

路由控制与转发

为了使路由转发效率更高,系统使用了路由的控制平面和转发平面。路由的控制平面负责路由计算和维护,而路由的转发平面负责 IP 数据报文的转发。本章介绍控制平面和转发平面的关系,并对路由表、FIB 表的生成和作用进行了详解。

3.1 本章目标

学习完本章,应该能够:

- (1) 了解控制平面和转发平面的区别;
- (2) 了解 FIB 表项作用与生成过程;
- (3) 掌握快速转发工作原理;
- (4) 掌握快速转发的配置。

3.2 路由的控制平面与转发平面

路由器、交换机承担着路由学习、MAC 地址学习、数据报文转发等重要的工作,其系统的稳定性是非常重要的。系统的设计者们尽力从架构上使系统工作稳定、可靠,其中很重要的一点就是控制平面和转发平面相对独立,以减少相互影响,如图 3-1 所示。本章将学习控制平面、转发平面的定义、作用及相关表项内容和配置。

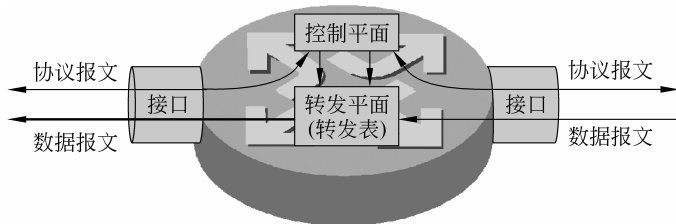


图 3-1 路由的控制平面与转发平面

(1) 控制平面。控制平面是指系统中用来传送信令、计算表项的部分。诸如协议报文收发、协议表项计算、维护等都属于控制平面的范畴。例如,在路由系统中,负责路由协议学习、路由表项维护的进程就属于控制平面;而在交换系统中,负责 MAC 地址学习的进程则属于控制平面。

(2) 转发平面。转发平面是指系统中用来进行数据报文的封装、转发的部分。诸如数据报文的接收、解封装、封装、转发等都属于转发平面的范畴。例如,系统接收到 IP 报文后,需要进行解封装、查路由表、从出接口转发等,系统中负责以上行为的进程则属于转发平面。

控制平面与转发平面相对独立又协同工作。系统的控制平面进行协议交互、路由计算后,生成若干表项,下发到转发平面,指导转发平面对报文进行转发。例如,路由器通过 OSPF 协

议建立了路由表项,再进一步生成 FIB(Forwarding Information Base)表、快速转发表等,指导系统进行 IP 报文转发。

良好的系统设计应该是使控制平面与转发平面尽量分离,互不影响。当系统的控制平面暂时出现故障时,转发平面还可以继续工作,这样可以保证网络中原有的业务不受系统故障的影响,提高整个网络的可靠性。

控制平面与转发平面分离的另外好处是系统设计能够做到模块化,易于维护,便于扩展。例如,原有系统仅能够支持 RIP 协议,伴随着网络设备数量的增长,RIP 协议不能够满足需求。系统开发者可以在不改变转发平面的基础上升级控制平面至支持 OSPF 协议,减少升级的代价。

控制平面与转发平面可以是物理分离,也可以是逻辑分离。高端设备如核心交换机、核心路由器,一般采用控制平面与转发平面物理分离。其主控板上的 CPU 不负责报文转发,而专注于系统的控制;而业务板则专注于数据报文转发。如果主控板损坏,业务板仍然能够转发报文。而对于入门级的网络设备,受限于成本,一般只做到逻辑分离。即设备启动后,系统将 CPU 和内存资源划分给不同的进程,有的进程负责路由学习,有的进程负责报文转发。

在 H3C 的系列路由器中,核心路由器 SR8800、SR6600 是典型的控制平面与转发平面物理分离的架构。以 SR8800 为例,其主控板上有 CPU,负责路由计算、协议交互、管理维护等功能;有专门的硬件转发芯片,芯片有很强的数据交换能力,负责在业务板间进行数据转发。业务板也有 CPU,并通过管理通道连接到主控板的 CPU 上;业务上还集成了专门负责数据转发的硬件芯片。

如果业务板的端口收到邻居发送的路由协议报文,则业务板的 CPU 通过管理通道将其上送到主控板的 CPU,由主控板的 CPU 进行相应的处理。如果有路由更新,则由主控板 CPU 生成相应的 FIB 表,下发到业务板的芯片中。

当业务板的端口收到邻居发送来的数据报文后,根据报文中的目的 IP 地址,进行 FIB 查表操作,如果查表发现目的主机位于本业务板上,则直接由业务板的芯片转发到另外一个物理端口上;如果发现目的主机位于其他业务板上,则由业务板的芯片通过数据通道转发到主控板的硬件转发芯片,再转发到其他业务板上。

通过以下流程可以看出,控制平面与转发平面是物理分开的,其中一个平面出现故障,不会影响另一个平面的运行。

3.3 路由表和 FIB 表

3.3.1 路由表转发

当路由表中存在多个路由项可以同时匹配目的 IP 地址时,路由查找进程会选择其中掩码最长的路由项用于转发,此为最长匹配原则。

在图 3-2 中,路由器接收到目的地址为 40.0.0.2 的数据包,经查找整个路由表,发现与路由 40.0.0.0/24 和 40.0.0.0/8 都能匹配。但根据最长匹配的原则,路由器会选择路由项 40.0.0.0/24,根据该路由项转发数据包。

由以上过程可知,路由表中路由项数量越多,所需查找及匹配的次数则越多,其转发效率就越低。

Comware 平台提供了丰富的命令来查看路由表信息。可以使用 display ip routing-table 命令来查看路由表中当前激活路由的摘要信息。其输出如下:

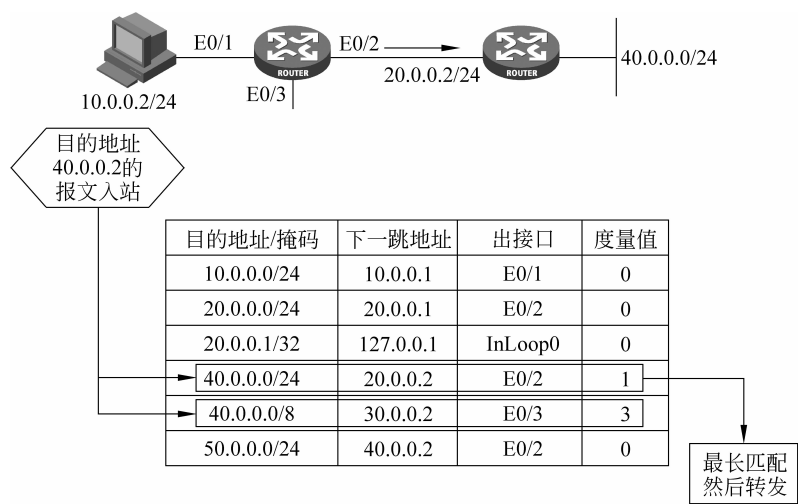


图 3-2 路由表转发

[H3C] display ip routing-table

Destinations : 16		Routes : 16			
Destination/Mask	Proto	Pre	Cost	NextHop	Interface
0.0.0.0/32	Direct	0	0	127.0.0.1	InLoop0
1.1.1.1/32	O_INTRA	10	2	20.0.0.1	GE0/0
2.2.2.2/32	O_INTRA	10	1	20.0.0.1	GE0/0
3.3.3.3/32	Direct	0	0	127.0.0.1	InLoop0
10.0.0.0/24	O_INTRA	10	2	20.0.0.1	GE0/0
20.0.0.0/24	Direct	0	0	20.0.0.2	GE0/0
20.0.0.0/32	Direct	0	0	20.0.0.2	GE0/0
20.0.0.2/32	Direct	0	0	127.0.0.1	InLoop0
20.0.0.255/32	Direct	0	0	20.0.0.2	GE0/0
127.0.0.0/8	Direct	0	0	127.0.0.1	InLoop0
127.0.0.0/32	Direct	0	0	127.0.0.1	InLoop0
127.0.0.1/32	Direct	0	0	127.0.0.1	InLoop0
127.255.255.255/32	Direct	0	0	127.0.0.1	InLoop0
224.0.0.0/4	Direct	0	0	0.0.0.0	NULL0
224.0.0.0/24	Direct	0	0	0.0.0.0	NULL0
255.255.255.255/32	Direct	0	0	127.0.0.1	InLoop0

其中各参数含义如表 3-1 所示。

表 3-1 display ip routing-table 命令显示信息描述表

字 段	描 述
Destinations	目的地址的个数
Routes	路由条数
Destination/Mask	目的地址/掩码长度
Proto	发现该路由的路由协议

续表

字 段	描 述
Pre	路由的优先级
Cost	路由的度量值
NextHop	此路由的下一跳地址
Interface	出接口,即到该目的网段的数据包将从此接口发出

如果想查看当前所有的路由,包括处于 Active 和 Inactive 状态的路由及其详细描述,可以使用命令 `display ip routing-table verbose`。其输出如下:

```
<h3c>display ip routing-table verbose
```

```
Destinations : 13          Routes : 13
```

```
Destination: 0.0.0.0/32
```

```

  Protocol: Direct          Process ID: 0
  SubProtID: 0x0           Age: 08h34m37s
    Cost: 0                Preference: 0
    Tag: 0                 State: Active NoAdv
  OrigTblID: 0x0           OrigVrf: default-vrf
  TableID: 0x2             OrigAs: 0
    NBRID: 0x10000000      LastAs: 0
  AttrID: 0xffffffff       Neighbor: 0.0.0.0
    Flags: 0x1000c         OrigNextHop: 127.0.0.1
    Label: NULL            RealNextHop: 127.0.0.1
  BkLabel: NULL            BkNextHop: N/A
  Tunnel ID: Invalid       Interface: InLoopBack0
  BkTunnel ID: Invalid     BkInterface: N/A
```

```
Destination: 1.1.1.0/24
```

```

  Protocol: Static          Process ID: 0
  SubProtID: 0x0           Age: 04h20m37s
    Cost: 0                Preference: 60
    Tag: 0                 State: Active Adv
  OrigTblID: 0x0           OrigVrf: default-vrf
  TableID: 0x2             OrigAs: 0
    NBRID: 0x10000003      LastAs: 0
  AttrID: 0xffffffff       Neighbor: 0.0.0.0
    Flags: 0x1008c         OrigNextHop: 192.168.47.4
    Label: NULL            RealNextHop: 192.168.47.4
  BkLabel: NULL            BkNextHop: N/A
  Tunnel ID: Invalid       Interface: Ethernet1/1
  BkTunnel ID: Invalid     BkInterface: N/A
```

...(省略部分显示信息)

其中各参数含义如表 3-2 所示。

表 3-2 display ip routing-table verbose 命令显示信息描述表

字 段	描 述
Destinations	目的地址的个数
Routes	路由条数
Destination	目的地址/掩码
Protocol	发现该路由的路由协议类型
Process ID	进程号
SubProtID	路由子协议 ID
Age	此路由在路由表中存在的时间
Cost	路由的度量值
Preference	路由的优先级
Tag	路由标记
State	路由状态描述： Active：有效的单播路由 Adv：允许对外发送的路由 Inactive：非激活路由标志 NoAdv：不允许发布的路由 Vrrp：VRRP 产生的路由 Nat：NAT 产生的路由 TunE：Tunnel 隧道的标志
OrigTblID	原始路由表 ID
OrigVrf	路由所属的原始 VPN
TableID	路由所在路由表的 ID
OrigAs	初始 AS 号
NBRID	邻居 ID
LastAs	最后 AS 号
AttrID	路由属性 ID 号
Neighbor	路由协议的邻居地址
Flags	路由标志位
OrigNextHop	此路由的下一跳地址
Label	标签
RealNextHop	路由真实下一跳
BkLabel	备份标签
BkNextHop	备份下一跳地址
Tunnel ID	隧道 ID
Interface	出接口，即到该目的网段的数据包将从此接口发出
BkTunnel ID	备份隧道 ID
BkInterface	备份出接口

在大规模网络路由中，网络设备的路由表项可能会十分庞大。如果采用以上命令，则系统会输出全部路由表项的信息。此时，如果想查看指定目的地址的路由信息，可以使用 display ip routing-table ip-address 命令。其输出如下：