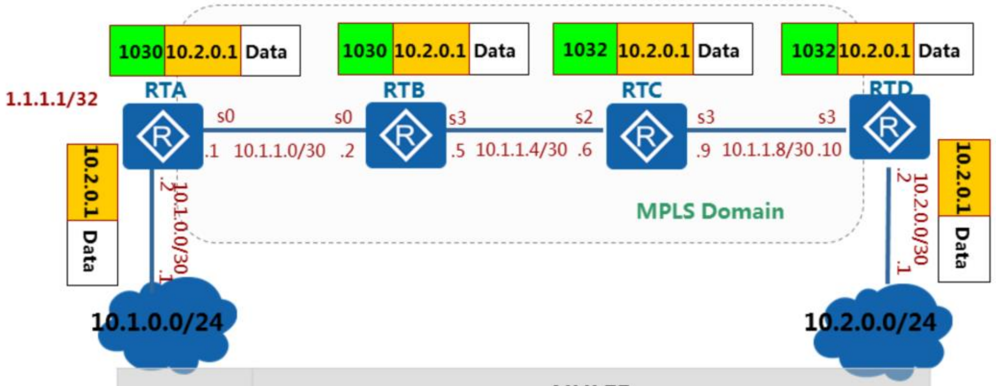
InLabel	NHLFE			
	NextHop	Out Interface	Label Operation	Others
1030	10.1.1.10	Serial3	SWAP	......

HUAWEI

范例

MPLS转发过程-Egress LER(RTD)

- 在Egress，通过查询ILM 表指导MPLS 报文的转发。



InLabel	NHLFE			
	NextHop	Out Interface	Label Operation	Others
1032	10.2.0.2	--------	POP	......

HUAWEI

思考题

1. MPLS头部哪个字段用于标示栈底标签（ ）。
   - A. Label
   - B. EXP
   - C. S
   - D. TTL
2. 基于MPLS标签转发报文时，数据包可能需要执行的动作有哪些（ ）。
   - A. Push
   - B. Pop
   - C. Swap
   - D. Switch

3.

- 参考答案：

  1： C。

  2：ABC。

本章总结

- MPLS产生背景
- MPLS基本概念
- MPLS报文格式
- MPLS转发原理


谢 谢
www.huawei.com

SR基本原理

前言

- SR（Segment Routing）段路由作为SDN的关键技术，同时也是对IP网络倡导的极简协议的关键技术，在IP网络得到越来越多的应用，课程将会对SR技术及在5G承载网中的应用进行分析讲解。

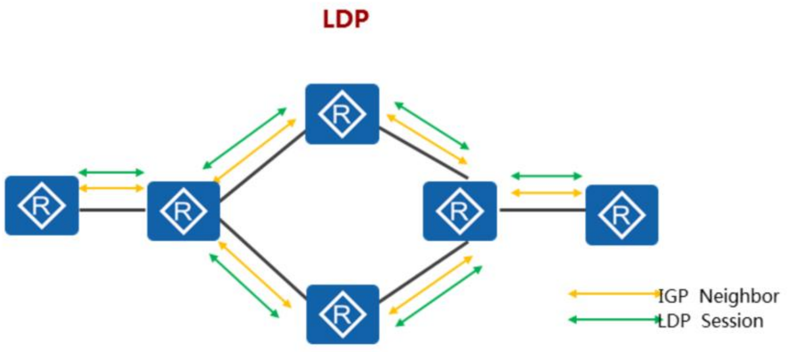
- SDN（Software-Defined Networking）:软件定义网络，是Emulex网络一种新型网络创新架构，其核心技术OpenFlow通过将网络设备控制面与数据面分离开来，从而实现了网络流量的灵活控制。

目标

- 学完本课程后，您应该能：
  - 理解SR的优势
  - 理解SR的的基本概念和工作原理
  - 了解SR技术的应用

目录

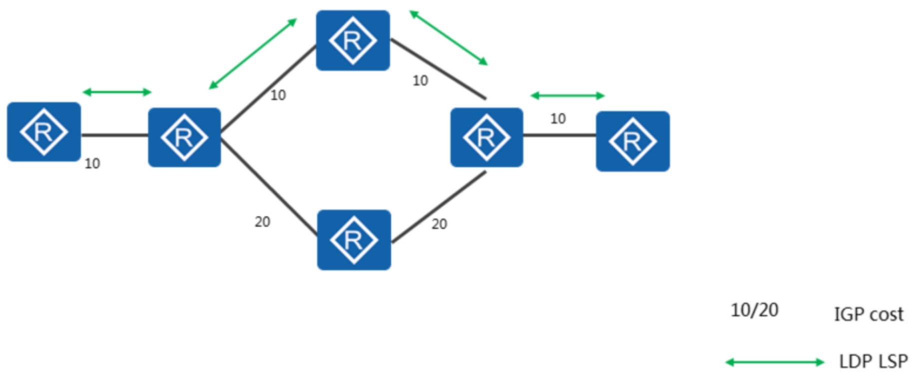
SR产生背景-LDP面临的问题（一）



- LDP协议依赖于IGP协议，使用本地标签交换，支持ECMP。
- 缺点：LDP本身有11种协议报文，在应用时大大增加了链路带宽的消耗和设备CPU利用率。

- LDP（Label Distribute Protocol）：标签分发协议，根据不同的目的地址来分配标签，使用标签转发代替了IP转发，其优点为可以隔离公网私网路由，跟随路由的最佳路径选择转发路径，支持ECMP，配置简单。
- 语音、数据、视频、VR等不同类型的业务对网络的要求不尽相同，对带宽的要求也不断增长，网络规模呈爆炸式增长。
- 传统MPLS技术需要专门的标签分发协议，需要每条LSP分配标签，资源占用大；状态维护协议报文占据大量带宽；需要与IGP协议同步，部署维护复杂，可扩展性差。这种网络的运营模式难以满足服务商按需快速部署网络业务的需要，而且OPEX（OPerating EXpense ）随着规模线性增长。

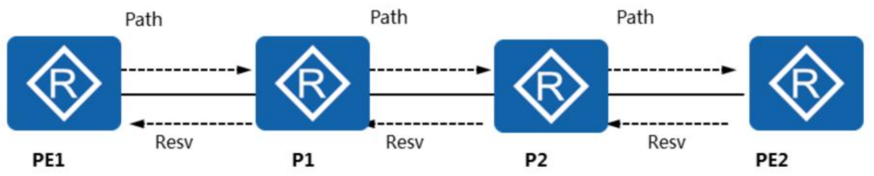
SR产生背景-LDP面临的问题（二）



- LDP只支持IGP最短路径（最小开销）进行路径计算，不支持TE(流量工程)。

- LDP进行LSP路径计算时依赖与IGP路由表

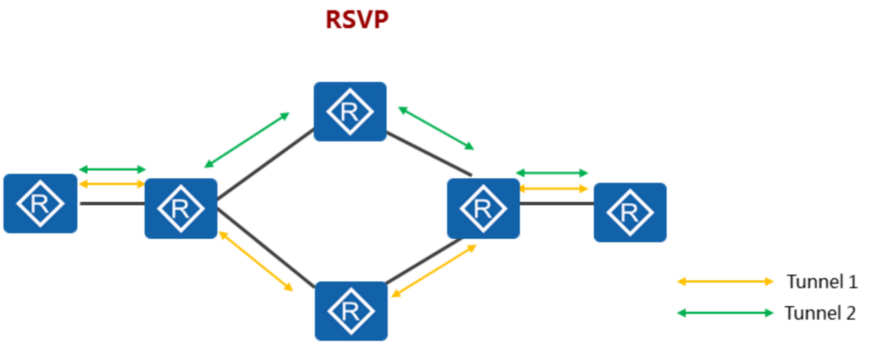
SR产生背景-RSVP面临的问题（一）



- RSVP依据CSPF算路结果逐跳形成端到端的LSP，标签为本地标签
- 逐跳，途经节点均会维护该Tunnel的状态信息，即便SDN化，仍需维护
- RSVP TE配置复杂
- ECMP实现复杂
- 8 种RSVP协议报文占据网络带宽和CPU处理

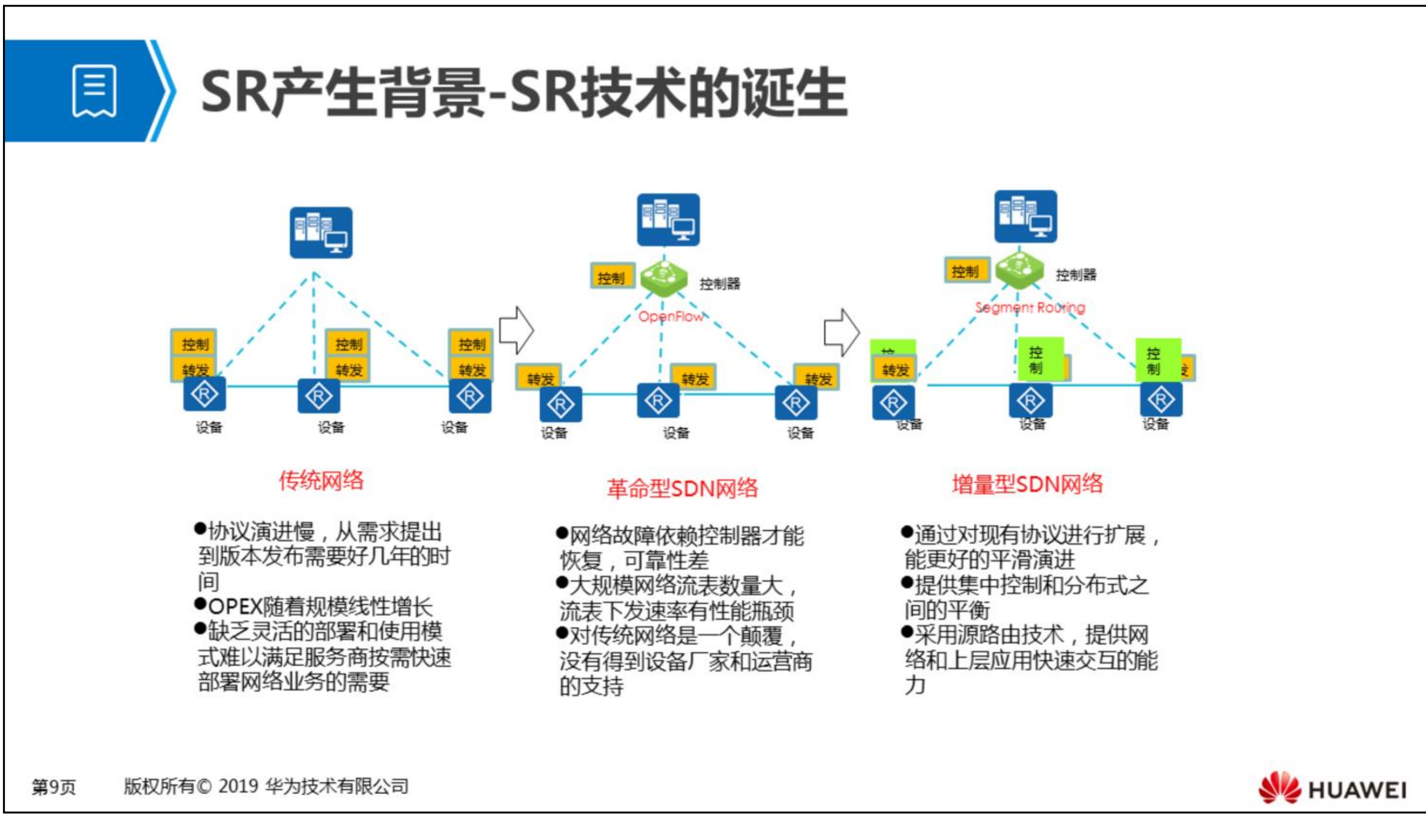
- RSVP(Resource Reservation Protocol)：资源预留协议，为了解决传统IP网络只能最优路径转发，无法规划路径的问题，引入了RSVP-TE。
- RSVP-TE引入带来了很多好处，可以做路径的显式规划、带宽资源预留以及多种的保护。
- RSVP依赖IGP建立后，会维护自己的邻居、链路状态，控制面复杂化，增加网络维护、问题定位复杂度。
- RSVP-TE需要建立多条隧道来完成ECMP功能。
- LSP状态维护需要PATH、RESV不断刷新，业务增多后，带来了中间点的性能问题（承受业务叠加），如图中的C/G。
- 而且RSVP-TE隧道的配置量，一直被运营商所诟病。平均一个隧道要8条配置。

SR产生背景-RSVP面临的问题（二）



- RSVP在负载分担方面实现复杂，需要配置多条Tunnel来实现，给配置和维护带来巨大的工作量。

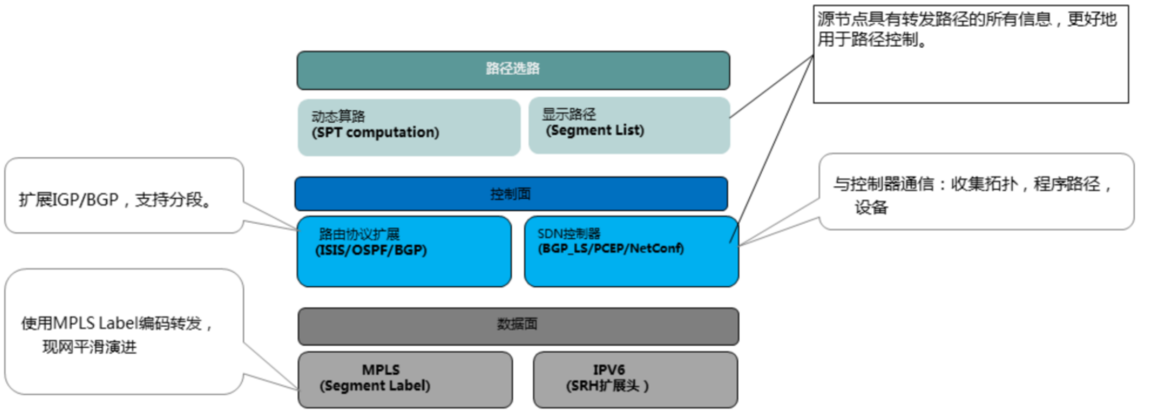
HUAWEI



- SDN（Software Defined Network），即软件定义网络，是由美国斯坦福大学clean slate研究组提出的一种新型网络创新架构。其核心技术是通过将网络**设备控制面**与**数据面**分离开来，从而实现了网络流量的灵活控制，为核心网络及应用的创新提供了良好的平台。
- 革命型SDN网络采用集中式控制架构，网络设备的控制功能(如路由计算)集中到一个控制器上去处理，转发表由控制器生成下发到设备上。设备功能大大简化，只负责转发，变成了一个傻瓜式的设备，Openflow是Controller和设备之间的控制接口。
- 增量型SDN网络是将现有网络进行拓展，现有设备向SDN演进思路，设备保留部分控制功能，把部分需要集中控制的功能放到控制器上去处理，强调设备平滑演进能力。

SR产生背景-SR的技术框架

- SR是一种只需在源（显式路径加载的节点）节点给报文增加一系列的段标识，便可指导报文转发的技术方案。



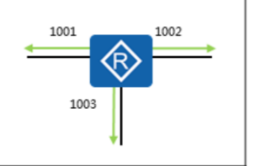
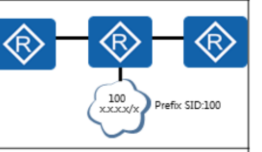
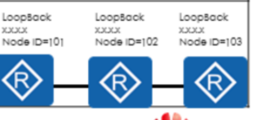
目录

SR基本概念

- **SR域（Segment Routing Domain）**：SR节点的集合。
- **Segment**：节点对入口报文执行的指令（如：依据最短路径转发报文到目的、或通过指定接口转发报文、或将报文转发至指定的应用/业务实例）。
- **SID（ Segment ID）**： Segment标识。Segment Routing将网络中的目的地址前缀/节点和邻接定义为一个个段，并且为这些目的地址前缀/节点和邻接分配段ID。段ID相当于传统MPLS技术中的MPLS标签，在转发层面映射为MPLS标签。
- **SRGB（segment routing global block）**： 为全局segment预留的本地标签集合。在MPLS中，SRGB为全局标签预留的本地标签集合，在IPv6中，SRGB为全局还未被使用IPv6的地址集合。
- **标签栈（Segment List）**：标签栈是目的地址前缀SID /节点SID和邻接SID有序列表的排序集合，用于标识一条完整的标签交换路径LSP(Label Switched Path)。在MPLS架构中为标签栈，封装于报文头中指导转发。

Segment

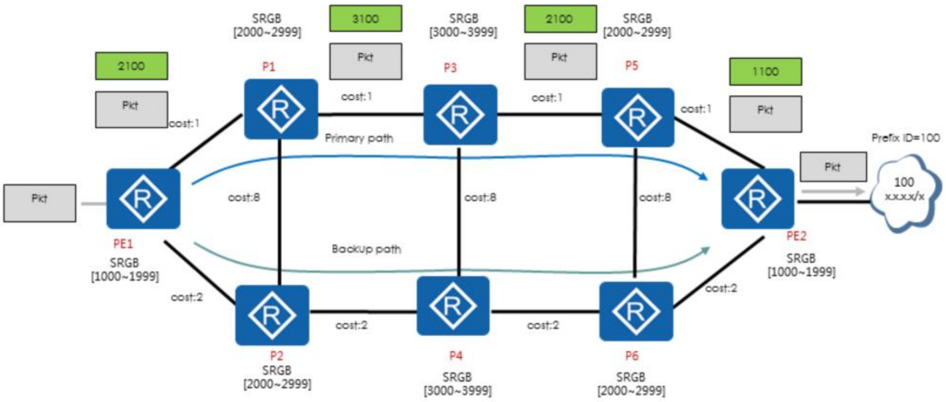
- SR的基本理念是把网络分为不同的段（Segment），然后拼接起来，用于指导报文按指定的路径转发。
- SID：即Segment ID，用来标识唯一的段。在转发层面，映射为MPLS标签。
- SR技术中三种基本Segment：邻接段，前辍段和节点段。

标签段	生成方式	作用	示例
Adjacency Segment（邻接段）	源节点通过协议动态分配	• Adjacency Segment用于标识网络中某节点的某个邻接。 • 由源节点IGP协议分配，只在本地有效，并且具有一定的方向性。 • Adjacency Segment通过IGP协议扩散到其他网元，全局可见，本地有效。 • 通过Adjacency Segment ID（SID）标识。Adjacency SID为SRGB范围外的本地SID。	
Prefix Segment（前缀段）	手工配置	• Prefix Segment用于标识网络中的某个目的地址前缀（Prefix）。 • Prefix Segment通过IGP协议扩散到其他网元，全局可见，全局有效。 • Prefix SID是源端发布的SRGB范围内的偏移值，接收端会根据自己的SRGB计算实际标签值用于生成MPLS转发表项。	
Node Segment（节点段）	手工配置	Node Segment是特殊的Prefix Segment，用于标识特定的节点（Node）。在节点的Loopback接口下配置IP地址作为前缀，这个节点的Prefix SID实际就是Node SID。	

- Prefix Segment代表目的地址，Adjacency Segment代表数据包的外发链路，可以分别类似于传统IP转发中的目的IP地址和出接口。在IGP区域内，网元设备使用扩展IGP消息将自身的Node SID以及Adjacency SID进行泛洪，这样任意一个网元都可以获得其他网元的信息。
- 通过按序组合前缀（节点）SID和邻接SID，可以构建出网络内的任何路径。在路径中的每一跳，使用栈顶段信息区分下一跳。段信息按照顺序堆叠在数据头的顶部。当栈顶段信息包含另一个节点的标识时，接收节点使用等价多路径（ECMP）将数据包转发到下一跳。当栈顶段信息是本节点的标识时，接收节点弹出顶部段并执行下一个段所需的任务。
- 实际应用中Adjacency Segment、Prefix Segment、Node Segment可以单独使用，也可以结合使用。
- 主要有如下三种场景：**Prefix Segment，Adjacency Segment，Adjacency Segment + Node Segment**

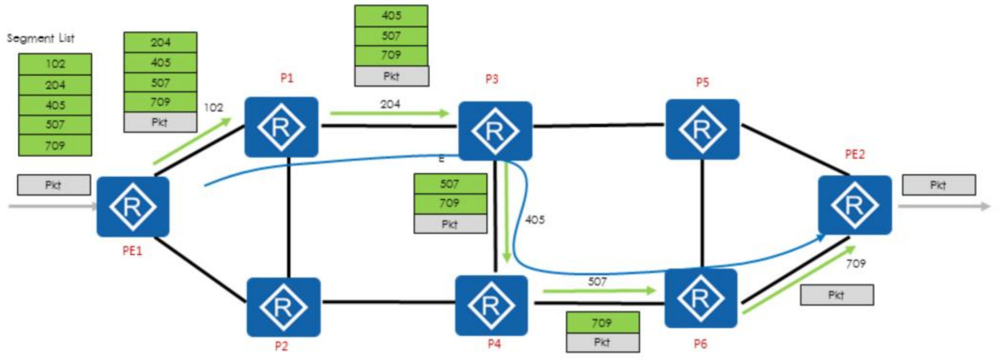
Segment —Prefix Segment

- 基于Prefix Segment：由IGP使用SPF算法计算最短路径，所以也称为SR-BE（Best Effort）。
- 如下图所示，以节点Z为目的节点，其Prefix SID是100。通过IGP扩散之后，整个IGP域的所有设备学习到节点PE2的Prefix SID，之后都会使用SPF算法得出一条到节点PE2的最短路径。



Segment —Adjacency Segment

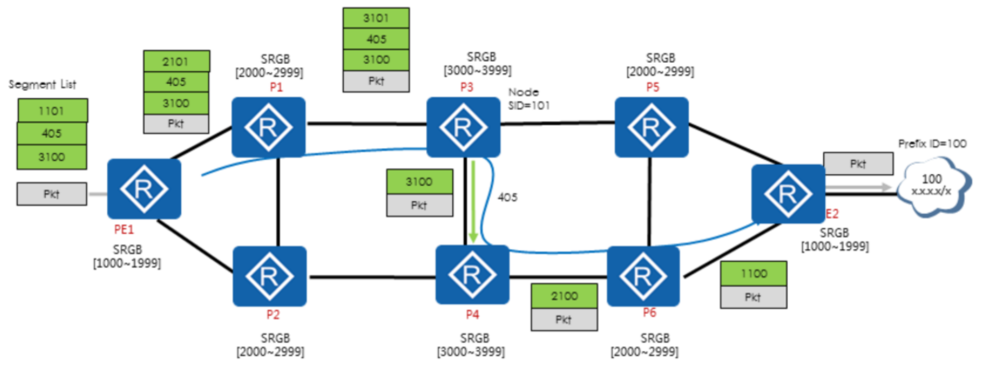
- 基于Adjacency Segment：头节点指定严格显式路径（Strict Explicit）。这种方式可以集中进行路径调整和流量调优，因此可以更好的配合实现SDN（Software-Defined Networking）。Adjacency Segment主要用于SR-TE（Traffic Engineering）。



  HUAWEI

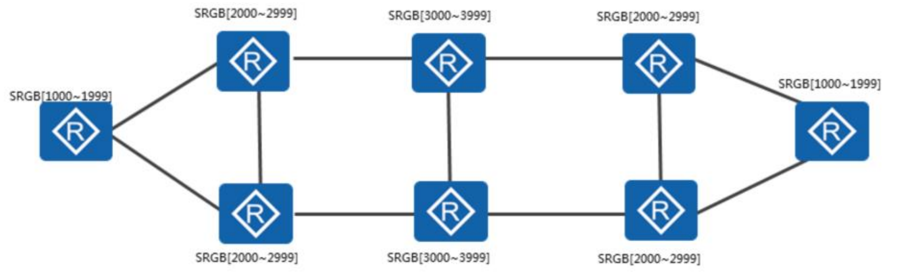
Segment —Adjacency Segment + Node Segment

- 基于Adjacency Segment +Node Segment：显式路径与最短路径相结合，称作松散路径（Loose Explicit）。主要用于SR-TE。



  HUAWEI

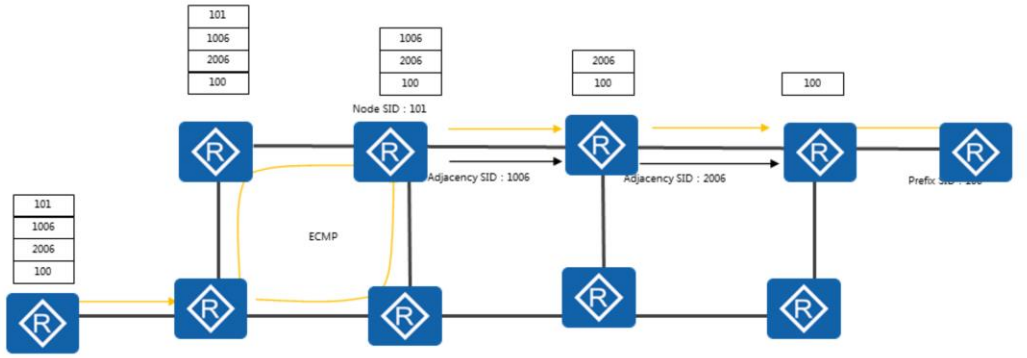
SRGB



- SRGB是从本地标签资源隔离出来的一段区间，专门用于SR，以使SR的全局标签和传统MPLS在本地共存。
- 为节点SID基础，全局唯一的节点标签与本地标签不得冲突：
- SRGB范围和起始值需配置，在起始值基础上对节点SID的索引值进行偏移，得出本地标签
- 各设备上的SRGB范围为何不配置为一样？
  - 标准中未统一规定SRGB范围，各设备商实现不同
  - 与传统MPLS共存时，各设备上空余的标签空间不能保证一样

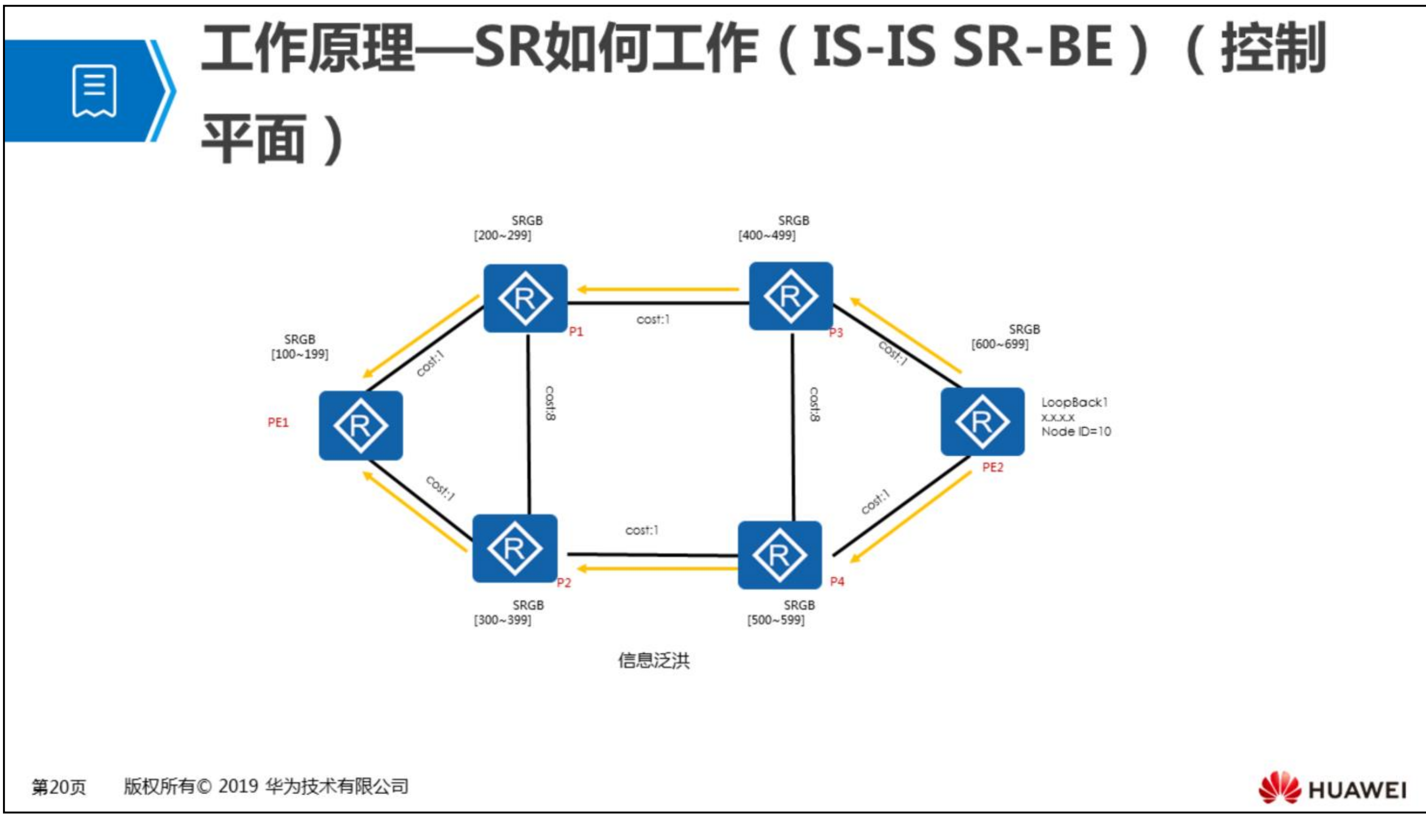
目录

SR的原理



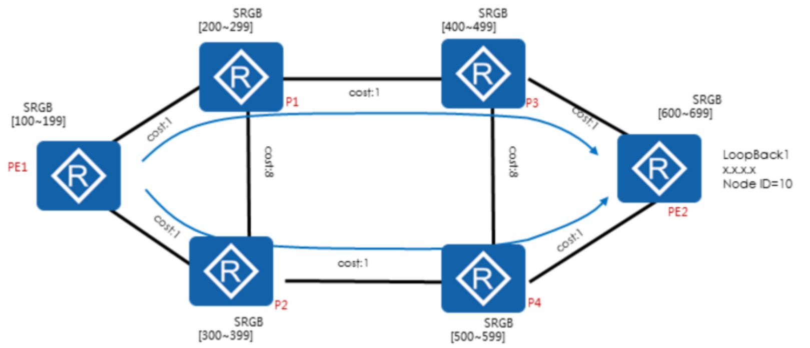
- SR（Segment Routing）段路由是对现有IGP协议进行扩展，基于MPLS协议，采用源路由技术而设计的在网络上转发数据包的一种协议。
- SR将网络中的目的地址前缀/节点和邻接定义为段，并且为这些段分配SID（Segment ID）。通过对Adjacency SID（邻接段）和Prefix/Node SID（目的地址前缀/节点段）进行有序排列（Segment List），就得到一条转发路径。

- Segment Routing将代表转发路径的段序列封装在数据包头部，随数据包传输。网络节点收到数据包后，对段序列进行解析，如果段序列的顶部段标识是Node SID，根据SPF计算的最短路径（如果存在等价路径，则可以实现ECMP）转发到该节点，如果是Adjacency SID，则根据Adjacency SID转发到下一节点，直到数据报文到达目的节点。



- ISIS：是与OSPF并列的两大IGP协议之一。也采用链路状态算法。
- **以图中PE1到PE2的LSP为例：**
- 1、Egress PE2设备上LoopBack1接口的环回地址为x.x.x.x/x，为该地址分配的SID为10，并将该信息泛洪到整个IS-IS域。

工作原理—SR如何工作（IS-IS SR-BE）（控制平面）



PE1→P1→P3→PE2路径上各节点为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

PE1→P2→P4→PE2路径上各节点为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4>PE2
PE2	610	NA	NA

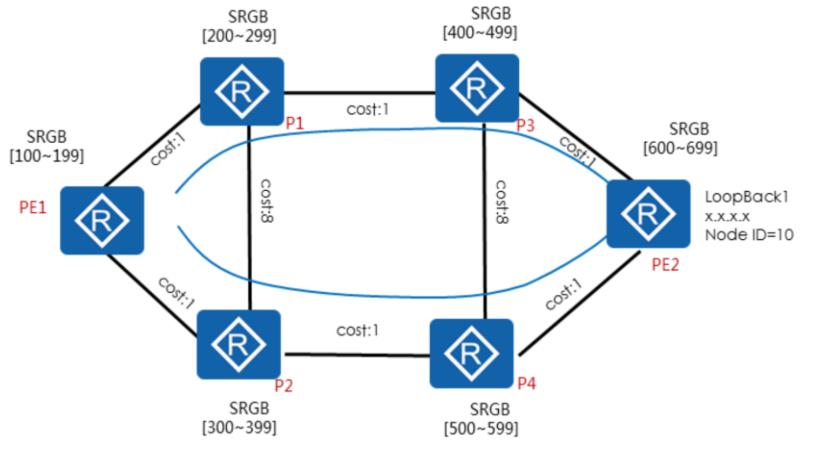
- 2、所有节点收到PE2的Node SID，生成标签转发表：

  -入标签：本地SRGB起始值+发布的偏移值

  -出标签：下一跳SRGB起始值+发布的偏移值

  -出接口下一跳：IGP计算的最短路径的出接口下一跳

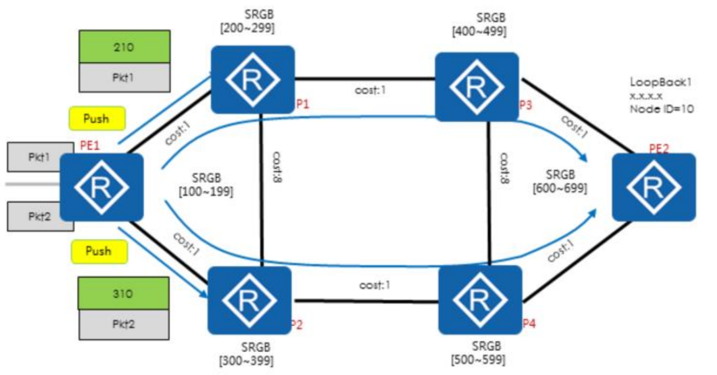
工作原理—SR如何工作（IS-IS SR-BE）（控制平面）



  HUAWEI

- 3、Ingress PE1进行IS-IS SPF计算，得到ECMP到PE2的两条最短路径。

工作原理—SR如何工作（IS-IS SR-BE）（转发平面）



PE1→P1→P3→PE2路径上各节点上为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

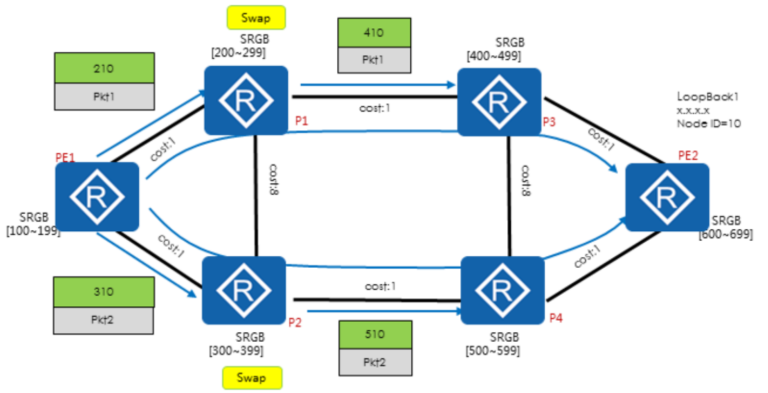
PE1→P2→P4→PE2路径上各节点为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4>PE2
PE2	610	NA	NA

- SR的标签操作类型和MPLS相同，包括标签栈压入（Push）、标签栈交换（Swap）和标签弹出（Pop）。

- **以图中PE1到PE2的SR-BE Tunnel为例：**
- 1、**Ingress节点PE1：**接收到业务报文，依据目的地址、标签转发表和ECMP分流算法，封装下一跳P节点分配的出标签，依据目的前缀SID和标签转发表从对应的接口转发出去。
  - 举例：
  - Push 210 to P1
  - Push 310 to P2

工作原理—SR如何工作（IS-IS SR-BE）（转发平面）



PE1→P1→P3→PE2路径上各节点为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

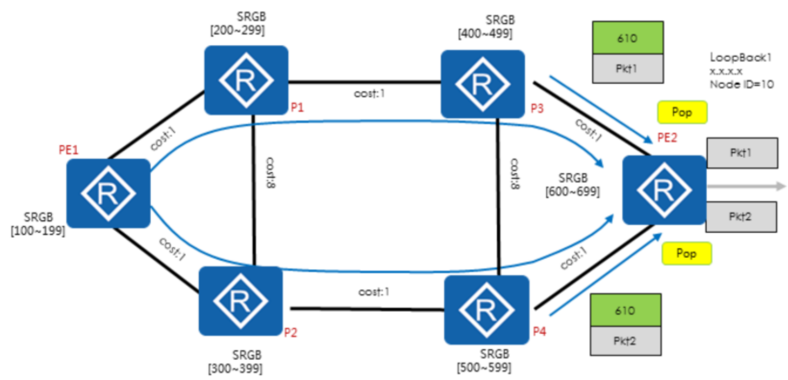
PE1→P2→P4→PE2路径上各节点为Node SID 10生成的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4>PE2
PE2	610	NA	NA

  HUAWEI

- 2、**Transit节点：**依据外层标签及标签转发表，交换出标签为下一跳分配的SR出标签，并从对应的接口转发出去。
  - 举例：
  - Swap 210 to 410 on P1
  - Swap 310 to 510 on P2

工作原理—SR如何工作（IS-IS SR-BE）（转发平面）



PE1→P1→P3→PE2路径上各节点为Node SID 10生成的的标签转发表项

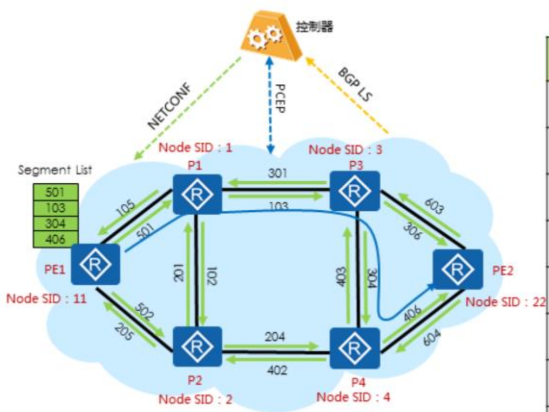
Node	InLabel	Outlabel	Interface
PE1	110	210	PE1->P1
P1	210	410	P1->P3
P3	410	610	P3->PE2
PE2	610	NA	NA

PE1→P2→P4→PE2路径上各节点为Node SID 10生成的的标签转发表项

Node	InLabel	Outlabel	Interface
PE1	110	310	PE1->P2
P2	310	510	P2->P4
P4	510	610	P4>PE2
PE2	610	NA	NA

- 3、**Egress节点PE2：**发现外层标签是自己，弹出外层标签，依据外层地址将业务报文转发给CE设备。
  - 举例：
  - Pop 610 from P3
  - Pop 610 from P4

工作原理—SR如何工作（SR-TE）（控制平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

控制器上的标签表

Node	Adjacency	label
PE1	NA	1011
PE1	Link PE1 to P1	501
PE1	Link PE1 to P2	502
P1	NA	1001
P1	Link P1 to PE1	105
P1	Link P1 to P2	102
P1	Link P1 to P3	103
......	......	......

  HUAWEI

1、标签、标签转发表和拓扑信息生成

网络中各节点通过IS-IS SR协议为其邻接分配邻接SID，泛洪到整个网络中，并为其分配的本地邻接SID生成标签转发表项。同时生成带有邻接SID信息的拓扑信息。

另外，手动配置的SRGB/前缀SID/节点SID、 SR能力、SR算法等信息通过IGP协议报文在IGP域内泛洪。每个节点运行IS-IS SPF来计算每个节点标签的最短转发路径，并生成标签转发表。

2、标签和拓扑信息上报

网络中各节点通过BGP-LS（BGP Link-State）将带有邻接SID信息的拓扑信息上报给控制器。

3、路径计算

控制器采用PCEP（Path Computation Element）完成标签转发路径计算。

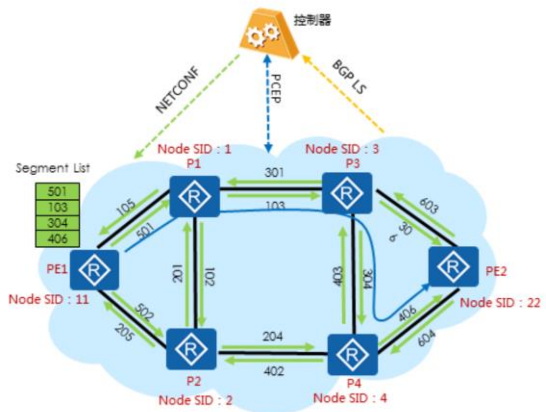
4、路径下发

控制器通过PCEP下发隧道信息、通过Netconf下发隧道属性信息给隧道首节点。 隧道首节点通过PCEP上报隧道状态给控制器。

5、隧道创建

隧道首节点根据控制器下发的标签栈建立SR-TE 隧道。

工作原理—SR如何工作（SR-TE）（控制平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

PE1上的标签转发表项

SID/Inlabel	Outlabel	Interface
1011	NA	NA
501	NA	PE1->P1
502	NA	PE1>P2
1001	1001	PE1->P1
1002	1002	PE1->P2
1003	1003	PE1->P1
1004	1004	PE1->P2
1022	1022	PE1->P1
1022	1022	PE1->P2

PE1上的SR-TE Tunnel标签栈封装表格

SR-TE Tunnel	Segment List
Tunnel 1	501/103/304/406
......	......

  HUAWEI

1、标签、标签转发表和拓扑信息生成

网络中各节点通过IS-IS SR协议为其邻接分配邻接SID，泛洪到整个网络中，并为其分配的本地邻接SID生成标签转发表项。同时生成带有邻接SID信息的拓扑信息。

另外，手动配置的SRGB/前缀SID/节点SID、 SR能力、SR算法等信息通过IGP协议报文在IGP域内泛洪。每个节点运行IS-IS SPF来计算每个节点标签的最短转发路径，并生成标签转发表。

2、标签和拓扑信息上报

网络中各节点通过BGP-LS（BGP Link-State）将带有邻接SID信息的拓扑信息上报给控制器。

3、路径计算

控制器采用PCEP（Path Computation Element）完成标签转发路径计算。

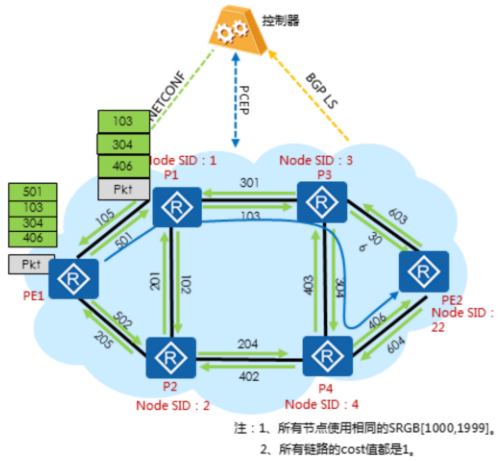
4、路径下发

控制器通过PCEP下发隧道信息、通过Netconf下发隧道属性信息给隧道首节点。 隧道首节点通过PCEP上报隧道状态给控制器。

5、隧道创建

隧道首节点根据控制器下发的标签栈建立SR-TE 隧道。

工作原理—SR如何工作（SR-TE Strict Explicit Path）（转发平面）



SR设备根据报文头中SR-TE隧道对应的标签栈，对报文进行标签操作。根据栈顶标签逐跳查找转发出接口，指导数据报文转发到隧道目的地址。

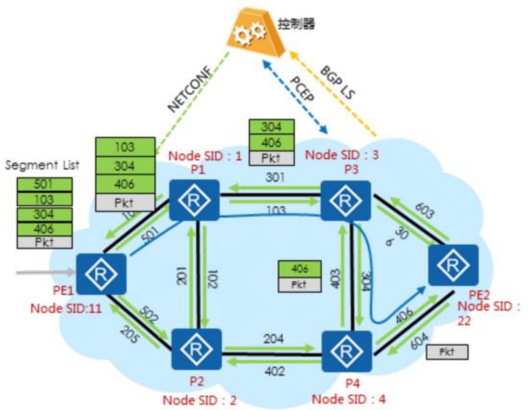
图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

  HUAWEI

- **以图中PE1到PE2的严格路径SR-TE Tunnel为例：**
- 1、**Ingress节点PE1：**接收到业务报文，依据业务配置的路由策略或隧道策略确定的SR-TE Tunnel，封装SR-TE Tunnel对应的标签栈。判断对应的栈顶标签501为邻接SID，则弹出栈顶标签501，封装标签栈其余部分[103,304,406]到业务报文，并依据栈顶标签和标签转发表从对应接口PE1->P1转发出去。

工作原理—SR如何工作（SR-TE Strict Explicit Path）（转发平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

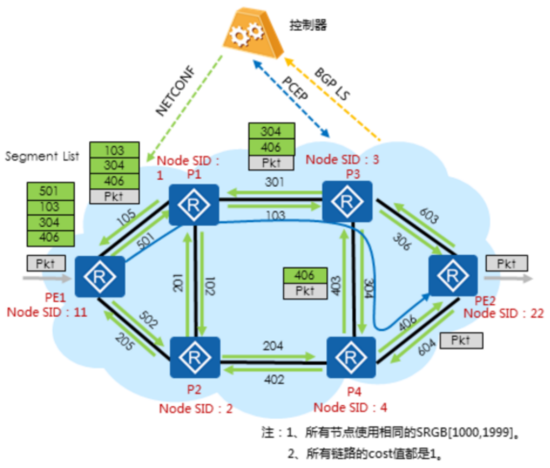
图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

HUAWEI

- 2、**Transit节点：**判断栈顶标签为邻接SID，则弹出栈顶标签，并依据栈顶标签及标签转发表从对应接口转发出去。
- 以P1节点为例：判断栈顶标签103为邻接SID，则弹出栈顶标签103，并依据栈顶标签103及标签转发表从对应接口P1->P3转发出去。

工作原理—SR如何工作（SR-TE Strict Explicit Path）（转发平面）

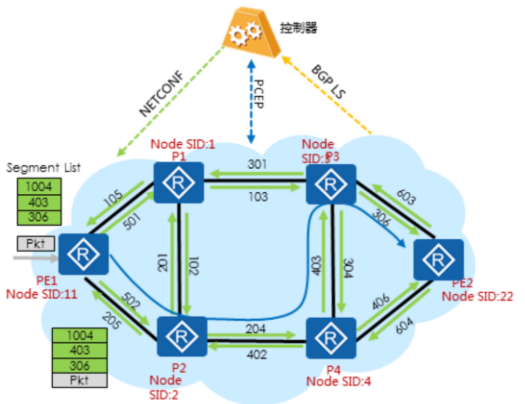


图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	501	NA	PE1->P1
P1	103	NA	P1->P3
P3	304	NA	P3->P4
P4	406	NA	P4->PE2

- 3、**Egress节点PE2：**依据业务报文外层地址将业务报文转发给CE设备。

工作原理—SR如何工作（SR-TE Loose Explicit Path）（转发平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

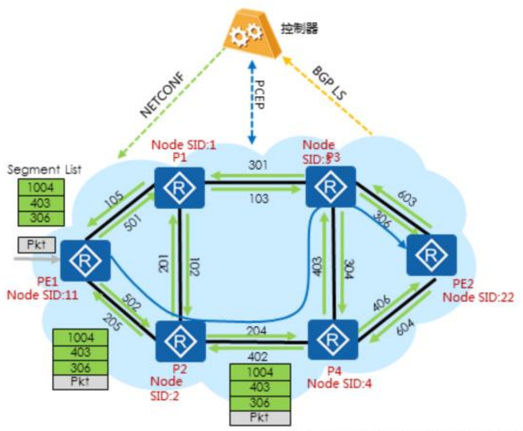
  HUAWEI

- SR设备根据报文头中SR-TE隧道对应的标签栈，对报文进行标签操作。根据栈顶标签逐跳查找转发出接口，指导数据报文转发到隧道目的地址。

- **以图中PE1到PE2的松散路径SR-TE Tunnel为例：**

- 1、**Ingress节点 PE1**：接收到业务报文，依据业务配置的路由策略或隧道策略确定的SR-TE Tunnel，封装SR-TE Tunnel对应的标签栈[1004,403,306] 。依据外层标签和标签转发表封装下一跳P节点分配的SR出标签1004，并依据标签转发表从对应接口PE1->P2转发出去。

工作原理—SR如何工作（SR-TE Loose Explicit Path）（转发平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

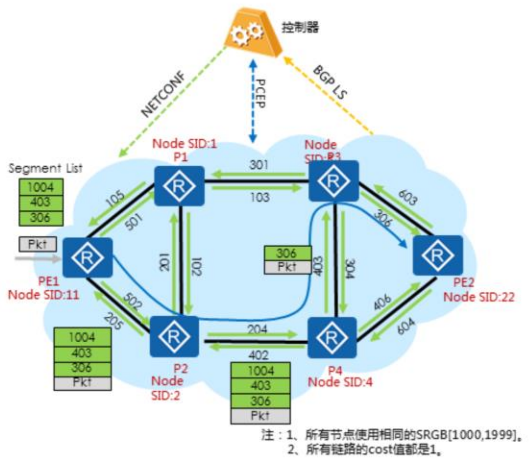
图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

  HUAWEI

2、 **Transit节点P2：**依据外层标签1004和标签转发表，交换外层标签为下一跳分配的出标签1004，并从对应的接口P2->P4转发出去。

工作原理—SR如何工作（SR-TE Loose Explicit Path）（转发平面）

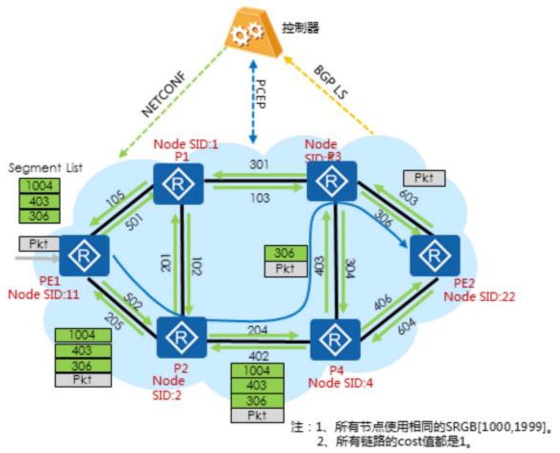


图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

3、 **Transit节点P4：**发现外层标签1004是自己，弹出外层标签1004。判断下一层标签403为邻接SID，将其弹出，并依据标签转发表从对应接口P4->P3转发出去。

工作原理—SR如何工作（SR-TE Loose Explicit Path）（转发平面）

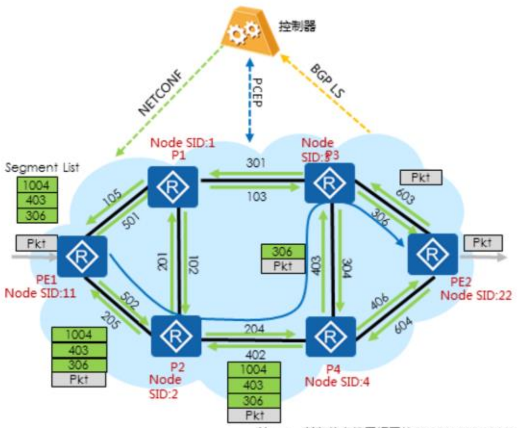


图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

4、Transit节点P3：判断外层标签306为邻接标签，弹出外层标签306，并依据标签转发表从对应接口P3->PE2转发出去。

工作原理—SR如何工作（SR-TE Loose Explicit Path）（转发平面）



注：1、所有节点使用相同的SRGB[1000,1999]。
2、所有链路的cost值都是1。

图中PE1到PE2的SR-TE Tunnel途经各节点上的标签转发表项

Node	SID/Inlabel	Outlabel	Interface
PE1	1004	1004	PE1->P2
P2	1004	1004	P2->P4
P4	1004	NA	NA
P4	403	NA	P4->P3
P3	306	NA	P3->PE2

5、**Egress节点PE2：**依据业务报文外层地址将业务报文转发给CE设备。

思考题

1. SR隧道类型有哪些（ ）。
   - A. SR-BE最短路径隧道
   - B. SR-TE严格显示路径隧道
   - C. SR-BE严格显示路径隧道
   - D. SR-TE松散显示路径隧道
2. 基于SR标签转发报文时，数据包可能需要执行的动作有哪些（ ）。
   - A. Push
   - B. Pop
   - C. Swap
   - D. Switch

- 参考答案：
- 1：ABD
- 2：ABC

本章总结

- SR产生背景
- SR基本概念
- SR技术架构
- SR工作原理


谢 谢
www.huawei.com

学习推荐

- 华为培训与认证官方网站
  - http://learning.huawei.com/cn/
- 华为在线学习
  - https://ilearningx.huawei.com/portal/#/portal/ebg/26
- 华为职业认证
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=zh
- 查找培训入口
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=zh

更多信息

- 华为培训APP