1. 网络

1.1 什么是网络

他们为什么能打电话？电话是怎么让不在同一个地方的人交流的？

通俗一点来说，网络就是一种辅助双方或者多方能够连接在一起的工具

1.2 网络的作用

就是为了联通多方然后进行通信用的，即把数据从一方传递给另外一方

前面的学习编写的程序都是单机的，不能和别的电脑上的程序进行通信

为了让在不同的电脑上运行的软件，之间能够互相传递数据，就需要借助网络的功能

1.3 小结

- 使用网络能够把多方链接在一起，然后可以进行数据传递
- 所谓的网络编程就是，让在不同的电脑上的软件能够进行数据传递，即进程之间的通信

2. 网络协议

2.1 什么是协议?

协议，协议，其实就是一个规定

我们中国这么大，大家都来自五湖四海，每个地方的方言都不一样吧,只有同一个地方的人才能听得懂，现在为了解决这个问题，国家规定用普通话作为通用语言，这就是一个规定，这就是协议

2.2 网络协议

现在的生活中，不同的计算机只需要能够联网（有线无线都可以）那么就可以相互进行传递数据。那么不同种类之间的计算机到底是怎么进行数据传递的呢？就像说不同语言的人沟通一样，只要有一种大家都认可都遵守的协议即可，那么这个计算机都遵守的网络通信协议叫做 TCP/IP 协议

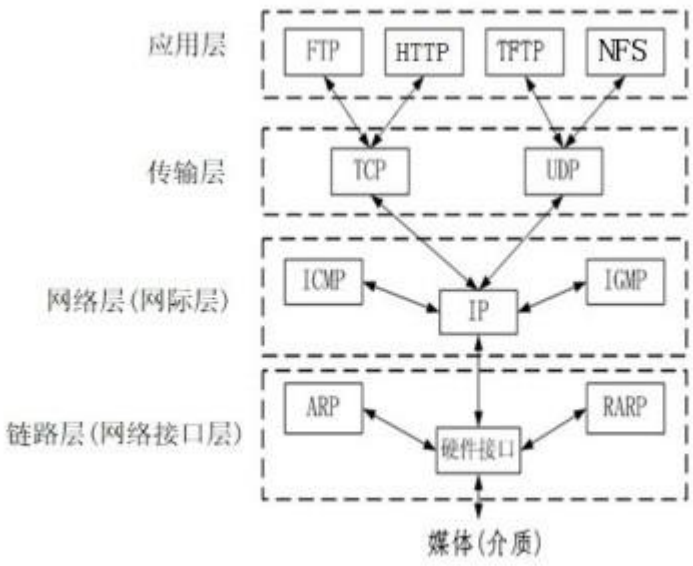
2.3 TCP/IP协议(族)

早期的计算机网络，都是由各厂商自己规定一套协议，IBM、Apple 和 Microsoft 都有各自的网络协议，互不兼容。

为了把全世界的所有不同类型的计算机都连接起来，就必须规定一套全球通用的协议，为了实现互联网这个目标，互联网协议族（Internet Protocol Suite）就是通用协议标准。

因为互联网协议包含了上百种协议标准，但是最重要的两个协议是 TCP 和 IP 协议，所以，大家把互联网的协议简称 TCP/IP 协议(族)。

TCP/IP 协议是一个大家族，不仅仅只有 TCP 和 IP 协议，它还包括其它的协议，如下图：



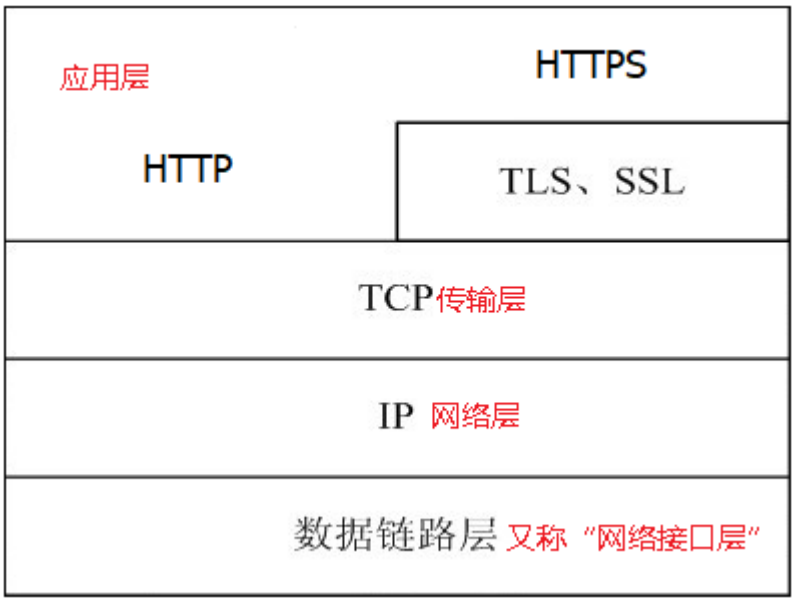
- ARP：是正向地址解析协议（Address Resolution Protocol），通过已知的 IP，寻找对应主机的 MAC 地址。
- RARP：是反向地址转换协议，通过 MAC 地址确定 IP 地址。
- IP：是因特网互联协议（Internet Protocol）
- ICMP：是 Internet 控制报文协议（Internet Control Message Protocol）它是 TCP/IP 协议族的一个子协议，用于在 IP 主机、路由器之间传递控制消息。
- IGMP：是 Internet 组管理协议（Internet Group Management Protocol），是因特网协议家族中的一个组播协议。该协议运行在主机和组播路由器之间。
- TCP：传输控制协议（Transmission Control Protocol），是一种面向连接的、可靠的、基于字节流的传输层通信协议。
- UDP：用户数据报协议（User Datagram Protocol），是 OSI 参考模型中一种无连接的传输层协议，提供面向事务的简单不可靠信息传送服务。
- HTTP：超文本传输协议（Hyper Text Transfer Protocol），是互联网上应用最为广泛的一种网络协议。
- FTP：文件传输协议（File Transfer Protocol）

2.4 HTTP协议

浏览器通过网络和 web 服务器通信,，那么服务器如何知道浏览器想要什么数据？服务器返回给浏览器的数据，浏览器如何区分数据到底是图片、还是音乐、还是普通文本数据？这一些都需要约定好，不然不同的浏览器和服务器就无法通信。负责解决该通信数据格式的技术叫做“HTTP 协议”。

HTTP协议（Hyper Text Transfer Protocol，超文本传输协议）是互联网上应用最为广泛的一种网络协议,是用于从web服务器传输超文本到本地浏览器的传输协议，是因特网中的“多媒体信使”。它可以使浏览器更加高效，使网络传输减少。它不仅保证计算机正确快速地传输超文本文档，还确定传输文档中的哪一部分，以及哪部分内容首先显示(如文本先于图形)等。同时，HTTP使用的是可靠的数据传输协议，即使是来自于地球另一端的数据，它也可以确

保数据在传输的过程中不会丢失和损坏，保证了用户在访问信息时的完整性。HTTP是互联网上应用最为广泛的一种协议，后面还会介绍其他的互联网常用协议（https,ftp,file,mailto等）。按照上述点餐流程理解的话就是厨师具备煎、炒、烹、炸、溜、爆、煸、蒸、烧、煮等多种烹调技法，你需要告诉厨师这道菜怎么做。



2.5 无状态协议是什么

无状态协议是指比如客户获得一张网页之后关闭浏览器， 然后再一次启动浏览器，再登录 该网站，但是服务器并不知道客户关闭了一次浏览器。

HTTP 协议是无状态的，指的是协议对于事务处理没有记忆能力，服务器不知道客户端是什 么状态。也就是说，打开一个服务器上的网页和你之前打开这个服务器上的网页之间没有任 何联系。

HTTP 是一个无状态的面向连接的协议，无状态不代表 HTTP 不能保持 TCP 连接，更不能 代表 HTTP 使用的是 UDP 协议（无连接）。

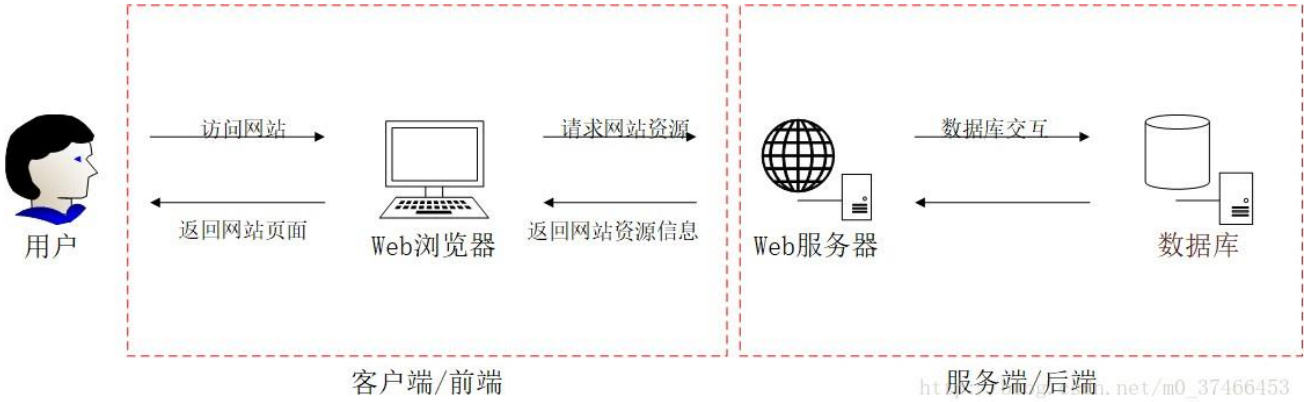
2.5.1主要特点

1. 简单快速：客户向服务器请求服务时，只需传送请求方法和路径。请求方法常用的有GET、HEAD、POST。每种方法规定了客户与服务器联系的类型不同。由于HTTP协议简单，使得HTTP服务器的程序规模小，因而通信速度很快。
2. 灵活：HTTP允许传输任意类型的数据对象。正在传输的类型由Content-Type加以标记。
3. 无连接：无连接的含义是限制每次连接只处理一个请求。服务器处理完客户的请求，并收到客户的应答后，即断开连接。采用这种方式可以节省传输时间。
4. 无状态：HTTP协议是无状态协议。无状态是指协议对于事务处理没有记忆能力。缺少状态意味着如果后续处理需要前面的信息，则它必须重传，这样可能导致每次连接传送的数据量增大。另一方面，在服务器不需要先前信息时它的应答就较快。
5. 支持B/S及C/S模式。

3. web基础

3.1. Web工作流程

首先，我们先来思考一下我们平常在上网浏览网页时候的场景，大致就是打开一个web浏览器，输入某一个网站的地址，然后转到该网址，在浏览器中得到该网址的页面。从这个场景中我们可以抽象出来几个基本对象，我们（用户）、web浏览器（客户端）和发送过来页面的地方（服务端），这些对象其实就是整个web工作流程中的重要组成部分。



为了加强理解，其实可以将这个工作流程看做去吃饭时点餐的流程，web浏览器就是服务员，而服务端就是厨房。你给服务员说你要点什么菜，然后服务员将你点的菜端上来，具体厨房里是怎么忙活的也并不知道，其实web服务器就相当于厨师，有着各种各样的技能，根据你的成菜要求，为你进行服务，数据库在这里可以认为是个菜窖，需要什么菜去拿什么菜。

3.2 域名

我们在访问一台服务器的时候，需要记住该服务器的 IP 地址，由于 IP 地址不利于人们记忆，所以推出的域名技术。

域名是由一串用点分隔的名字组成的 Internet 上某一台计算机或计算机组的名称，用于在数据传输时标识计算机的位置。

域名可以用来表示一个单位、机构或可以利用个人在 Internet 上 的确定的名称或位置。域名是惟一的，客户可以利用这个名字找寻有关的产品和服务信息。

3.3 DNS

由于我们用域名来标识计算机的位置，但是我们前面讲过，网络上标识主机的唯一标识是IP 地址，所以就需要记录一下，一个域名和 IP 地址的对应关系，这个对应关系就存储在 DNS服务器中，当我们向 DNS 发出请求时，DNS会返回给我们域名所对应的 IP 地址。

3.4 ip地址

IP地址是指互联网协议地址（英语：Internet Protocol Address，又译为网际协议地址），是IP Address的缩写。IP地址是IP协议提供的一种统一的地址格式，它为互联网上的每一个网络和每一台主机分配一个逻辑地址，以此来屏蔽物理地址的差异。

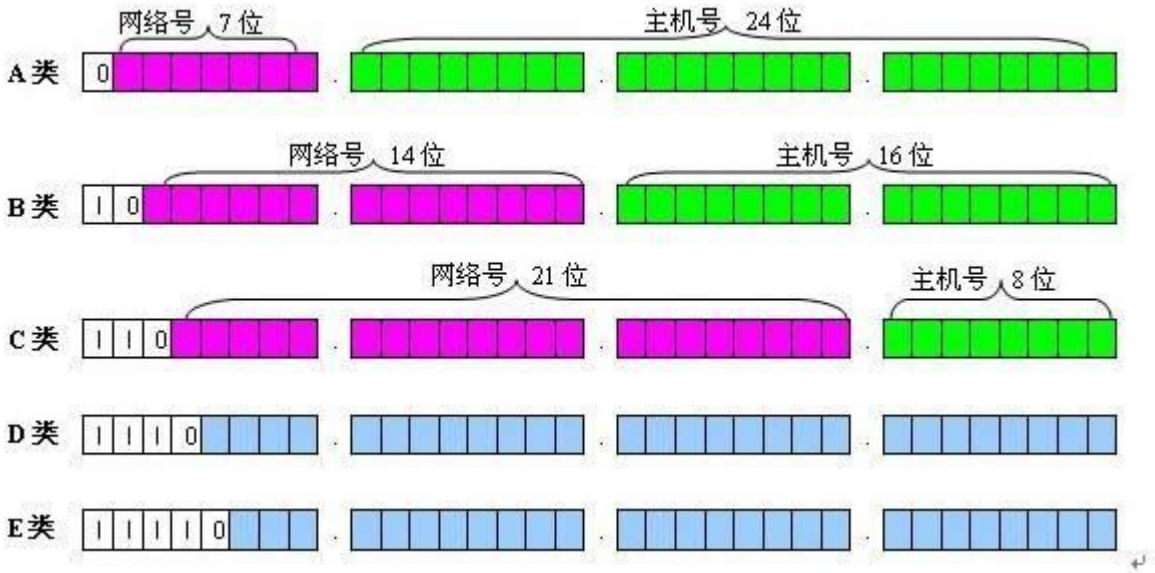
3.4.1 ip地址的作用

很显然了，ip地址的作用就跟生活中地址的作用一样的，你买东西要填了地址别人才知道把东西发到哪

ip地址是用来在网络中标记一台电脑的一串数字，比如192.168.1.1；在本地局域网上是独一无二的。

3.4.2 ip地址的分类

最初设计互联网络时，为了便于寻址以及层次化构造网络，每个IP地址包括两个标识码（ID），即网络ID和主机ID。同一个物理网络上的所有主机都使用同一个网络ID，网络上的一个主机（包括网络上工作站，服务器和路由器等）有一个主机ID与其对应。Internet委员会定义了5种IP地址类型以适合不同容量的网络，即A类~E类。



1) A类IP地址

一个A类IP地址由1字节的网络地址和3字节主机地址组成，网络地址的最高位必须是“0”。

网络标识长度为8位，主机标识的长度为24位，子网掩码为255.0.0.0

网络地址数量较少，有126个网络，每个网络支持的最大主机数为256的3次方-2=16777214台。

地址范围 1.0.0.1 - 126.255.255.254

二进制表示为：00000001 00000000 00000000 00000000 - 01111110 11111111 11111111 11111111

2) B类IP地址

一个B类IP地址就由2字节的网络地址和2字节主机地址组成，网络地址的最高位必须是“10”。

网络标识长度为16位，主机标识的长度为16位，子网掩码为255.255.0.0

适用于中等规模的网络，有16384个网络，每个网络支持的最大主机数为256的2次方-2=65534台。

地址地址范围 128.0.0.1—191.255.255.254

二进制表示为：10000000 00000000 00000000 00000000 - 10111111 11111111 11111111 11111111

3) C类IP地址

一个C类IP地址就由3字节的网络地址和1字节主机地址组成，网络地址的最高位必须是“110”。

网络标识长度为24位，主机标识的长度为8位，子网掩码为255.255.255.0

网络地址数量较多，有209万余个网络。适用于小规模的局域网络，每个网络支持的最大主机数为256-2=254台。

地址范围192.0.0.0-223.255.255.255

二进制表示为: 11000000 00000000 00000000 00000000 - 11011111 11111111 11111111 11111111

4) D类IP地址

D类IP地址在历史上被叫做多播地址(multicast address)，即组播地址。

在以太网中，多播地址命名了一组应该在这个网络中应用接收到一个分组的站点。多播地址的最高位必须是“1110”，

地址范围从224.0.0.1到239.255.255.254

5) E类IP地址

以“1111”开始，为将来使用保留

地址范围：240.0.0.1—255.255.255.254

E类地址保留，仅作实验和开发用

6) 私有ip

在这么多网络IP中，国际规定有一部分IP地址是用于我们的局域网使用，也就

是属于私网IP，不在公网中使用的，它们的范围是：

```
10.0.0.0～10.255.255.255
172.16.0.0～172.31.255.255
192.168.0.0～192.168.255.255
```

7) 注意

IP地址127．0．0．1~127．255．255．255用于回路测试，

如：127.0.0.1可以代表本机IP地址，用 http://127.0.0.1 就可以测试本机中配置的Web服务器。

4. 端口

如果把IP地址比作一间房子 ，端口就是出入这间房子的门。

在Internet上，各主机间通过TCP/IP协议发送和接收数据包，各个数据包根据其目的主机的ip地址来进行互联网络中的路由选择,把数据包顺利的传送到目的主机。大多数操作系统都支持多程序（进程）同时运行，那么目的主机应该把接收到的数据包传送给众多同时运行的进程中的哪一个呢？显然这个问题有待解决，端口机制便由此被引入进来。

4.1 端口的作用

我们知道，一台拥有IP地址的主机可以提供许多服务，比如HTTP（万维网服务）、FTP（文件传输）、SMTP（电子邮件）等，这些服务完全可以通过1个IP地址来实现。那么，主机是怎样区分不同的网络服务呢？显然不能只靠IP地址，因为IP地址与网络服务的关系是一对多的关系。实际上是通过“IP地址+端口号”来区分不同的服务的。

4.2 端口号

端口是通过端口号来标记的，端口号只有整数，范围是从0到65535

4.3 端口分配

端口号不是随意使用的，而是按照一定的规定进行分配。

端口的分类标准有好几种，我们这里不做详细讲解，只介绍一下知名端口和动态端口

4.3.1 知名端口（Well Known Ports）

知名端口是众所周知的端口号，范围从0到1023

```
80端口分配给HTTP服务
21端口分配给FTP服务
```

可以理解为，一些常用的功能使用的号码是估计的，好比 电话号码110、10086、10010一样,一般情况下,如果一个程序需要使用知名端口的需要有root权限

4.3.2 动态端口（Dynamic Ports）

动态端口的范围是从1024到65535

之所以称为动态端口，是因为它一般不固定分配某种服务，而是动态分配。

动态分配是指当一个系统进程或应用程序进程需要网络通信时，它向主机申请一个端口，主机从可用的端口号中分配一个供它使用。

当这个进程关闭时，同时也就释放了所占用的端口号。

4.3.4 怎样查看端口？

用“netstat －an”查看端口状态

4.4 小结

端口并不是一一对应的。比如你的电脑作为客户机访问一台WWW服务器时，WWW服务器使用“80”端口与你的电脑通信，但你的电脑则可能使用“4573”这样的端口。

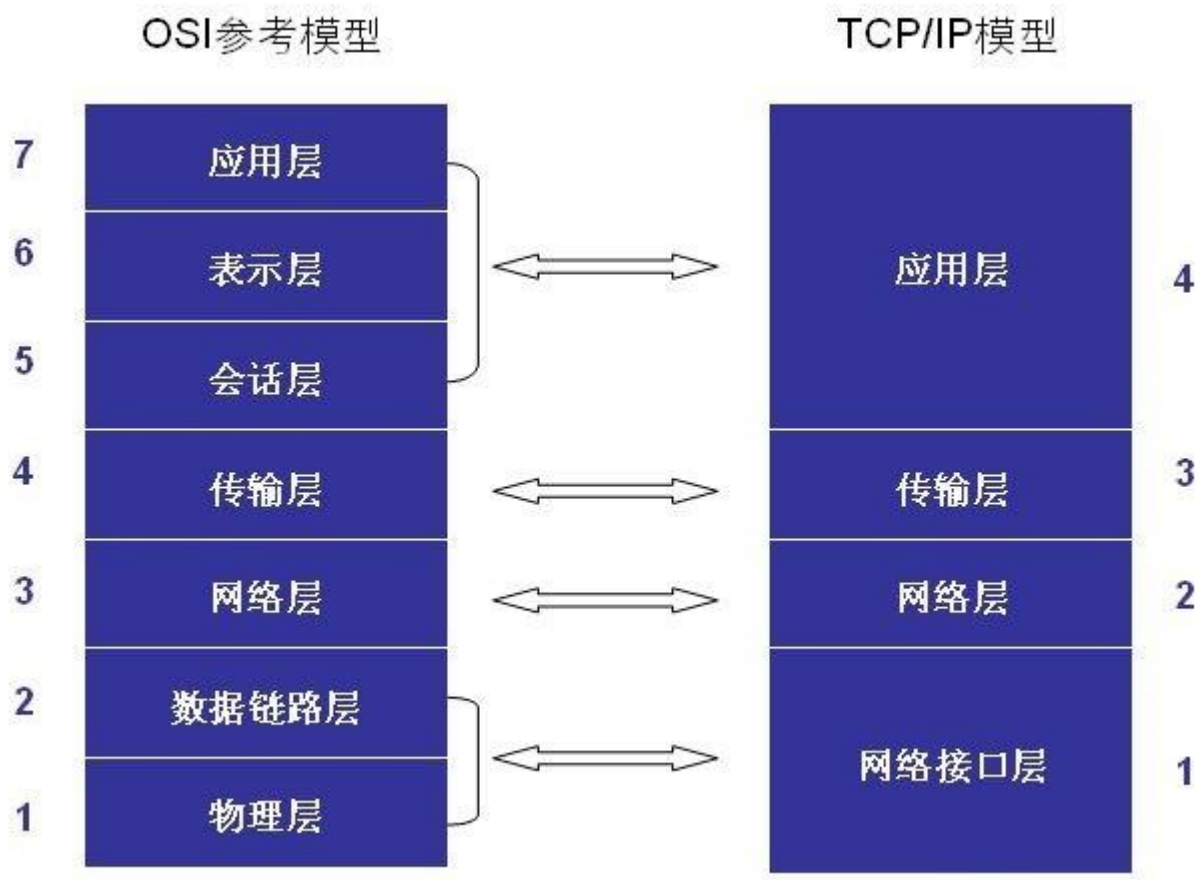
5. 分层模型

5.1 网络分层架构

OSI/RM（Open System Interconnection/Reference Model），全称为开放系统互连参考模型，是由国际标准化组织（ISO，International Standard Organization）提出来的一种网络互连模型（旨在成为一个所有销售商都能实现的开放网络模型，来解决众多私有网络模型带来的困难和低效性）。虽然ISO参考模型的实际应用意义不是很大，但是对我们理解网络协议的内部运作很有帮助。

OSI模型把网路通信工作分为七层，每一层为上一层提供服务，并为其上一层提供一个访问接口或者界面。

不同主机之间的相同层称为对等层，对等层之间的通信需要需要遵循一定的规则，如通信内容，通信方式，称为协议（Protocol）。网络中各层的协议总和称为协议栈，其形象的反应了一个网络中文件的传输过程：从上层协议到底层协议，再由底层协议到上层协议。使用最广泛的就是TCP /IP协议栈。



越下面的层，越靠近硬件；越上面的层，越靠近用户。至于每一层叫什么名字，其实并不重要（面试的时候，面试官可能会问每一层的名字）。只需要知道，互联网分成若干层即可。

5.2 OSI七层模型分层

- 物理层（Physical Layer）：为上层协议提供了一个传输数据的物理媒体。

  数据单位：比特（bit）

  典型设备：光纤、双绞线、同轴电缆、集线器

- 数据链路层（Data Link Layer）：在不可靠的物理媒介上进行可靠的传输，该层的作用包括：物理地址寻址、数据的成帧、流量控制、数据的检错、重发等。

  数据单位：帧（frame）

  典型设备：交换机、网卡

- 网络层（Network Layer）：负责对子网间的数据包进行路由选择。此外，网络层还可以实现拥塞控制、网际互连等功能。

  数据单位：数据包（packet）

  典型设备：路由器

- 传输层（Transport Layer）：负责将上层数据分段，并提供端到端，可靠或不可靠的传输。此外，传输层还要处理端到端的差错控制和流量控制问题。

  数据单位：数据段（segment）

  典型设备：网关

- 会话层（Session Layer）：管理主机之间的会话进程，即负责建立、管理、终止进程之间的会话。会话层还利用在数据中插入校验点来实现数据的同步。
- 表现层（Presentation Layer）：对上层数据或信息进行变换以保证一个主机应用层信息可以被另一个主机的应用程序理解。表示层的数据转换包括数据的加密、压缩、格式转换等。
- 应用层（Application　Layer）：为操作系统或网络应用程序提供访问网络服务的接口。

5.3 TCP/IP协议族四层模型分层

- 网络接口层 负责接收IP数据包并通过网络发送，或者接收网络中的物理帧，提取出IP包，交给上一层。网
- 络层：完成不同主机之间的通信，包括三方面的功能
    1. 处理来自传输层的分组发送请求，收到请求后，将分组装入IP数据包，填充报头，选择去往信宿机的路径，然后将数据报发往适当的网络接口。
    2. 处理输入数据包：首先检查其合法性，然后进行寻径--假如该数据报已到达信宿机，则去掉报头，将剩下部分交给适当的传输协议；假如该数据报尚未到达信宿，则转发该数据报。
    3. 处理路径、流控、拥塞等问题。
- 传输层：提供应用程序间的通信。
- 应用层：TCP/IP模型将OSI参考模型中的会话层和表示层的功能合并到应用层实现。

在TCP/IP协议族中，传输层位于网络层之上，传输层为主机上运行的不同应用程序（进程）提供逻辑通信，而网络层则是为不同主机之前提供逻辑通信。

网络层中有一个IP（Internet Protocol）协议，提供了主机之间的逻辑通信。IP服务模型是一个尽力传送服务。这就意味着IP尽它最大的努力在通信主机之间传送数据段，但是却不提 供任何保障。特别是，它不能保证数据段传输的安全性，不能保证数据段的顺序传输，不能保证数据段传输的数据完整性。基于这些原因，IP服务被称为不可靠服务。

传输层包含两个主要的协议TCP（Transmission Control Protocol）和UDP（User Datagram Protocol）协议。TCP为应用程序提供了一种可靠的、面向连接的服务。UDP则是一种不可靠、无连接的协议。当设计一种网络应用程序时，应用程序外发着需要指定这两种协议中的一种。

6. socket

6.1 网络中进程之间如何通信

本地的进程间通信（IPC）有很多种方式，但可以总结为下面4类：

消息传递（管道、FIFO、消息队列）

同步（互斥量、条件变量、读写锁、文件和写记录锁、信号量）

共享内存（匿名的和具名的）

远程过程调用（Solaris门和Sun RPC）

网络中进程之间如何通信?

首要解决的问题是如何唯一标识一个进程，否则通信无从谈起！在本地可以通过进程PID来唯一标识一个进程，但是在网络中这是行不通的。其实TCP/IP协议族已经帮我们解决了这个问题，网络层的“ip地址”可以唯一标识网络中的主机，而传输层的“协议+端口”可以唯一标识主机中的应用程序（进程）。这样利用三元组（ip地址，协议，端口）就可以标识网络的进程了，网络中的进程通信就可以利用这个标志与其它进程进行交互。

使用TCP/IP协议的应用程序通常采用应用编程接口：UNIX BSD的套接字（socket）和UNIX System V的TLI（已经被淘汰），来实现网络进程之间的通信。就目前而言，几乎所有的应用程序都是采用socket，而现在又是网络时代，网络中进程通信是无处不在。

6.2 什么是socket

网络上的两个程序通过一个双向的通信连接实现数据的交换，这个连接的一端称为一个socket。

建立网络通信连接至少要一对端口号(socket)。socket本质就是编程接口(API)，对TCP/IP的封装，TCP/IP也要提供可供程序员做网络开发所用的接口，这就是Socket编程接口；HTTP是轿车，提供了封装或者显示数据的具体形式；Socket是发动机，提供了网络通信的能力。

socket是进程间通信的一种方式，它与其他进程间通信的一个主要不同是：

它能实现不同主机间的进程间通信，我们网络上各种各样的服务大多都是基于 Socket 来完成通

信的例如我们每天浏览网页、QQ 聊天、收发 email 等等

7. socket的创建

在 Python 中 使用 socket 模块的函数 socket 就可以完成：

```
import socket
socket.socket(AddressFamily, Type)
```

函数 socket.socket 创建一个 socket，该函数带有两个参数：

- AddressFamily：可以选择 AF_INET（用于 Internet 进程间通信） 或者 AF_UNIX（用于同一台机器进程间通信），实际工作中常用 AF_INET
- Type：套接字类型，可以是 SOCK_STREAM（流式套接字，主要用于 TCP 协议）或者 SOCK_DGRAM（数据报套接字，主要用于 UDP 协议）

示例代码：

```
import socket
s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM) #创建 udp 的套接字
#s = socket.socket(socket.AF_INET, socket.SOCK_STREAM) #创建 tcp的套接字
# ...这里是使用套接字的功能（省略）...
# 不用的时候，关闭套接字
s.close()
```

套接字使用流程与文件的使用流程很类似：

1. 创建套接字
2. 使用套接字收/发数据
3. 关闭套接字