# 三层交换机基础

第一节 简介

第二节 交换原理

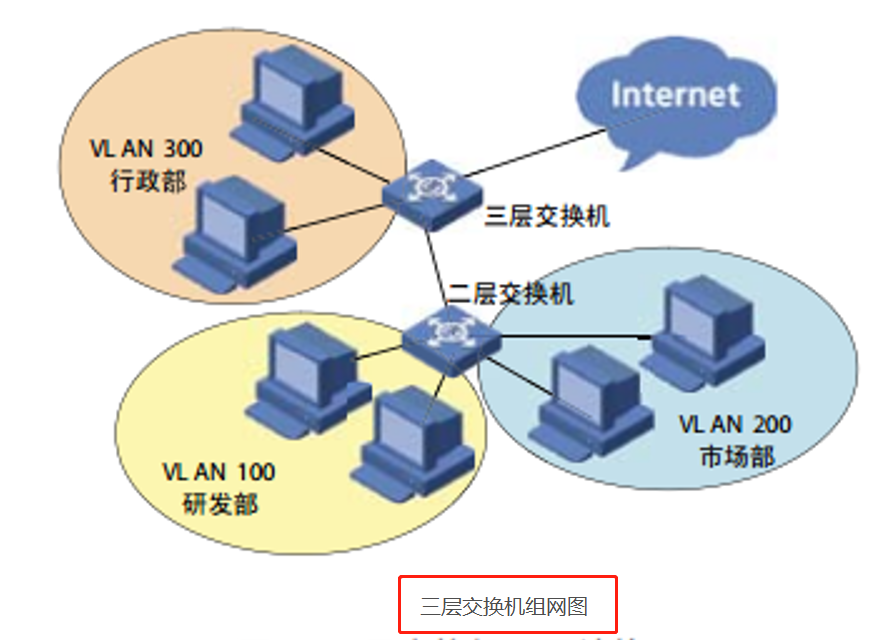
第三节 实例

### 第一节 简介

早期的网络一般采用二层交换机搭建局域网，在不同的局域网中间使用路由器来通信，那时局域网间的通信访问量比较小，使用路由器已经足够应付。但是随着通信范围的扩大，通信量的增多，而路由器由于自身的成本高、转发性能低、端口数据量少等特点无法很好的满足网络发展的需求。

三层交换机是一种实现高速三层转发的设备。

大多数三层交换机采用ASIC硬件芯片来完成转发， ASIC芯片内部集成了IP三层转发的功能，包括检查IP报文头、修改存活时间（ TTL）参数、重新计算IP头校验和、 IP包的数据链路封装等等。



路由器的三层转发主要依靠CPU进行，而三层交换机的三层转发依靠ASIC芯片完成，这就决定了两者在转发性能上的巨大差别。当然，三层交换机并不能完全替代路由器，路由器所具备的丰富的接口类型、良好的流量服务等级控制、强大的路由能力等仍然是三层交换机的薄弱环节

### 第二节 交换原理

目前的三层交换机一般是通过VLAN来划分二层网络并实现二层交换，同时能够实现不同VLAN间的三层IP互访。

1、位于同一网段（二层交换机）

源主机直接向目的主机发送ARP请求，在收到目的主机的ARP应答以后获取目的MAC，然后以目的主机的MAC作为目的MAC发送报文

2、位于不同网段（三层交换机）

通过网关来递交报文，即发送ARP获取网关IP对应的MAC，以网关MAC作为目的MAC进行报文发送。注意：源IP是源主机IP，目的IP仍然是目的主机IP

三层交换机转发有如下特点：

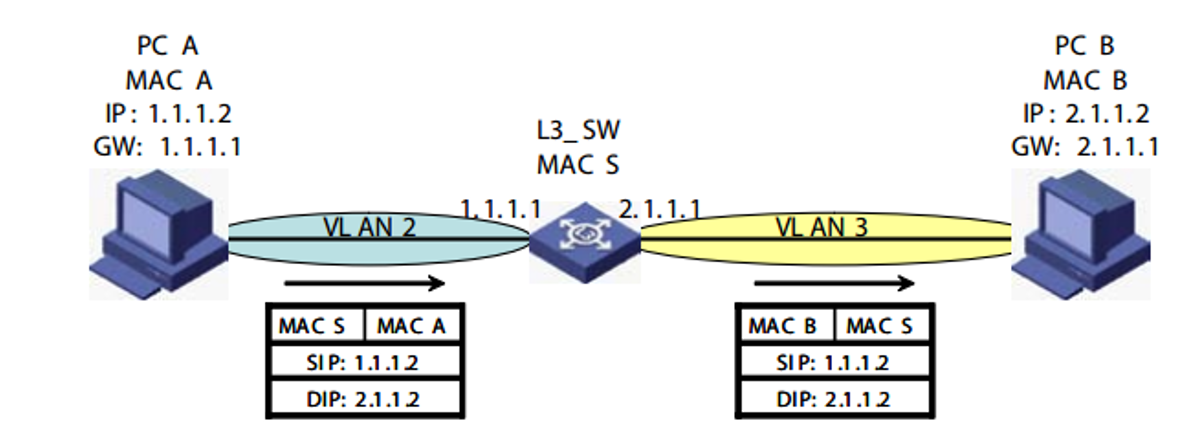
首包通过CPU转发，同时建立交换芯片硬件表项；后续包由交换芯片直接硬件转发，即常说的“一次路由、多次交换”；

交换芯片的硬件转发并不关心路由的具体下一跳IP地址是多少，硬件三层表项中只包含了目的地址（或网段）、目的IP（或下一跳IP）对应的MAC、出口VLAN、出端口；（这里说明一下，并不是所有的三层交换机的硬件三层表项都会包含“出端口”的）

IP报文每经过一次三层转发，它的源、目的MAC都会变化，但是源、目的IP是始终不变的。

### 第三节 实例

如图所示，通信的源、目的主机连接在同一台三层交换机上，但它们位于不同VLAN（网段）。对于三层交换机来说，这两台主机都位于它的直连网段内，它们的IP对应的路由都是直连路由。



1、 PC A首先检查出目的IP地址2.1.1.2（ PC B）与自己不在同一网段，于是把要发送的数据包先缓存起来，然后查找自己的ARP表，如果在自己的ARP表中没有找到网关MAC地址，就先发出请求网关地址1.1.1.1对应MAC的ARP请求，以获取网关IP地址1.1.1.1所对应的MAC地址。此ARP请求报文的“源MAC地址”字段是PC A的MAC地址“MAC A”，“目的MAC地址”因为未知，以全0格式填充；“源IP地址”和“目的IP地址”字段分别填上PC A的IP地址（1.1.1.2）和网关的IP地址（1.1.1.1）。

2、 L3\_SW收到PC A的ARP请求后，检查请求报文发现被请求IP是自己的三层接口IP，因此发送ARP应答并将自己的三层接口MAC（ MAC S）包含在其中。同时它还会把PC A的IP地址与MAC地址对应（ 1.1.1.2<==>MAC A）关系记录到自己的ARP表项中去（因为ARP请求报文中包含了发送者的IP和MAC），然后把PC A的IP地址（作为“目的IP地址”）、MAC地址（作为“下一跳MAC地址”），以及与交换机直接相连的端口号等信息下发到三层交换机ASIC芯片中的三层硬件转发表。此时在三层硬件转发表中就有了第一个转发表项，即PC A的转发表项。

3、 PC A在收到网关的ARP应答报文后，把要发送给PC B的数据包经过帧封装后的 “目的MAC地址”修改为网关MAC地址（MAC S1），其它不变，先把数据包发给网关（三层交换机）。

4、 L3\_SW收到报文后，首先根据报文的源MAC+VID（即VLAN ID）更新MAC地址表。然后，根据报文的目的MAC＋VID查找MAC地址表，发现匹配了自己三层接口MAC的表  
项。这里说明一下，三层交换机为VLAN配置三层接口IP后，会在交换芯片的MAC地址表  
中添加三层接口MAC＋VID的表项，并且为表项的三层转发标志置位。当报文的目的MAC  
匹配这样的表项以后，说明需要作三层转发，于是继续查找交换芯片的三层表项；  
5、交换芯片根据报文的目的IP去查找其三层表项，由于之前未建立任何表项，因此查找失  
败，于是将报文送到CPU去进行软件处理；

5、CPU同样会根据包中的“目的IP地址”去查找其软件路由表，发现匹配了一个直连网段（PC B对应的网段），于是继续查在ARP表中查找对应的MAC地址项。同样是由于是第一次查找，所以仍然查找失败。如果在ARP表中找到了对应的MAC地址，则数据可以直接由软件路由表转发了。  
 6、 PC B收到L3\_SW发送的ARP请求后，检查发现被请求IP是自己的IP，因此发送ARP应  
答并将自己的MAC（ MAC B）包含在其中。同时，将L3\_SW的IP与MAC的对应关系  
（ 2.1.1.1<==>MAC S）记录到自己的ARP表中去；

7、三层交换机CPU在收到PC B的ARP应答报文后，将其IP地址和MAC地址对应关系（2.1.1.2<==>MAC B）记录到自己的ARP表中去，把PC B的IP地址、MAC地址、进入交换机的端口号等信息下发到三层交换机的三层转发中。此时转发表中就有到达PC A和PC B这两条对应的表项了。

8、三层交换机CPU根据获得的PC B的MAC地址和端口信息，以及软件路由表信息，把由PC A发来的IP数据包转发给PC B，这样就完成了PC A到PC B的第一次单向通信。

由于芯片内部的三层引擎中已经保存了从PC A到达PC B的完整转发路径信息，所以以后PC A与PC B之间进行通信，或其它网段的站点想要与PC A或PC B进行通信时，三层交换机的ASIC芯片就会直接把包从对应的三层硬件转发表项中指定的端口转发出去，而不必再把包交给CPU进行路由处理。这就是所谓的“一次路由（指通过CPU路由表查到了对应的直连网段），多次交换”的原理，大大提高了转发速度。