二层网络基础知识

周威光整理[*]

2017-06-24

[*]简介：恒天云 FTE

目 录

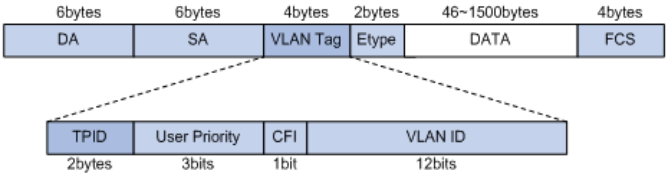
1 vlan 基础知识

1.1 vlan 的含义

LAN 表示 Local Area Network，本地局域网，通常使用 Hub 和 Switch 来连接 LAN 中的计算机。一般来说，当你将两台计算机连入同一个 Hub 或者 Switch 时，它们就在同一个 LAN 中。同样地，你连接两个 Switch 的话，它们也在一个 LAN 中。一个 LAN 表示一个广播域，它的意思是，LAN 中的所有成员都会收到 LAN 中一个成员发出的广播包。可见，LAN 的边界在路由器或者类似的 3 层设备。

VLAN 表示 Virutal LAN。一个带有 VLAN 功能的 switch 能够同时处于多个 LAN 中。最简单地说，VLAN 是一种将一个交换机分成多个交换机的一种方法。

IEEE 802.1Q 标准定义了 VLAN Header 的格式。它在普通以太网帧结构的 SA（src addr）之后加入了 4bytes 的 VLAN Tag/Header 数据，其中包括 12-bits 的 VLAN ID。VLAN ID 最大值为 4096，但是有效值范围是 1 - 4094。

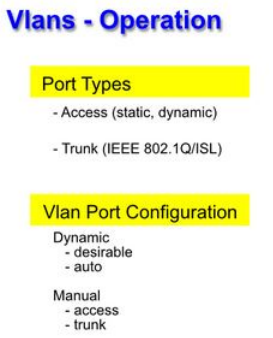


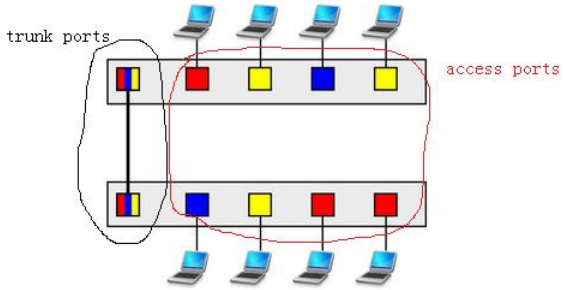
1.2 vlan 交换机端口类型

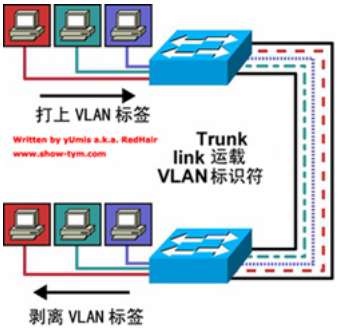
带 VLAN 的交换机的端口分为两类：

(1). Access port：这些端口被打上了 VLAN Tag。离开交换机的 Access port 进入计算机的以太帧中没有 VLAN Tag，这意味着连接到 access ports 的机器不会觉察到 VLAN 的存在。离开计算机进入这些端口的数据帧被打上了 VLAN Tag。

(2). Trunk port：有多个交换机时，组 A 中的部分机器连接到 switch 1，另一部分机器连接到 switch 2。要使得这些机器能够相互访问，你需要连接两台交换机。要避免使用一根电缆连接每个 VLAN 的两个端口，我们可以在每个交换机上配置一个 VLAN trunk port。Trunk port 发出和收到的数据包都带有 VLAN header，该 header 表明了该数据包属于那个 VLAN。因此，只需要分别连接两个交换机的一个 trunk port 就可以转发所有的数据包了。通常来讲，只使用 trunk port 连接两个交换机，而不是用来连接机器和交换机，因为机器不想看到它们收到的数据包带有 VLAN Header。







1.3 数据包进出交换机不同类型端口表现

交换机端口类型	tagged（进）	untagged（进）	出（把帧发给包含其 VID 的端口）
Access 端口	丢弃	打上 PVID	剥离 VID
Trunk 端口	若允许，则不变；否则丢弃	打上 PVID	若 VID 与 PVID 不同，则剥离 VID

1.4 vlan 的不足

(1). VLAN 使用 12-bit 的 VLAN ID，所以 VLAN 的第一个不足之处就是它最多只支持 4096 个 VLAN 网络（当然这还要除去几个预留的），对于大型数据中心的来说，这个数量是远远不够的。

(2). VLAN 是基于 L2 的，所以很难跨越 L2 的边界，在很大程度上限制了网络的灵活性。

(3). VLAN 操作需手工介入较多，这对于管理成千上万台机器的管理员来说是难以接受的。

2 二层交换的基础知识

2.1 二层交换机最基本的功能

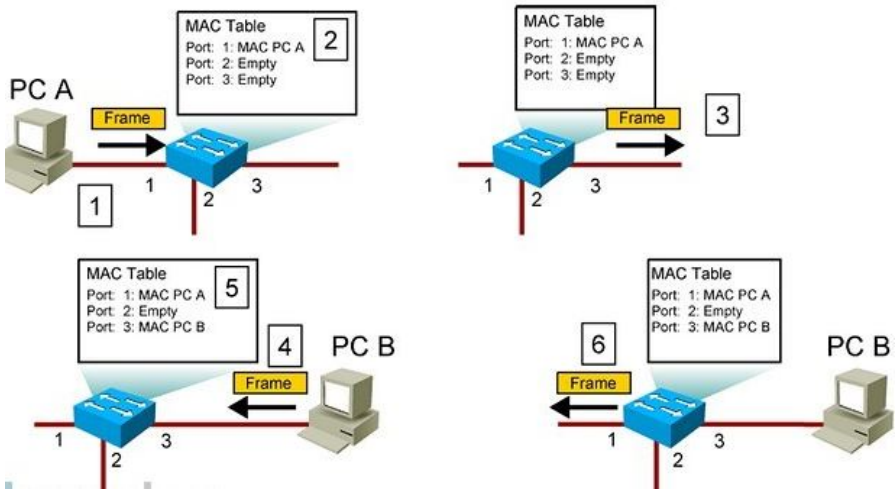
(1). mac 地址学习

(2). 数据帧的转发

(3). 添加 vlan 标签和剥离 vlan 标签

2.2 数据帧转发和 mac 学习的过程

配置普通的交换机:



(1). PC A 发一个帧到交换机的 1 端口，其目的 MAC 地址为 PC B 的 MAC。

(2). 交换机比较其目的 MAC 地址和它的内部 MAC Table，发现它不存在（此时表为空）。在决定泛洪之前，它把端口 1 和 PC A 的 MAC 地址存进它的 MAC Table。

(3). 交换机将帧拷贝多份，分别从 2 和 3 端口发出。

(4). PC B 收到该帧以后，发现其目的 MAC 地址和他自己的 MAC 地址相同。它发出一个回复帧进入端口 3。

(5). 交换机将 PC B 的 MAC 地址和端口 3 存在它的 MAC 表中。

(6). 因为该帧的目的地址为 PC A 的 MAC 地址它已经在 MAC 表中，交换机直接将它转发到端口 1，达到 PC A。

配置了 vlan 的交换机的该机制类似:

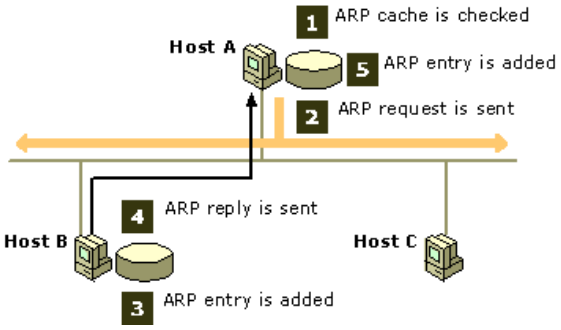
(1). MAC 表格中每一行有不同的 VLAN ID。做比较的时候，拿传入帧的目的 MAC 地址和 VLAN ID 和此表中的行数据相比较。如果都相同，则选择其 Ports 作为转发出口端口。

(2). 如果没有吻合的表项，则将此帧从所有有同样 VLAN ID 的 Access ports 和 Trunk ports 转发出去。

2.3 Address Resolution Protocol(ARP) 原理

功能：ARP 通过 IP 地址获取到 MAC 地址

情况一：目的 IP 地址在同一网段的话



该示例中，Host A 和 B 在同一个网段中。A 的 IP 地址是 10.0.0.99，B 的 IP 地址是 10.0.0.100。当 A 要和 B 通信时，A 需要知道 B 的 MAC 地址。该过程经过以下步骤:

(1). A 上的 IP 协议栈知道通过 B 的 IP 地址可以直接到达 B。A 检查它的本地 ARP 缓存来看 B 的 MAC 地址是否已经存在。

(2). 如果 A 没有发现 B 的 MAC 地址，它发出一个 ARP 广播请求，来询问 10.0.0.100 的 MAC 地址是什么？该数据包:

```
1    SRC MAC:   A 的 MAC
2    DST MAC:   FF:FF:FF:FF:FF:FF
3    SRC IP:   A 的 IP
4    DST IP:   B 的 IP
```

(3). 该网段中所有的电脑都将收到该包，并且会检查 DST IP 和自己的 IP 是否相同。如果不同，则丢弃该包。Host B 发现其 IP 地址和 DST IP 相同，它将 A 的 IP/MAP 地址加入到自己的 ARP 缓存中。

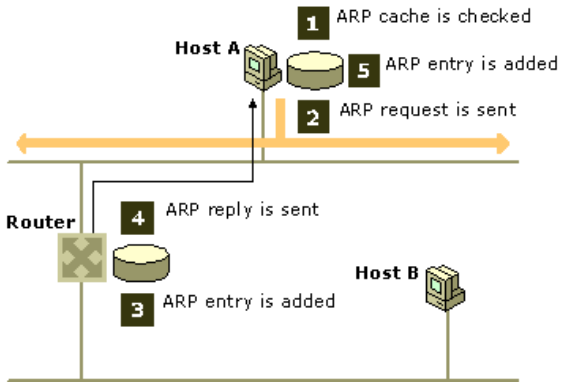
(4). B 发出一个 ARP 回复消息

```
SRC MAC:  B 的 MAC
DST MAC: A 的 MAC
SRC IP:  B 的 IP
DST IP: A 的 IP
```

(5). 交换机直接将该包交给 host A。A 收到后，将 B 的 MAC/IP 地址缓存到 ARP 缓存中。

(6). A 使用 B 的 MAC 作为目的 MAC 地址发出 IP 包。

情况二：目的 IP 地址不在同一个网段的话



本例子中，A 的地址是 10.0.0.99，B 的地址是 192.168.0.99。Router 的 interface 1 和 A 在同一个网段，其 IP 地址为 10.0.0.1；interface 2 和 B 在同一个网段，其 IP 地址为 192.168.0.1。A 使用下面的步骤来获取 Router 的 interface 1 的 MAC 地址。

(1). 根据其路由表，A 上的 IP 协议知道需要通过它上面配置的 gateway 10.0.0.1 才能到达到 B。经过上面例子中的步骤，A 会得到 10.0.0.1 的 MAC 地址。

(2). 当 A 收到 Router interface 1 的 MAC 地址后，A 发出了给 B 的数据包：

```
SRC MAC:  A 的 MAC
DST MAC: Router 的 interface 1 的 MAC 地址
SRC IP:  A 的 IP
DST IP: B 的 IP
```

(3). 路由器的 interface1 收到该数据包后，根据其路由表，首先经过同样的 ARP 过程，路由器根据 B 的 IP 地址通过 ARP 获得其 MAC 地址，然后将包发给它。

```
SRC MAC:  Router interface 2 的 MAC
DST MAC: B 的 MAC
SRC IP:  A 的 IP
DST IP: B 的 IP
```

2.4 总结：两个主机通信的大致过程

(1). 主机 A 通过主机 B 的 ip 访问 B

(2). 主机 A 首先查看自己的路由表，目的 ip 是在本网段，还是在不同网段

(3). 如果目的 ip 是在本网段，查看本机 arp 缓存是否含有目的 ip 对应的 mac 地址。

(4). 如果有, 则直接发送到 B 的包。如果没有，则需要通过 arp 广播获取 B 的 mac 包之后，再进行发送。

arp 广播包头如下：

```
SRC MAC:  A 的 MAC
DST MAC: FF:FF:FF:FF:FF:FF
SRC IP:  A 的 IP
DST IP: B 的 IP
```

A 发送到 B 的包头如下：

```
SRC MAC:  A 的 MAC
DST MAC: B 的 MAC
SRC IP:  A 的 IP
DST IP: B 的 IP
```

(5). 如果目的 ip 不是在同网段，则通过路由表获取下一跳网关 ip。并查看本机 arp 缓存是否含有该 ip 对应的 mac 地址。

(6). 如果有，则开始向网关发送到 B 的包。如果没有，则需要通过 arp 广播获取网关 ip 对应的 mac 地址 arp 广播包头如下：

```
SRC MAC:  A 的 MAC
DST MAC: FF:FF:FF:FF:FF:FF
SRC IP:  A 的 IP
DST IP:  下一跳 的 IP
```

A 向网关发送到 B 的包头如下：

```
SRC MAC:  A 的 MAC
DST MAC:  下一跳 的 MAC
SRC IP:  A 的 IP
DST IP: B 的 IP
```

(7). 路由器收到 A 发送到 B 的包，查看路由表，B 所处网段是否与其某个接口直连，还是需要转到另一个路由器进行转发

(8). 若需要通过另一个路由器的转发，则需要查看本地 arp 缓存或 arp 广播获取下一跳路由的 mac，然后再发送包 arp 广播包头如下：

```
1    SRC MAC: 路由器1 发送端口的 MAC
2    DST MAC: FF:FF:FF:FF:FF:FF
3    SRC IP:  路由器1 发送端口的 IP
4    DST IP:  路由器2 接收端口 的 IP
```

路由器 1 转发到路由器 2 的包头如下:

```
1    SRC MAC:  路由器1 发送端口的 MAC
2    DST MAC: 路由器2 接收端口 的 MAC
3    SRC IP:  A 的 IP
4    DST IP: B 的 IP
```

(9). 若不需要另一个路由器的转发直接到 B，则需要查看本地 arp 缓存或 arp 广播获取 B 的 mac，然后再发送包 arp 广播包头如下:

```
1    SRC MAC: 路由器1 发送端口的 MAC
2    DST MAC: FF:FF:FF:FF:FF:FF
3    SRC IP:  路由器1 发送端口的 IP
4    DST IP:  B 的 IP
```

路由器转发到 B 的包头如下:

```
1    SRC MAC:  路由器1 发送端口的 MAC
2    DST MAC:  B的 MAC
3    SRC IP:  A 的 IP
4    DST IP: B 的 IP
```

(10). 在包的发送过程中，伴随着主机和路由器添加 ip 和 mac 映射关系的 arp 缓存条目，以及交换机添加 mac 和端口映射的 mac 表条目。这里不一一赘述。