

三层交换机报文转发过程解析

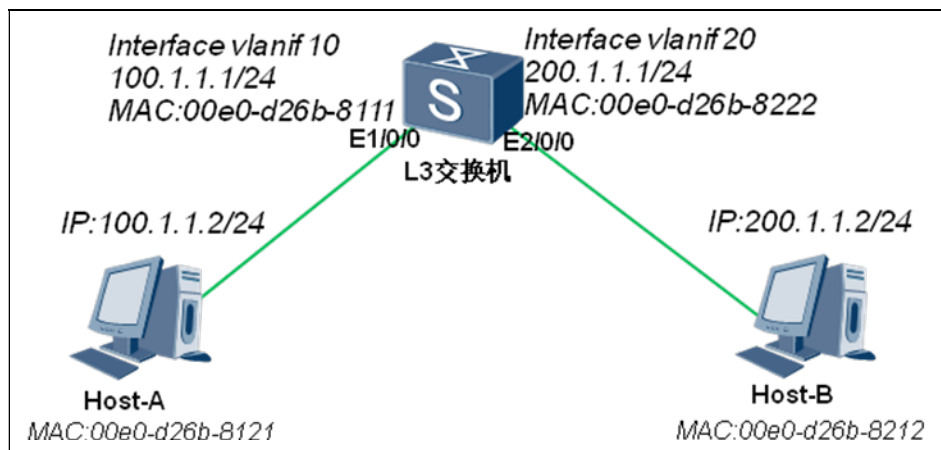


图 1

如图 1 所示，假如主机 A 想访问主机 B，首先主机 A 会将自己的 IP 地址和子网掩码做与操作，得出网路地址（如：Host-A 的 IP 地址 100.1.1.2 与自身掩码 255.255.255.0 做与操作后，得到的网络号是 100.1.1.0）。然后判断目的 IP 地址（即 Host-B 的 IP 地址）与自己的网络地址是不是在同一个子网。因为图中主机 A 和主机 B 不在同一子网内，所以需要进行三层转发。

1、 主机 A 发送 ARP 广播获取网关 MAC 地址

主机 A 想访问主机 B 首先要有主机 B 的 MAC 地址，由于主机 A 和主机 B 不在同一子网，所以主机 A 首先会向缺省网关发送 ARP 广播报文来获取网关的 MAC 地址。ARP 报文格式如下：

Ethernet 头		ARP 头			
D-MAC	S-MAC	S-MAC	S-IP	D-MAC	D-IP
ff:ff:ff:ff:ff:ff	00e0-d26b-8121	00e0-d26b-8121	100.1.1.2	00:00:00:00:00:00	100.1.1.1

2、 交换机形成主机 A 的 MAC 表项，并用网关 MAC 地址回应主机 A 的 ARP 请求

交换机收到 ARP 广播报文后,首先学习 ARP 报文 *Ethernet* 头部的源 MAC 地址，交换机芯片将自动记录主机 A 的 MAC 地址（00e0-d26b-8121）、接收该 ARP 报文的交换机接口号（E1/0/0）及此接口所属的 VLAN(VLAN 10)等信息,并形成一条 MAC 表项放入交换机 MAC 表中.同时，交换机也会通过软件把主机 A 的 IP、MAC、上连到交换机的接口等信息保存到交换机的硬件转发表里（三层硬件表项,MAC 表是没有 IP 的）。

由于主机 A 发送的 ARP 广播报文中的目的 IP 地址(100.1.1.1)就是交换机上接收该 ARP 广播报文的接口(E1/0/0)所属 VLAN(VLAN 10)的 IP 地址，所以交换机将使用 vlan10 的 MAC 地址回复主机 A 的 ARP 请求。ARP 回复报文如下：

Ethernet 头		ARP 头			
D-MAC	S-MAC	S-MAC	S-IP	D-MAC	D-IP
00e0-d26b-8121	00e0-d26b-8111	00e0-d26b-8111	100.1.1.1	00e0-d26b-8121	100.1.1.2

3、 主机 A 把网关 MAC 当作主机 B 的 MAC 访问主机 B

主机 A 收到网关的 ARP 回应报文后，会把网关的 MAC 地址当成是主机 B 的 MAC 地址，这样主机 A 发送数据给主机 B 时就会使用网关 MAC 作为目的 MAC 来封装数据帧,帧格式如下：

Ethernet 头		IP 头		数据
D-MAC	S-MAC	S-IP	D-IP	DATA
00e0-d26b-8111	00e0-d26b-8121	100.1.1.2	200.1.1.2	

4、 交换机查找硬件转发表/路由表进行三层转发

交换机收到主机 A 发来的数据报文后，仍然会首先学习数据报文 Ethernet 头部的源 MAC 地址，然后根据 Ethernet 头部的目的 MAC 查找交换机的 MAC 表，此时发现目的 MAC 地址就是本地 VLAN 的 MAC 地址，这种情况下交换机会把该报文上送到交换芯片的三层引擎处理。三层引擎将首先查找硬件转发表。

(1) 如果硬件转发表中有匹配项，则根据报文目的 IP 地址作相应处理：

①、目的 IP 地址就是本地的 IP 地址，则交相应模块处理。

②、目的 IP 是其他设备 IP 地址,本地只是转发，则根据硬件转发表项使用出接口 MAC 替换报文的源 MAC 地址，用下一跳 MAC 替换报文的目的 MAC 地址，同时 TTL 值减 1，继续转发。（此处与二层的区别在于需要对报文进行源、目的 MAC 进行替换）

(2) 如果不匹配则查找路由表，有匹配项则按照上面（2）中的操作进行，同时把相应的目的 IP、下一跳 MAC、出接口等信息存储到硬件转发表项中，下次就不需要查路由表了，这就是所谓的一次路由多次交换。

(3) 如果路由表也没有匹配项，则丢弃报文。

此处，由于主机 A 数据报文的目的 IP 是主机 B 的地址，并且主机 B 和交换机直连，交换机的硬件转发表/路由表已经有主机 B 的路由表项，所以交换机将主机 A 发来的报文中的源、目的 MAC 地址作替换，

同时 TTL 值减 1，然后发给主机 B。经过 MAC 替换后的报文格式如下：

Ethernet 头		IP 头		数据
D-MAC	S-MAC	S-IP	D-IP	DATA
00e0-d26b-8212	00e0-d26b-8222	100.1.1.2	200.1.1.2	

这样主机 A 就把数据报文发到了不同网段的主机 B。同时交换机上也保存了关于主机 A 和主机 B 的硬件转发表项，以后主机 A 和主机 B 互访，以及其他网段主机访问主机 A 或主机 B 交换机就可以根据硬件转发表项直接转发，而不需要查找路由表。