SCGY Network and PXE

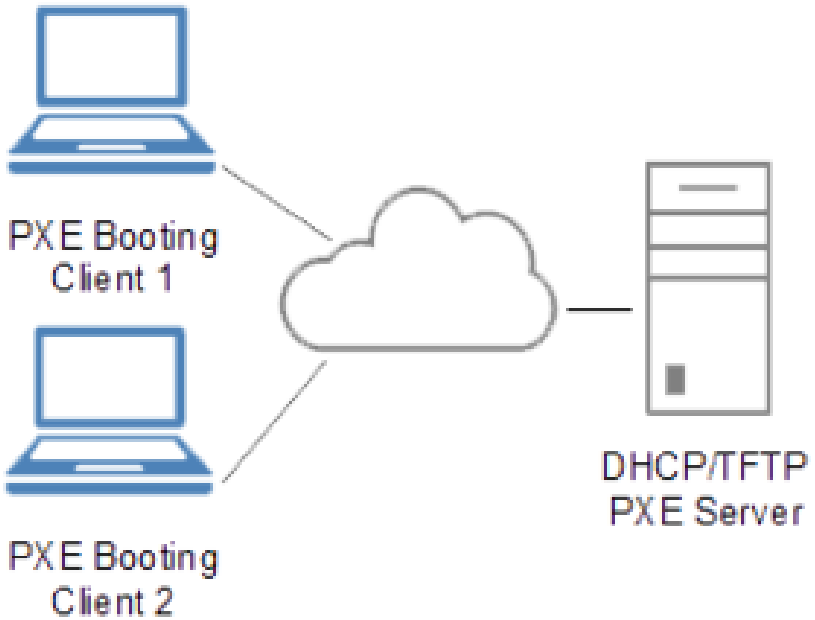
Jauntyliu @ SCGY-Tech

2019 年 3 月 9 日

What is PXE?

Preboot eXecution Environment:

- 赋予计算机通过网络启动合适计算机系统的能力
- 在客户端除了设置为 PXE 启动优先外，不需要任何额外设置
- 通过 DHCP、TFTP 和网卡中的 PXE ROM 实现



Network Prerequisites

先来学习一些网络基础知识。为了清晰，我们从网线中传输的 1010 开始，一步步构建到应用层。

What's inside the twisted wire cable?

双绞线中传输的 1 和 0：

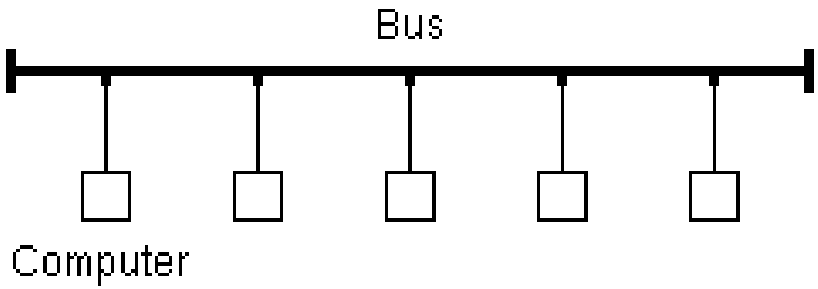
- 基本数据传输单位是**一个帧**，在以太网规范 IEEE 802.3 中有详细描述，包括帧的格式（Premable，MAC，Payload，Checksum）和编码（4b/6b 等）
- 帧有广播帧（目标 MAC 为 FF:FF:FF:FF:FF:FF）和普通帧（单播帧）的区别，请注意
- 规定最大帧长度 MTU，传统规范为 1500；也要注意 Jumbo Frame
- 通过 MAC 地址识别接收和发送方；MAC 地址理论上全球网卡唯一
- 采用**载波监听与碰撞检测（CSMA/CD）**技术，允许通过共享媒介通信

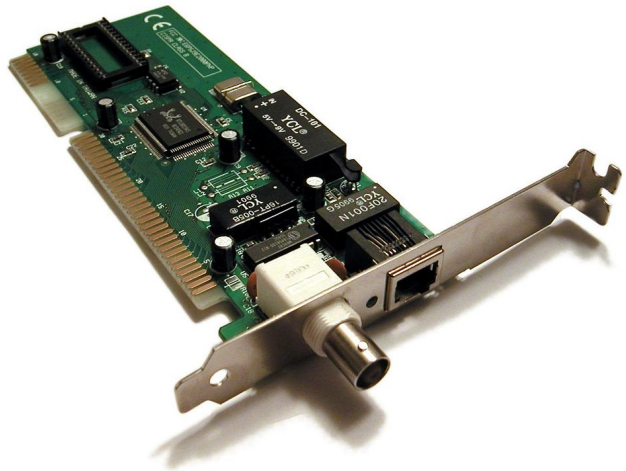
  问题：如何确定网络中计算机网卡的 MAC 地址？

Network Topology, Hubs and Switches

早期的互联网通过共享总线通信，所有计算机连接到同一根“网线”或同轴电缆上：

- 节约成本，不需要额外硬件
- 但是，如果网线中任意一点故障，将导致整个网络不可用





图：总线型网络 与 一块同时支持同轴电缆和双绞线的 ISA 网卡

Network Topology, Hubs and Switches

为了提高容错，大家想到把拓扑结构改成星型，提高网络可靠性。中间的节点叫做 **集线器（Hub）**：

- 一个 Hub 基本可以理解为把几个端口的双绞线通过中继器放大信号后，直接连接在一起
- 好处就是便宜！
- 连接过远或过多 Hub 的情况下，CSMA/CD 机制中的碰撞检测失效，导致帧不能正确传输；参见 5-4-3 规则等
- 在高负载时，由于大量帧碰撞，系统吞吐量和延迟急剧恶化

图：四口以太网 Hub，制造商 NETGEAR

为了解决高负载下帧碰撞降低网络性能的问题，人们想到了**交换机（Switch）**：

- 交换机使用了**存储-转发策略**，隔离每个节点之间的普通帧。
- 交换机内部维护一张端口和 MAC 地址的映射表，通过映射表转发包
- 广播帧[1]照例广播（除非用 VLAN 划分更小的子网），所以广播域与使用 Hub 相同

问题 1：当交换机上电或连接新设备时，映射表如何建立？
问题 2：可不可以仅仅使用交换机，构成互联网呢？

[1]为了方便，认为广播帧可达的计算机在同一个**广播域（Broadcast Domain）**中，单播帧可达的计算机在同一个**碰撞域（Collision Domain）**中

只使用交换机连广播域都没法隔离，并且只能单线连接[2]。对此，人们想到了 **IP 协议（Internet Protocol）**：

> *The Internet Protocol is designed for use in interconnected systems of packet-switched computer communication networks. Such a system has been called a "catenet" [1]. The internet protocol provides for transmitting blocks of data called datagrams from sources to destinations, where sources and destinations are hosts identified by fixed length addresses. The internet protocol also provides for fragmentation and reassembly of long datagrams, if necessary, for transmission through "small packet" networks.*

节选自 RFC 791，上面部分描述了 IP 协议的设计动机——为了包转发系统的互联，包的路由和不同系统允许最大帧长度不同的处理。

[2]如果网络中有环，帧就会在环路中反复发送，直到耗尽交换机的全部交换能力。少院机房之前就搞过这个乌龙，还以为是交换机坏了

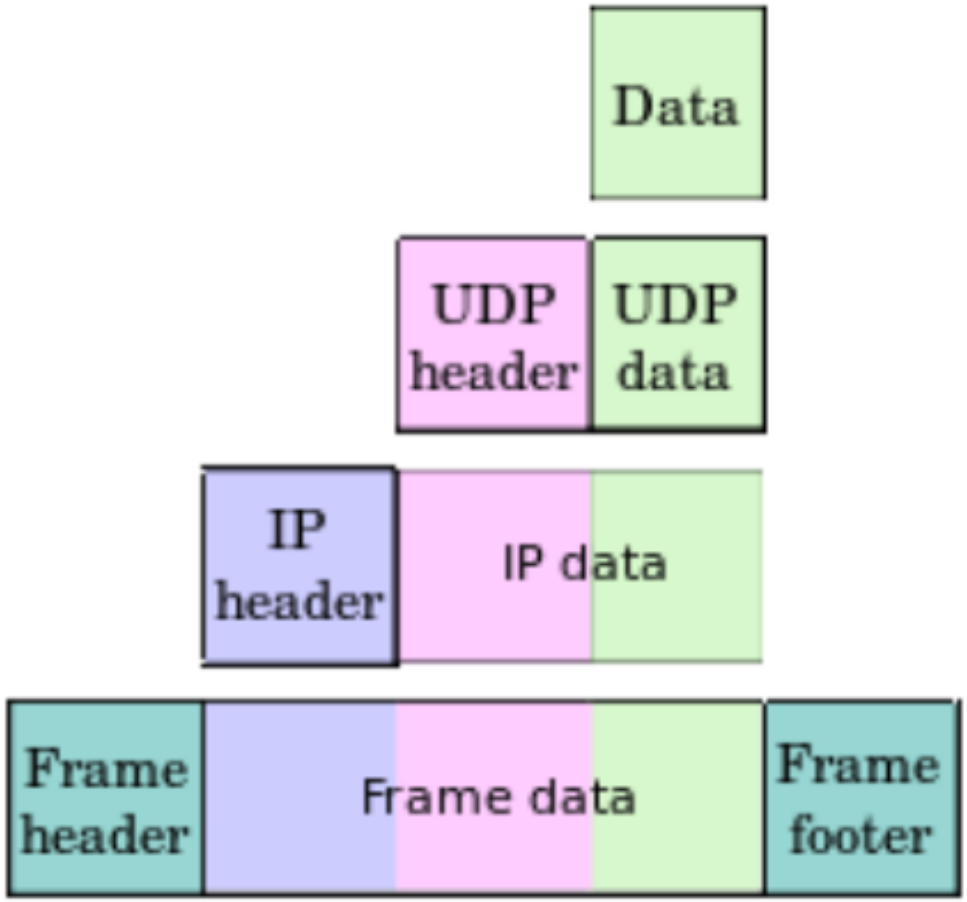
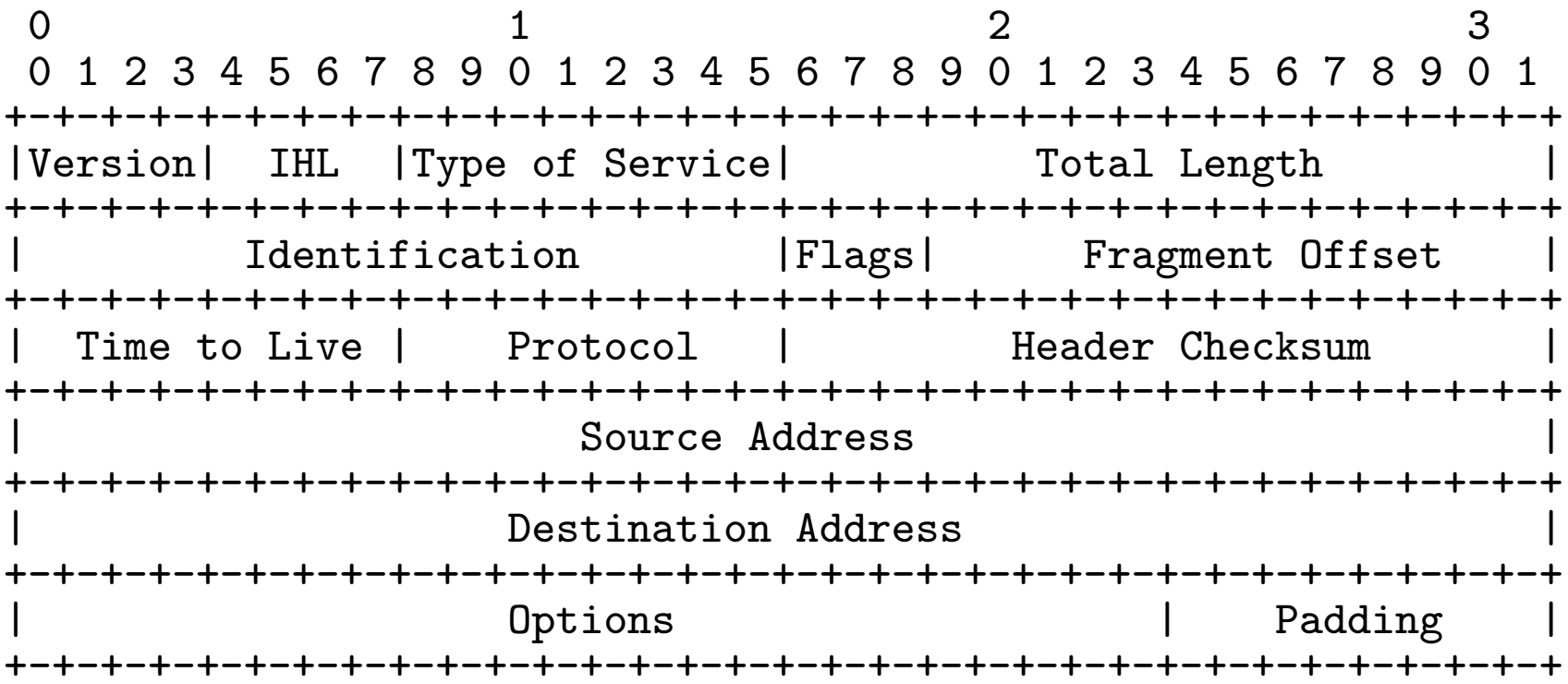

Data
UDP header
UDP data
IP header
IP data
Frame header
Frame data
Frame footer

图: 协议之间的封装关系

Routers and the IP Protocol



上面是一个 IP 包头，节选自 RFC 791。

Routers and the IP Protocol

对于每一个在 IP 层工作的设备，都需要一个 IP 地址。

- 对于 IPv4 协议，地址长度为 32 位，采用点分二进制表示
  例：**192.168.1.254** 对应的二进制地址
- 定义一些特殊地址，用来实现特殊的功能
  - 10.0.0.0/8 [3] 172.16.0.0/12 192.168.0.0/16 用于私有地址
  - 127.0.0.0/8 用于回环地址
  - 其它请参见 https://en.wikipedia.org/wiki/IPv4 中的 Special Use Addresses 一节
- 私有地址约定不能参与 Internet 上的路由[4]。这表示，如果 Internet 上的路由器发现某个包的目的地址是私有地址，它就会被退回。

[3] 此处/8 表示子网掩码的前 8 位是 1，即 255.0.0.0。子网掩码一会介绍。
[4] 一会介绍。

Routers and the IP Protocol

路由（Routing），即为每个 IP 数据包寻找道路。典型的可以处理路由的设备是**路由器（Router）**和配有多网卡的计算机。

- 路由器自己拥有数个 IP 地址（一般每个端口一个 IP 地址）
- 对于发送到路由器的数据包，路由器会查找其维护的**路由表**，根据路由表来决定将把数据包发送到哪个 IP 地址[5]（即**下一跳（next hop）**）
- 两个 IP 地址分属同一**子网（subnet）**，指的是两个 IP 与子网掩码进行 AND 操作后的结果相同
- 如果自己的 IP 地址和数据包目标 IP 在同一子网下，那么路由器就会直接转发到自己的 IP 绑定的那个端口上

[5]你可能会疑问，此时的 MAC 地址应该如何设置？

那么，我们如何得到 IP 地址和 MAC 地址的关系？假设我们要发送一个 IP 数据包到地址 a：

1. 对于不在同一子网的目标地址，计算机将发送**指在 MAC 帧中填入默认网关的 MAC 地址，详见第二条**到其配置的**默认网关**[6]**（Default Gateway）**b
2. 对于同一子网的目标地址，计算机将查询其 **IP 地址-MAC 地址对应表（ARP 表）**，如果表中没有，则通过 ARP 协议发送广播帧寻找目标主机的 MAC 地址
3. 如果寻找到对应 MAC 地址，则将数据包底层的 MAC 帧填入对应 MAC 地址，发送出去；如果没有找到，则发送 ICMP 包“Destination Not Reachable”到源 IP 地址

请注意，以上所有的“发送 IP 包”操作，均为递归过程（233）

[6] 一般来说，默认网关应该和本机 IP 位于同一子网

How we configure routers

小练习：

1. 查看本机的 IP 地址，子网掩码和默认网关；看看你们是不是在一个子网里面
2. 查看本机的 ARP 表和路由表
3. 查看 H3C Magic R2+ 的管理界面

TCP Protocol

TCP，Transmission Control Protocol：提供基于流的网络数据传输

- 连接是流式的，这意味着我们不用关心哪个数据包先到，哪个数据包后到；从网络获得数据就像从文件读取数据一样自然
- 支持对错误/丢失的部分重新传输，不需要应用层关心
- 提供 65535 个端口号[7]，用来区分不同的连接

正因为如此，HTTP、FTP 等多基于 TCP 构建。

[7]请注意，端口号是这里引入的

DHCP Protocol

传统上，我们可以手工配置 IP 地址，但是效率很低；两台同一局域网的机器使用了相同 IP 地址，则无法通信。使用 **DHCP (Dynamic Host Configuration Protocol)**：

- 在每台机器（“**DHCP 客户机**”）发出 DHCP 请求时，DHCP 服务器提供一个 IP 地址、子网掩码和默认网关
- 在 Windows 中，使用“自动获得 IP 地址”来使用 DHCP
- 常用的 DHCP 服务器软件：ISC dhcpd，dnsmasq

Network Simulator

用 Cisco Packet Tracer 来模拟网络配置。
问题：观察 Hub、Switch、Router 和 DHCP、IP 数据包的行为

PXE Stack

- **ssh tempuser@192.168.4.233** - initiate a connection to remote machine, with username tempuser.
  When username was not specified, the ssh client will attempts to connect with local username that's using.
- **scp -r tempuser@192.168.4.233:/home/tempuser /home/tempuser/archive_at_remote** - copy files from remote machine by using *Secure Copy* command

Mission 4: Try copying files **flag.txt** from my laptop by using the account I given.