

路由协议基础

文档版本

01

发布日期

2020-11-18



版权所有 © 华为技术有限公司 2020。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址：深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址：<https://www.huawei.com>

客户服务邮箱：support@huawei.com

客户服务电话：4008302118

目 录

1 简介.....1

2 路由迭代.....3

3 路由器及路由基本原理.....4

4 静态路由与动态路由..... 5

5 路由表和 FIB 表..... 6

6 路由协议的优先级..... 10

7 路由的度量..... 12

8 负载分担与路由备份..... 13

1 简介

路由（Routing）是数据通信网络中一个基本的概念。路由就是通过互联的网络把信息从源地址传输到目的地址的活动。路由发生在OSI网络参考模型中的第三层（即网络层）。我们将具有路由转发功能的设备称为广义上的路由器。

当路由器收到一个IP数据包，路由器会根据目的IP地址在设备上的路由表（Routing Table）中进行查找，找到“最匹配”的路由条目后，将数据包根据路由条目所指示的出接口或下一跳IP转发出去。路由表中装载着路由器通过各种途径获知的路由条目（Routes）。路由器可通过静态、动态等方式获取路由条目并维护自己的路由表。

什么是路由协议

较小的网络通常可以手动设置路由表（即静态方式），但较大且拥有复杂拓扑的网络可能常常变化，若要手动创建、维护路由表是不切实际的。因此，人们希望路由器可以动态的（即动态方式）按照某种协议来自动创建维护路由表以解决这个问题，从而使得网络能够近自主的适应变化，避免故障。这些协议被称为路由协议。

路由协议的分类

路由协议可以有多种分类方式，常见的分类方式如下：

- 按照路由协议使用的算法分：
 - 距离矢量路由协议，例如：RIP（Routing Information Protocol）
 - 链路状态路由协议，例如：OSPF（Open Shortest Path First）
- 按照路由协议作用的区域划分：
 - 内部网关协议（Interior Gateway Protocol），在单一的自治系统中交换路由信息，例如：OSPF（Open Shortest Path First）
 - 外部网关协议（Exterior Gateway Protocol），在不同的自治系统中交换路由信息，例如：BGP（Border Gateway Protocol）

常见的路由协议有哪些

目前常用的路由协议有：

- OSPF（Open Shortest Path First）
- RIP（Routing Information Protocol）
- IS-IS（Intermediate System to Intermediate System）

- BGP (Border Gateway Protocol)

其中，最多被使用到的路由协议为OSPF和BGP。

2 路由迭代

路由必须有直连的下一跳才能够指导转发，但是路由生成时下一跳可能不是直连的，因此需要计算出一个直连的下一跳和对应的出接口，这个过程就叫做路由迭代。BGP路由、静态路由和UNR路由的下一跳都有可能不是直连的，都需要进行路由迭代。

例如，BGP路由的下一跳一般是非直连的对端loopback地址，不能指导转发，需要进行迭代。即根据以BGP学习到的下一跳为目的地址在IP路由表中查找，当找到一条具有直连的下一跳、出接口信息的路由后（一般为一条IGP路由），将其下一跳、出接口信息填入这条BGP路由的IP路由表中并生成对应的FIB表项。

对于BGP私网路由，需要隧道进行转发，路由的下一跳一般是远端PE的Loopback地址，不能指导转发，也需要进行路由迭代，即在隧道列表中查找到达该Loopback地址的隧道，将该隧道信息填入路由表中并生成对应的FIB表项。

3 路由器及路由基本原理

在因特网中，网络连接设备用来控制网络流量和保证网络数据传输质量。常见的网络连接设备有集线器（Hub）、网桥（Bridge）、交换机（Switch）和路由器（Router）。这些设备的基本原理类似，下面就以路由器为例来介绍一下设备的基本原理。

路由器是一种典型的网络连接设备，用来进行路由选择和报文转发。路由器根据收到报文的目地址选择一条合适的路径（包含一个或多个路由器的网络），然后将报文传送到下一个路由器，路径终端的路由器负责将报文送交目的主机。

路由就是报文从源端到目的端的路径。当报文从路由器到目的网段有多条路由可达时，路由器可以根据路由表中最佳路由进行转发。最佳路由的选取与发现此路由的路由协议的优先级、路由的度量有关。当多条路由的协议优先级与路由度量都相同时，可以实现负载分担，缓解网络压力；当多条路由的协议优先级与路由度量不同时，可以构成路由备份，提高网络的可靠性。

4 静态路由与动态路由

静态路由与动态路由的区别

路由协议是路由器之间维护路由表的规则，用于发现路由，生成路由表，并指导报文转发。依据来源的不同，路由可以分为三类：

- 通过链路层协议发现的路由称为直连路由。
- 通过网络管理员手动配置的路由称为静态路由。
- 通过动态路由协议发现的路由称为动态路由。

静态路由配置方便，对系统要求低，适用于拓扑结构简单并且稳定的小型网络。缺点是不能自动适应网络拓扑的变化，需要人工干预。

动态路由协议有自己的路由算法，能够自动适应网络拓扑的变化，适用于具有一定数量三层设备的网络。缺点是配置对用户要求比较高，对系统的要求高于静态路由，并将占用一定的网络资源和系统资源。

动态路由的分类

对动态路由协议的分类可以采用以下不同标准：

根据作用范围不同，路由协议可分为：

- 内部网关协议IGP（Interior Gateway Protocol）：在一个自治系统内部运行。常见的IGP协议包括RIP、OSPF和IS-IS。
- 外部网关协议EGP（Exterior Gateway Protocol）：运行于不同自治系统之间。BGP是目前最常用的EGP协议。

根据使用算法不同，路由协议可分为：

- 距离矢量协议（Distance-Vector Protocol）：包括RIP和BGP。其中，BGP也被称为路径矢量协议（Path-Vector Protocol）。
- 链路状态协议（Link-State Protocol）：包括OSPF和IS-IS。

以上两种算法的主要区别在于发现路由和计算路由的方法不同。

5 路由表和 FIB 表

路由器转发数据包的关键是路由表和FIB表，每个路由器都至少保存着一张路由表和一张FIB（Forwarding Information Base）表。路由器通过路由表选择路由，通过FIB表指导报文进行转发。

路由表

每台路由器中都保存着一张本地核心路由表（即设备的IP路由表），同时各个路由协议也维护着自己的路由表。

- 本地核心路由表

路由器使用本地核心路由表用来保存决策优选路由，并负责把优选路由下发到FIB表，通过FIB表指导报文进行转发。这张路由表依据各种路由协议的优先级和度量值来选取路由。

说明

对于支持L3VPN（Layer 3 Virtual Private Network）的路由器，每一个VPN-Instance拥有一个自己的管理路由表（本地核心路由表）。

- 协议路由表

协议路由表中存放着该协议发现的路由信息。

路由协议可以引入并发布其他协议生成的路由。例如，在路由器上运行OSPF协议，需要使用OSPF协议通告直连路由、静态路由或者IS-IS路由时，要将这些路由引入到OSPF协议的路由表中。

路由表中的内容

在路由器中，执行命令**display ip routing-table**时，可以查看路由器的路由表概要信息，如下所示：

```
<HUAWEI> display ip routing-table
Proto: Protocol      Pre: Preference
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole route
-----
Routing Table: _public_
      Destinations : 14      Routes : 14

Destination/Mask    Proto   Pre  Cost   Flags NextHop        Interface
-----
0.0.0.0/0           Static  60   0       RD   10.137.216.1      Vlanif20
10.10.10.0/24       Direct  0     0       D    10.10.10.10       Vlanif20
```

```

10.10.10.10/32 Direct 0 0 D 127.0.0.1 InLoopBack0
10.10.10.255/32 Direct 0 0 D 127.0.0.1 InLoopBack0
10.10.11.0/24 Direct 0 0 D 10.10.11.1 LoopBack0
10.10.11.1/32 Direct 0 0 D 127.0.0.1 InLoopBack0
10.10.11.255/32 Direct 0 0 D 127.0.0.1 InLoopBack0
10.137.216.0/23 Direct 0 0 D 10.137.217.208 Vlanif20
10.137.217.208/32 Direct 0 0 D 127.0.0.1 InLoopBack0
10.137.217.255/32 Direct 0 0 D 127.0.0.1 InLoopBack0
127.0.0.0/8 Direct 0 0 D 127.0.0.1 InLoopBack0
127.0.0.1/32 Direct 0 0 D 127.0.0.1 InLoopBack0
127.255.255.255/32 Direct 0 0 D 127.0.0.1 InLoopBack0
255.255.255.255/32 Direct 0 0 D 127.0.0.1 InLoopBack0

```

路由表中包含了下列关键项：

- **Destination**：表示此路由的目的地址。用来标识IP包的目的地址或目的网络。
- **Mask**：表示此目的地址的子网掩码长度。与目的地址一起来标识目的主机或路由器所在的网段的地址。
将目的地址和子网掩码“逻辑与”后可得到目的主机或路由器所在网段的地址。
例如：目的地址为10.1.1.1，掩码为255.255.255.0的主机或路由器所在网段的地址为10.1.1.0。
掩码由若干个连续“1”构成，既可以用点分十进制表示，也可以用掩码中连续“1”的个数来表示。例如掩码255.255.255.0长度为24，即可以表示为24。
- **Proto**：表示学习此路由的路由协议。
- **Pre**：表示此路由的路由协议优先级。针对同一目的地，可能存在不同下一跳、出接口等多条路由，这些不同的路由可能是由不同的路由协议发现的，也可以是手工配置的静态路由。优先级高（数值小）者将成为当前的最优路由。各协议路由优先级请参见[6 路由协议的优先级](#)。
- **Cost**：路由开销。当到达同一目的地的多条路由具有相同的路由优先级时，路由开销最小的将成为当前的最优路由。

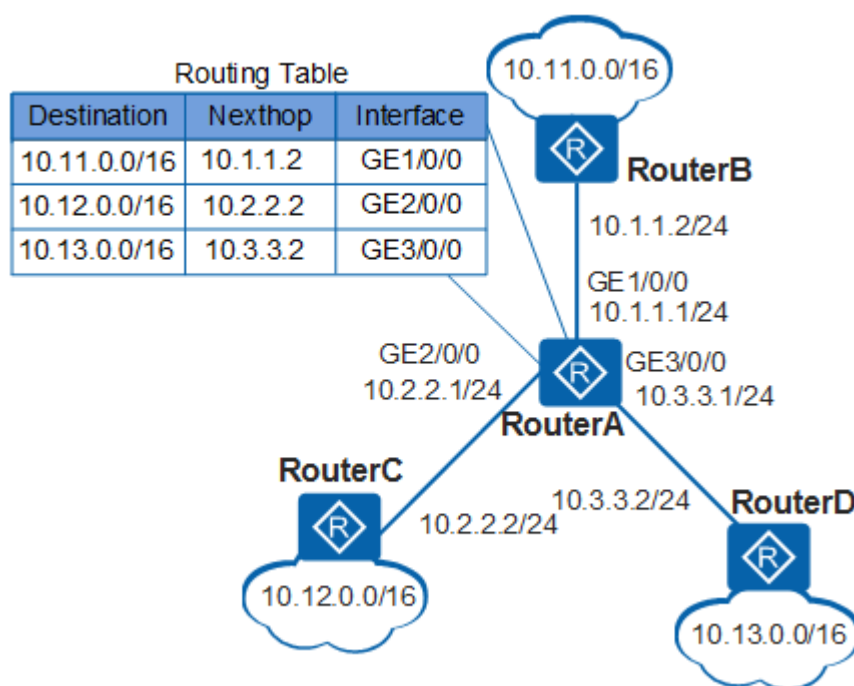
说明

Preference用于不同路由协议间路由优先级的比较，Cost用于同一种路由协议内部不同路由的优先级的比较。

- **NextHop**：表示此路由的下一跳地址。指明数据转发的下一个设备。
- **Interface**：表示此路由的出接口。指明数据将从本地路由器哪个接口转发出去。

在[图5-1](#)所示的网络中，路由器A与三个网络相连，因此有三个IP地址和三个出接口，其路由表如图所示。

图 5-1 路由表示意图



路由超限自动恢复

本地核心路由表里保存着各路由协议的路由，如果本地核心路由表里的路由数量达到系统上限，协议路由表将无法向本地核心路由表添加路由。本地核心路由表有以下几种路由限制：

- 整机路由限制：指定所有路由条数的上限值。
- 整机路由前缀限制：指定所有路由的地址前缀范围。
- 组播IGP路由限制：指定组播IGP路由条数的上限值。
- 多拓扑路由限制：指定多拓扑路由条数的上限值。
- 所有私网路由限制：指定所有私网路由条数的上限值。
- VPN路由限制：指定VPN路由条数的上限值。
- VPN路由前缀限制：指定VPN路由的地址前缀范围。

如果协议由于某种路由限制而向本地核心路由表添加路由失败，系统会记录本次添加路由的协议和对应的路由表Table ID。

当协议删除本地核心路由表里的路由释放了路由表的空间之后，路由超限解除，系统会通知所有向本地核心路由表添加路由失败的协议，重新向本地核心路由表添加路由，使得本地核心路由表中的路由能够得到最大程度的恢复。是否可以完全恢复，取决于释放的路由表空间的大小。

FIB 表的匹配

在路由表选择出路由后，路由表会将激活路由下发到FIB表中。当报文到达路由器时，会通过查找FIB表进行转发。

FIB表中每条转发项都指明到达某网段或某主机的报文应通过路由器的哪个物理接口或逻辑接口发送，然后就可到达该路径的下一个路由器，或者不再经过别的路由器而传送到直接相连的网络中的目的主机。

FIB表的匹配遵循最长匹配原则。查找FIB表时，报文的目的地址和FIB中各表项的掩码进行按位“逻辑与”，得到的地址符合FIB表项中的网络地址则匹配。最终选择一个最长匹配的FIB表项转发报文。

例如，一台路由器上的路由表如下：

Routing Tables:

| Destination/Mask | Proto | Pre | Cost | Flags | NextHop | Interface |
|------------------|--------|-----|------|-------|-------------|----------------------|
| 0.0.0.0/0 | Static | 60 | 0 | D | 192.168.0.2 | GigabitEthernet1/0/0 |
| 10.8.0.0/16 | Static | 60 | 3 | D | 192.168.0.2 | GigabitEthernet1/0/0 |
| 10.9.0.0/16 | Static | 60 | 50 | D | 172.16.0.2 | GigabitEthernet3/0/0 |
| 10.9.1.0/24 | Static | 60 | 4 | D | 192.168.0.2 | GigabitEthernet2/0/0 |
| 10.20.0.0/16 | Direct | 0 | 0 | D | 172.16.0.1 | GigabitEthernet4/0/0 |

一个目的地址是10.9.1.2的报文进入路由器，查找对应的FIB表。

FIB Table:

Total number of Routes : 5

| Destination/Mask | Nexthop | Flag | TimeStamp | Interface | TunnelID |
|------------------|-------------|------|-----------|----------------------|----------|
| 0.0.0.0/0 | 192.168.0.2 | SU | t[37] | GigabitEthernet1/0/0 | 0x0 |
| 10.8.0.0/16 | 192.168.0.2 | DU | t[37] | GigabitEthernet1/0/0 | 0x0 |
| 10.9.0.0/16 | 172.16.0.2 | DU | t[9992] | GigabitEthernet3/0/0 | 0x0 |
| 10.9.1.0/24 | 192.168.0.2 | DU | t[9992] | GigabitEthernet2/0/0 | 0x0 |
| 10.20.0.0/16 | 172.16.0.1 | U | t[9992] | GigabitEthernet4/0/0 | 0x0 |

首先，目的地址10.9.1.2与FIB表中各表项的掩码“0、16、24”作“逻辑与”运算，得到下面的网段地址：0.0.0.0/0、10.9.0.0/16、10.9.1.0/24。这三个结果可以匹配到FIB表中对应的三个表项。最终，路由器会选择最长匹配10.9.1.0/24表项，从接口GE2/0/0转发这条目的地址是10.9.1.2的报文。

6 路由协议的优先级

对于相同的目的地，不同的路由协议（包括静态路由）可能会发现不同的路由，但这些路由并不都是最优的。事实上，在某一时刻，到某一目的地的当前路由仅能由唯一的路由协议来决定。为了判断最优路由，各路由协议（包括静态路由）都被赋予了一个优先级，当存在多个路由信息源时，具有较高优先级（取值较小）的路由协议发现的路由将成为最优路由，并将最优路由放入本地路由表中。

路由器分别定义了外部优先级和内部优先级。其中，0表示直接连接的路由，255表示任何来自不可信源端的路由；数值越小表明优先级越高。外部优先级是指用户可以手工为各路由协议配置的优先级，缺省情况下如表6-1所示。

表 6-1 路由协议缺省时的外部优先级

| 路由协议的类型 | 路由协议的外部优先级 |
|-----------|------------|
| Direct | 0 |
| OSPF | 10 |
| IS-IS | 15 |
| Static | 60 |
| RIP | 100 |
| OSPF ASE | 150 |
| OSPF NSSA | 150 |
| IBGP | 255 |
| EBGP | 255 |

路由协议的内部优先级则不能被用户手工修改，如表6-2所示。

表 6-2 路由协议内部优先级

| 路由协议的类型 | 路由协议的内部优先级 |
|---------------|------------|
| Direct | 0 |
| OSPF | 10 |
| IS-IS Level-1 | 15 |
| IS-IS Level-2 | 18 |
| Static | 60 |
| RIP | 100 |
| OSPF ASE | 150 |
| OSPF NSSA | 150 |
| IBGP | 200 |
| EBGP | 20 |

选择路由时先比较路由的外部优先级，当不同的路由协议配置了相同的优先级后，系统会通过内部优先级决定哪个路由协议发现的路由将成为最优路由。例如，到达同一目的地10.1.1.0/24有两条路由可供选择，一条静态路由，另一条是OSPF路由，且这两条路由的外部优先级都被配置成5。这时路由器系统将根据表6-2所示的内部优先级进行判断。因为OSPF协议的内部优先级是10，高于静态路由的内部优先级60。所以系统选择OSPF协议发现的路由作为最优路由。

7 路由的度量

路由的度量标示出了这条路由到达指定的目的地址的代价，通常以下因素会影响到路由的度量。

- 路径长度

路径长度是最常见的影响路由度量的因素。链路状态路由协议可以为每一条链路设置一个链路开销来标示此链路的路径长度。在这种情况下，路径长度是指经过的所有链路的链路开销的总和。距离矢量路由协议使用跳数来标示路径长度。跳数是指数据从源端到目的端所经过的设备数量。例如，路由器到与它直接相连网络的跳数为0，通过一台路由器可达的网络的跳数为1，其余以此类推。

- 网络带宽

网络带宽是一个链路实际的传输能力。例如，一个10千兆的链路要比1千兆的链路更优越。虽然带宽是指一个链路能达到的最大传输速率，但这不能说明在高带宽链路上路由要比低带宽链路上更优越。比如说，一个高带宽的链路正处于拥堵的状态下，那报文在这条链路上转发时将会花费更多的时间。

- 负载

负载是一个网络资源的使用程度。计算负载方法包括CPU的利用率和它每秒处理数据包的数量。持续监测这些参数可以及时了解网络的使用情况。

- 通信开销

通信开销衡量了一条链路的运营成本。尤其是只注重运营成本而不在乎网络性能的时候，通信开销则就成了一个重要的指标。

8 负载分担与路由备份

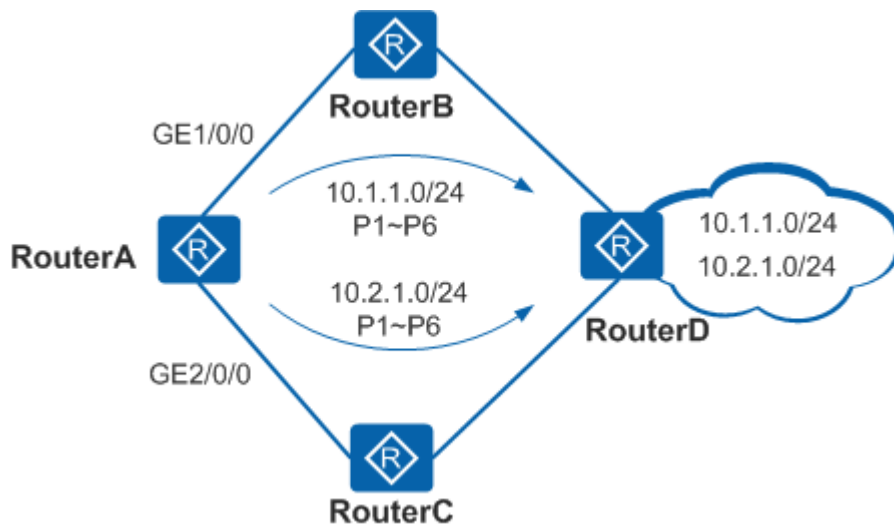
当多条路由的路由优先级和路由度量都相同时，这几条路由就称为等价路由，多条等价路由可以实现负载分担。当这几条路由为非等价路由时，就可以实现路由备份。

负载分担

路由器支持多路由模式，即允许配置多条目的地相同且优先级也相同的路由。当到达同一目的地存在同一路由协议发现的多条路由时，且这几条路由的开销值也相同，那么就满足负载分担的条件。

当实现负载分担时，路由器根据五元组（源地址、目的地址、源端口、目的端口、协议）进行转发，当五元组相同时，路由器总是选择与上一次相同的下一跳地址发送报文。当五元组不同时，路由器会选取相对空闲的路径进行转发。

图 8-1 负载分担组网图



如图8-1所示，RouterA已经通过接口GE1/0/0转发到目的地址10.1.1.0/24的第1个报文P1，随后又需要分别转发报文到目的地址10.1.1.0/24和10.2.1.0/24。其转发过程如下：

- 当转发到达10.1.1.0/24的第2个报文P2时，发现此报文与到达10.1.1.0/24的第1个报文P1的五元组一致。所以之后到达该目的地的报文都从GE1/0/0转发。
- 当转发到达10.2.1.0/24的第1个报文P1时，发现此报文与到达10.1.1.0/24的第1个报文P1的五元组不一致。所以选取从GE2/0/0转发，并且之后到达该目的地的报文都从GE2/0/0转发。

路由备份

路由备份功能，可以提高网络的可靠性。用户可以根据实际情况，配置到同一目的地的多条路由，其中一条路由的优先级最高，作为主路由，其余的路由优先级较低，作为备份路由。

正常情况下，路由器采用主路由转发数据。当主链路出现故障时，主路由变为非激活状态，路由器选择备份路由中优先级最高的路由转发数据。这样，也就实现了主路由到备份路由的切换。当主链路恢复正常时，由于主路由的优先级最高，路由器重新选择主路由来发送数据。这样，就实现了从备份路由回切到主路由。