网络知识

# MAC地址

1. MAC（Media Access Control，介质访问控制）地址，或称为物理地址，也叫硬件地址，用来定义网络设备的位置，MAC地址是网卡出厂时设定的，是固定的（但可以通过在设备管理器中或注册表等方式修改，同一网段内的MAC地址必须唯一）。MAC地址采用十六进制数表示，长度是6个字节（48位），分为前24位和后24位。
2. 前24位叫做组织唯一标志（OUI），是由IEEE的注册管理机构分配给不同厂家的代码，用于区分不同厂家。
3. 后三字节是厂家自己分配的，也叫拓展标识符，同一厂家的后24位Mac地址是不同的。
4. MAC地址作用于OSI参考模型的第二层（数据链路层），通常用于交换机（工作于数据链路层）数据的转发依据，即MAC地址和端口对应（决定该MAC地址的数据从哪个端口发送），交换机必须维护这个对应关系（MAC地址表），解析端口进入的数据中的目的MAC，然后查找维护的MAC地址表决定该数据的出端口（关于交换机数据转发将在交换机部分详解）。
5. 广播地址
6. 组播地址

# IP地址

* 1. IP（internet protocol，网际协议）地址作用于OSI参考模型的第三层（网络层），用于网络层设备识别通信对端的信息地址，路由器则根据其数据包的目的IP判断是否和源IP在同一网段，不在同一网段则转发该数据包（关于路由器的数据转发在路由部分详解）。
  2. IP地址格式：
     1. IP地址格式（ipv4）由32 bit二进制数组成，共4字节以‘.’分割，该4字节的地址又被分为网络部分和主机部分（D、E类地址则不区分），网络部分相同的字节称为同一“网段”（C类地址192.168.1网段，192.168.2网段则是不同网段）。
  3. IP地址分类：

IP地址分为A、B、C、D、E五类地址，不同类型地址的网络部分和主机部分的长度（所占字节）不同。

A类地址：前一字节为主机地址，并且第一位固定为0，故A类地址的网络号（也可以说是网段数量）范围是0~127（0 ~0111 1111），所以在该类地址同一网段上能连接的计算机数量是2的24次方（后三字节为主机地址），默认子网掩码为255.0.0.0也可写作8。

B类地址：前2字节为网络地址，并且前2bit固定为10，后2字节为主机地址，故网段号为(128.0~191.255)，默认掩码为255.255.0.0/16，同一网段的主机数量为2的16次方。

C类地址：前3字节为网络地址，并且前3bit固定为110，后一字节为主机地址，所以同一网段的主机数量为256个，可用主机地址为254（见注意），掩码为255.255.255.0/24。

D类地址：也叫组播地址，或称多播地址，没有主机地址，并且前4bit固定为1110（224.0.0.0~239.255.255.255），无子网掩码。

E类地址：前5bit固定为11110（140.0.0.0-255.255.255.255），无子网掩码，其中255.255.255.255为广播地址，其余的为保留地址。

注意：在主机号中，为全0的主机号用于IP地址不能使用的情况下才使用（称网段地址），主机号为1的通常作为默认的网关（通常为交换机端口的主机号），为全1的主机号则是广播地址，此外127.0.0.1是本地环回地址，通常测试使用，0.0.0.0若IP地址和其他的地址冲突，使用ipconfig命令后则会显示为该地址掩码为0.0.0.0，169.254.0.0-169.254.255.255为自动私有IP地址。

* 1. 子网掩码：
     1. 子网掩码也是由长度为32bit的二进制数组成的一个地址，与IP地址相对应，通常与IP地址做与运算，左边连续的全1字段表示IP地址的网络号字段，所以子网掩码左边连续的1的长度表示IP地址网络号的长度，右边连续的0的长度则表示主机号的长度。
     2. 子网掩码用于网络号和主机号的识别、用于子网超网的划分，其本质就是将左边连续的1的个数确定ip的网络号长度，其余的则为主机号长度，以此可以划分不同大小的子网，实现网络的有效利用（减少路由通告或减少ip地址浪费），合并不同的子网。
     3. 有类和无类网络、子网和超网：

有类网络：也称标准网络，即A-C类网络使用的是默认的子网掩码。

无类网络：即某一类网络中子网掩码多于或少于标准的掩码。

超网：子网掩码长度小于该类的子网掩码，即主机号长度增多。

子网：子网掩码长度大于该类子网掩码，即主机号长度减少。

* + 1. CIDR无类别域间路由：

CIDR使用8-30的网络号（也可以说是子网掩码1的个数）划分子网，其表示法为，IP地址/n（斜线表示法192.168.1.0/24,24表示子网掩码连续24个1即255.255.255.0），CIDR可将多个子网划分为较大的子网实现网络的聚合。

* + 1. VLSM可变长子网掩码：

规定一个有类（A-C）网络类包含多个子网掩码的能力，以及对子网再划分的能力，VLSM就是将一个子网划分为多个子网。

# 以太网

1. 简介：以太网（Ethernet）开创于1972年，是一种共享介质的局域网技术，多个站点连接到同一个共享介质上，因此同一时间只能有一个站点发送数据，以太网采用随机访问控制协议中的CSMA/CD（带有冲突检测的载波侦听多录访问）作为多路访问控制协议；

载波侦听：当一个设备需要发送数据时，会首先检测物理介质是否空闲，如果介质忙则会等待，推迟数据的发送，当发现介质空闲后也必须等待一个IFG（帧间隔）时间让物理信道平稳，之后若介质无冲突才将数据发送出去。

回退：若同一时间有多个设备要发送数据，那么就会产生冲突，此时这些设备要发送一段时间的干扰信号，确保在介质上的所有设备都能检测到以太网上已产生了冲突，正在发送数据的设备必须终止，等待一个随机时间，这个随机时间称之为回退时间，若设备连续16次发送数据都检测到冲突，则数据发送失败。

1. 标准以太网帧Ethernet II：

EthernetII协议是局域网中标准的通信协议，规定链路上的数据传送以以太网帧的标准格式进行传输，标准以太网帧结构包含如下几个部分：



Preamble：7字节的前导码，以10交替表示一个以太帧的开始；

SFD：1字节特定10101011，11告知下一字段为目的主机地址；

DA：6字节目的ip；

SA：6字节源ip；

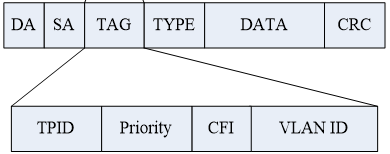
Type/Length：2字节，如果字段值在0x0000－0x05DC范围内，则表示该字段为Length，该帧为802.3 raw封装。如果字段值在0x0600－0xFFFF范围内，则表示该字段为Type字段，该帧为Ethernet II封装，0x05DD－0x05FF保留没有使用；

Payload：46~150字节可变，上层协议数据；

RFC：4字节校验和。

3、802.1Q以太网帧：

IEE802.1Q是用于规定vlan虚拟局域网实现方式的一种协议，其帧格式在标准以太网帧TYPE字段前面添加了一个TAG字段；



TPID：2字节固定为0x8100,，表示802.1Q封装；

Priority：3BIT，值表示帧的优先级；

CFI：1BIT（Canonical Format Indicator），0表示规范格式，1表示FDDI帧和令牌环网帧；

VlanID：12BIT，vlan ID。

# 二层交换机工作原理

1. 简介：二层交换机工作在OSI模型的第二层（数据链路层），对数据包的转发是建立在链路层信息MAC地址基础之上的（mac：媒体访问控制地址，也叫局域网地址（LAN地址）、以太网地址或物理地址每个计算机都有唯一的一个mac地址，IP地址则是第三层网络层负责），内部使用ASIC硬件芯片实现转发，性能很高。
2. 冲突域和广播域：

由于多个网络设备连接在同一个介质上，共享该介质的带宽（也就是多个设备连接到一条线上），多个设备同时发送数据时便会产生冲突，同一介质上的所有网络设备就处于一个冲突域；二层交换机的各个端口的发送和接收数据独立，并且各个端口可以同时收发数据，互不影响，各端口属于不同的冲突域，可有效的隔离网络中物理层的冲突域。

当在一个互联的网络中，某个设备发送一个广播报文，其余的设备都会接收到这个报文，这个范围就统称为一个广播域，当这个互联的网络设备比较多时，广播域就比较广会占用较多的带宽，普通二层交换机的所有端口都处于一个广播域（当某个端口收到广播时，其他端口都会发出该广播报文），为解决不必要的广播带宽占用，使用VLAN技术有效的隔离广播域，每一个VLAN属于一个广播域（关于VLAN将在后面介绍）。

1. MAC地址表：当二层交换机的某个端口进入一条报文时，交换机便会解析该报文的源MAC和目的MAC，然后将源MAC和端口绑定记录到硬件Mac转发表中，若该表向存在则更新mac老化时间（若在一定时间内未更新该mac则会从mac表中删除），交换机以此方式学习以太网帧的源MAC来维护MAC地址和端口的对应关系。
2. 二层交换机数据转发：交换机数据转发根据目的mac在mac表中查找与其对应的端口，若未查找到则除源数据的端口外向所有的端口转发；若目的mac是广播或组播，也除源端口外向所有的端口转发；若查询到的端口为源端口，则丢弃该帧。

# VLAN

1. VLAN简介：

VLAN(Virtual Bridged Local Area Network)虚拟桥接局域网，在交换机中通常将某些端口划分到某个或多个VLAN中(简单的说就是将交换机的多个端口划分在一个编号中)，使得不同VLAN中的端口之间广播报文无法通过，有效的分割广播域，一个VLAN等价于一个广播域，VLAN主要用于缩小广播域，抑制广播风暴（广播：当未找到mac对应的端口或mac地址为广播地址时，则会向除源端口外的所有端口转发报文），广播报文不能跨越该VLAN传送。

1. VLAN实现方式：
2. 基于端口（最常见）：根据以太网交换机的端口来划分，指明那个端口属于那个VLAN。
3. 基于mac地址：指明那个mac对应的端口属于那个VLAN。
4. 基于网络层协议：将物理网络划分成基于协议的逻辑VLAN，接收到帧时，该帧的VLAN由该帧的信息包中的协议类型决定。
5. 基于IP子网：交换机通过目的IP地址指定其端口所属的VLAN。
6. VLAN成员的三种连接方式;

Access：设备间相连，报文不带tag标签（见以太网章节802.1Q），不需要区分VLAN；

Trunk：设备相连，一般指交换机间相连，在同一VLAN间数据不带tag标签，夸VLAN的报文都必须带tag标签；

Hybrid：根据需要设置哪些VLAN报文带tag，可以设置多个VLAN不用带tag。

1. 报文tag处理：

TPID：协议标签帧标识，（820.1Q固定为0x8100）；

PRI：优先级；

CFI：0规范格式，1为非规范格式；

VID：VLAN ID（封装PVID-端口VLAN标识，每个端口可以加入多个VLAN，但只有一个PVID同端口的VID关联（0-4094））。

1. 当端口收到无TAG的帧时，便添加tag并封装PVID，对已含tag的帧则没有影响，tag模式用于链路中存在多个VLAN，当在无tag模式时，数据转发出去会去掉tag。
2. 报文接收：

无tag：允许报文进入该端口，并打上PVID的VLAN tag，与端口属性无关；

有tag：需要判断是否允许该报文进入端口；

Access端口：PVID和tag中封装的VLAN一致（无tag的须打上对应的PVID标签），否则丢弃；

Trunk/Hybrid端口：如果端口设置为仅允许tag中标明的VLAN通过，则只接受含VLAN tag的报文，否则丢弃。

1. 报文输出：

Access端口：去掉报文tag转发；

Trunk端口：报文所在的VLAN和PVID一致，则带tag转发，否则去掉tag转发；

Hybrid端口：报文所在VLAN配置为带tag，则带tag转发，否则去掉tag转发。

# 路由与三层转发

## 路由：

当设备（三层交换机或路由器）的接口接收到一个数据包时，根据数据包的目的地址查找并计算出该报文从哪个接口发送出去的一个过程（即为该目的ip的数据包查找一条通路的过程）。在OSI参考模型中，路由工作在第三层网络层，通过路由转发可以实现网络间的相互连接（也就是相同的协议间的通信）。

## 常见网络协议：

此处仅罗列出常用的几种协议（不含物理层），详细请参考<TcpIp学习笔记>(注：该笔记尚未出版)

### 数据链路层：

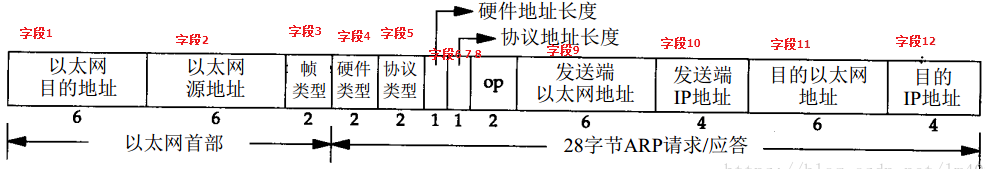
#### 令牌环：

#### 帧中继：

### 网络层：

#### ARP：

地址解析协议(address resolution protocol)，通过arp请求和应答获取目的ip所对应的mac地址，将目的ip和mac存储于arp表中，以此作为设备间通信的基本信息，arp报文格式如下：



字段1：6字节以太网目的地址，通常为0xFFFFFF广播mac地址；

字段2：6字节以太网源地址，通常为本机mac地址;

字段3：2字节帧类型为0x0806表示arp类型；

字段4：2字节硬件类型，0x0001表示以太网硬件类型；

字段5：2字节协议类型，0x0800表示是对ipv4类型进行地址映射；

字段6、7：分别占1字节，分别表示硬件地址的长度0x6字节和IP地址长度为0x4字节；

字段8：2字节分，0x0001 arp请求，0x0002 arp应答；

字段9、10：6字节和4字节分别表示，发送端的请求或应答mac，送端的请求或应答ip；

字段11、12： 6字节和4字节，表示接收端mac地址和ip地址，arp请求报文通常为全0。

#### ICMP：

ICMP（Internet Control Message Protocol）Internet控制报文协议，是TCP/IP的一个子协议，用于在IP主机、路由器之间传递控制消息（指网络通不通、主机是否可达、路由是否可用等网络本身的消息），该协议是一个面向无连接的协议，ping命令就是用的ICMP协议。

#### MPLS：

#### IP：

网际互连协议，规定数据报的传输格式，规定了数据报的传输方法和路由选择，设备间的数据传输必须要使用IP地址通信。

IPV4和IPV6：

### 传输层：

#### TCP：

#### UDP：

#### OSPF：

Open Shortest Path First(开放最短路径优先)，用于在单一自治系统AS（autonomous system）内决策路由，是一种基于SPF算法的链路状态路由的IGP内部网关（Interior getaway protocol）协议（这里的内部指的是一个AS系统），反应路由接口或链路参数（接口状态up/down、接口ip地址、接口的子网掩码、接口所连接的网路以及使用路由器连接的相关费用），ospf协议就是传输这些链路状态参数（所以ospf也叫接口链路状态协议）。每个路由器都有一个本地链路状态（也就是自己的这些链路状态信息），所有路由器的链路状态都会在整个OSPF域内传播，直到所有的路由器都有完整并相同的链路状态数据库为止，每个路由器都以自己的链路状态信息为根节点，其它分支信息表示可达最短的路径。

路由器ID：路由器IP地址；

Neighbor Router：邻居路由，连接到统一公共接口的路由器（也就是直接相互连接的路由ip）；

DR：Designated Router，DR在广播和NBMA网路中作为其他路由的邻居，路由器（DR）被指定的向公共网络中其他路由器传播链路状态信息；

BDR：BackupDR后备DR，DR路由器故障备用；

ABR：Area Border Router，连接了多个ospf自治系统的路由器，也叫自治系统边界路由器，或在一个AS中不允许IGP的路由器；

Adjacency：紧邻，两个点对点连接的路由形成紧邻，也可以在广播或NBMA网络的DR和非指定路由器之间形成，还可以在BDR和非指定路由器之间形成，ospf路由状态信息只能通过紧邻被传输和接收；

LSA：link State Advertisement链路状态通告，该通告向整个ospf区域扩散；

LSDB：链路状态数据库（LSA的集合），每个路由器以本地链路状态为根节点，以此计算出最佳的路由；

Flooding：在ospf区域内洪泛路由器之间的链路状态数据库；

External Routing：外部路由，从外部AS或另一个路由协议获取的路由可以作为外部路由放在ospf中；

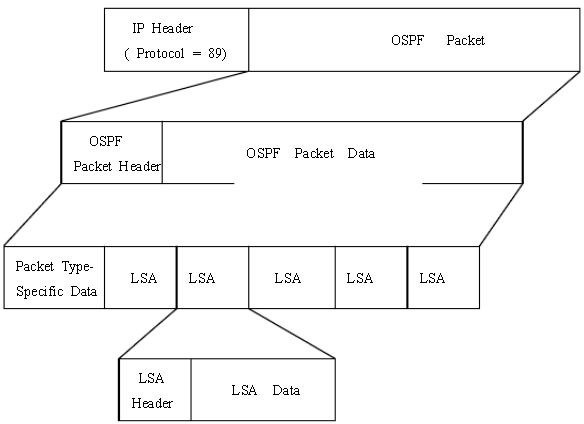
Intraarea：区域内路由，同一区域内路由；

Interarea：区域间路由，同一AS不同区域的路由；

StubArea：只有一个出口路径的区域。

Area：一个AS可划分为多个区域， area为0的为骨干区域，负责分发非骨干区域间的链路信息；

Ospf报文格式：ospf使用ip协议进行传输，协议号为0x89；



协议报文包含五种协议报文：

|  |  |  |  |
| --- | --- | --- | --- |
| 报文类型 | | 作用 | |
| Hell（问候报文） | | 发现和维护令居，10S发送一次，超时为其4倍时间，可修改。 | |
| DD（data description） | | 紧邻同步链路数据库 | |
| LSR（link state request） | | 链路状态请求分组 | |
| LSU（link state update） | 链路状态更新分组 | | |
| LSAck（link state acknowledge） | | | 链路状态确认分组 |

OSPF运行状态：

Down状态：

Init State：发出了hell报文，但未收到回复；

Two-way state：双向会话建立，收到了一个hell报文，并且报文中的router-id是自己的id；

Exstart-state：信息交换初始状态，本地路由和令居建立Master/Slave关系，并确定DD sequence Number，路由器id大的为master；

Exchange-state：信息交换状态，本地路由器和邻居交换一个或多个DBD分组（也叫PPD），DBD包含LSDB中LSA中条目的摘要信息；

Loading：信息加载状态，收到DBD后，使用LSACK分组确认已收到DBD，收到的信息同LSDB中的信息比较，如果DBD中有更新的链路状态条目，则向对方发送一个LSR请求更新LSA。

Full：完全邻接状态，这种邻接出现在router LSA和network LSA中。

### 应用层：

#### TFTP：

#### FTP：

#### NSF：

#### DNS：

#### SSH：

#### SIP：

#### HTTP：

#### BGP：

Border Getaway Protocol(边界网关协议)，该协议运行于TCP上，用于在不同自治系统(AS)之间交换路由信息（指网络可达信息，包括列出AS的信息，这些信息体现AS的拓扑图并清除环路，同时在AS级别上可实施策略决策），两个AS中用来交换信息的两个路由器称为边界网关(border getaway)或边界路由器(border router)。

Speaker：发送BGP报文的设备称为BGP发言者，接收或产生报文信息，并发布(advertise)给其他bgp speaker。

Peer：对等体，相互交换报文的speaker之间称为对等体，也叫做邻居；

BGP路由号：(router id)用于对等体之间的会话id，每个BGP设备有唯一的ID，缺省为路由器id；

BGP报文：

Open报文：建立bgp对等体；

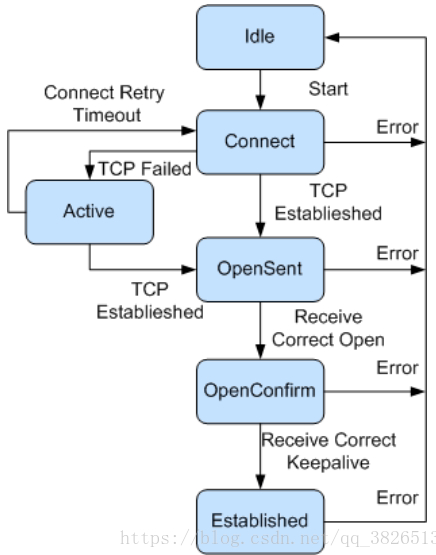
Update报文：在对等体之间交换路由信息；

Notification报文：终端BGP连接；

Keepalive报文：保持BGP的连接；

Route-refresh报文：用于改变路由策略后请求对等体重新发送路由信息。仅支持路由刷新(route-refresh)的设备才发送或响应此报文。

BGP状态机：



Idle：BGP初始状态，此时bgp不发送tcp请求连接，也不接收邻居的tcp请求，只有收到本设备start事件（设备配置启动bgp或重置bgp等启动操作引起）后才同其他BGP对等体建立tcp连接，并转至connect状态；

Connect：BGP启动重传定时器(connect retry，缺省为32s)，等待tcp的连接完成；

(1)Tcp连接成功，则发送open报文，转至OpenSent状态；

(2)失败转至active状态；(3)超时未收到bgp响应，则继续连接停留在connect状态；(4)其他事件出错转至idle状态；

Active：BGP总是试图建立tcp连接；(1)连接成功则发送open报文，关闭重连定时器，转至OpenSent状态；(2)连接失败停留在Active状态；(3)重传定时器超时，BGP没有收到对等体的响应，转至connect状态；

OpenSent：等待对等体的open报文，并对AS号、版本号、认证码等检查；(1)报文正确，发送Keepalive报文，转至Openconfirm状态；(2)报文错误发送Notification报文给对等体，转至idle状态；

OpenConfirm：等待Keepalive或Notification报文；(1)收到keepalive报文转至Established状态；(2)收到Notification报文则转至Idle状态；

Established：和对等体交换Update、Keepalive、Route-refresh和Notification报文；

1. 收到正确的update/keepalive报文，表示对端正常，保持bgp连接；
2. 收到错误的update/keepalive报文，则发送Notification报文给对端，转至Idle状态；
3. 收到Route-refresh报文，保持状态不变；
4. 收到Notification报文转至Idle状态；
5. 如果收到tcp拆链通知，则断开BGP连接转至Idle状态。