中山大學
SUN YAT-SEN UNIVERSITY

操作系统

笔记整理

姓名：刘斯宇
学号：17341110

目录

☒ 第一章--概述

1 概述

1.1 因特网概述

网络(network) 由若干结点 (node) 和连接这些结点的链路 (link) 组成。网络中的结点可以是计算机、集线器、交换机或路由器等。

网络和网络还可以通过路由器互连起来，这样就构成了一个覆盖范围更大的网络，即互联网(或互连网)，因此互联网是“网络中的网络”。

因特网 (Internet) 是世界上最大的互连网络。

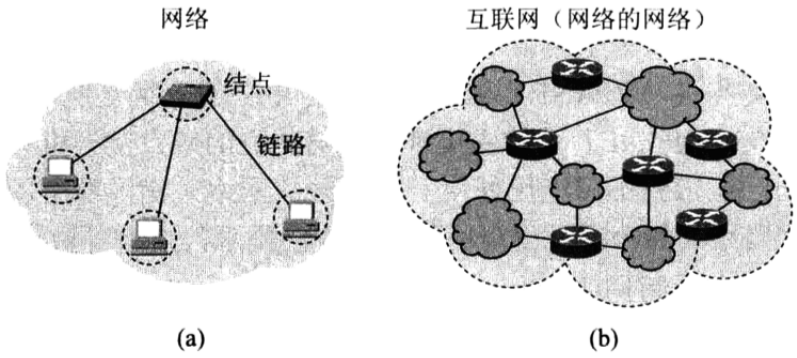

图 1: 网络和互联网的区别

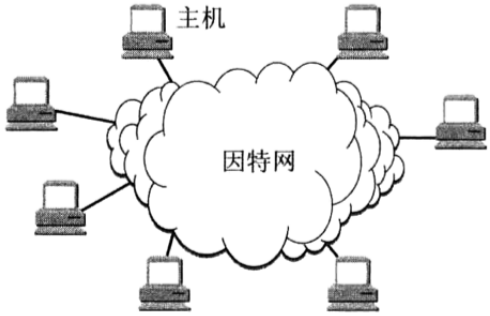

图 2: 因特网与连接的主机

以小写字母 i 开始的 internet (互联网或互连网) 是一个通用名词，它泛指由多个计算机网络互连而成的网络。在这些网络之间的通信协议 (即通信规则) 可以是任意的。

以大写字母 I 开始的 Internet (因特网) 则是一个专用名词，它指当前全球最大的、开放的、由众多网络相互连接而成的特定计算机网络，它采用 TCP/IP 协议族作为通信的规则，且其前身是美国的 ARPANET.

因特网服务提供者 ISP:ISP 可以从因特网管理机构申请到很多 IP 地址 (因特网上的主机都必须有 IP 地址才能上网，同时拥有通信线路以及路由器等连网设备，因此任何机构和个人只要向某个

ISP 交纳规定的费用，就可从该 ISP 获取所需 IP 地址的使用权，并可通过该 ISP 接入到因特网。所谓“上网”就是指 (通过某个 ISP 获得的 IP 地址) 接入到因特网。

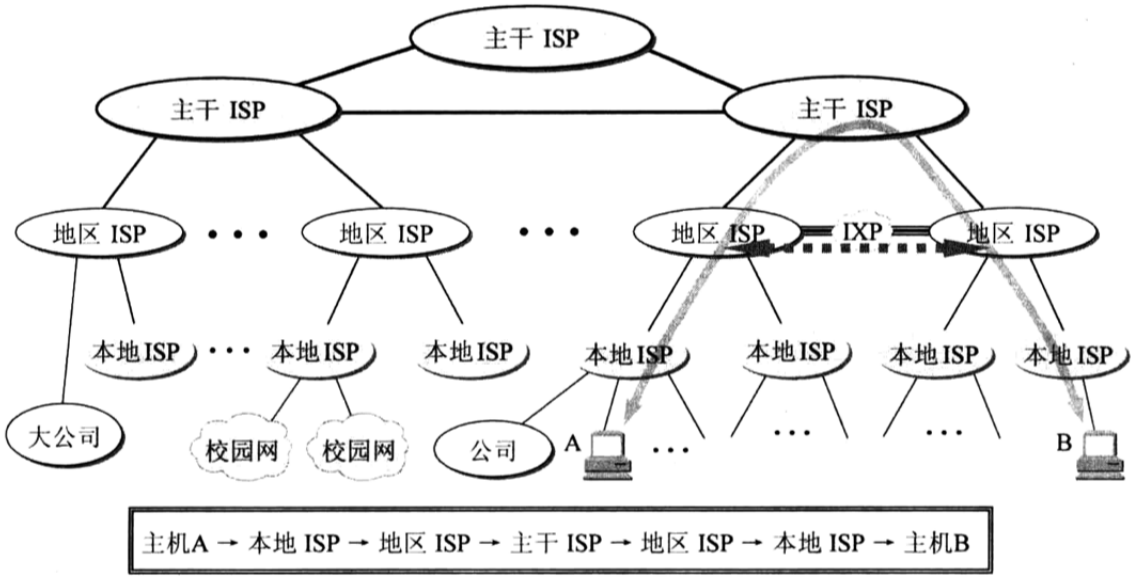

图 3: 基于 ISP 的多层结构的因特网的概念示意图

因特网交换点 IXP 的主要作用就是允许两个网络直接相连并交换分组，而不需要再通过第三个网络来转发分组。例如，在图中右方的两个地区 ISP 通过一个 IXP 连接起来了。这样，主机 A 和主机 B 交换分组时，就不必再经过最上层的主干 ISP，而是直接在两个地区 ISP 之间用高速链路对等地交换分组。这样就使因特网上的数据流量分布更加合理，同时也减少了分组转发的迟延时间，降低了分组转发的费用。

1.2 因特网的组成

(1) 边缘部分：由所有连接在因特网上的主机组成。这部分是用户直接使用的，用来进行通信 (传送数据、音频或视频) 和资源共享。

(2) 核心部分：由大量网络和连接这些网络的路由器组成。这部分是为边缘部分提供服务的 (提供连通性和交换)。

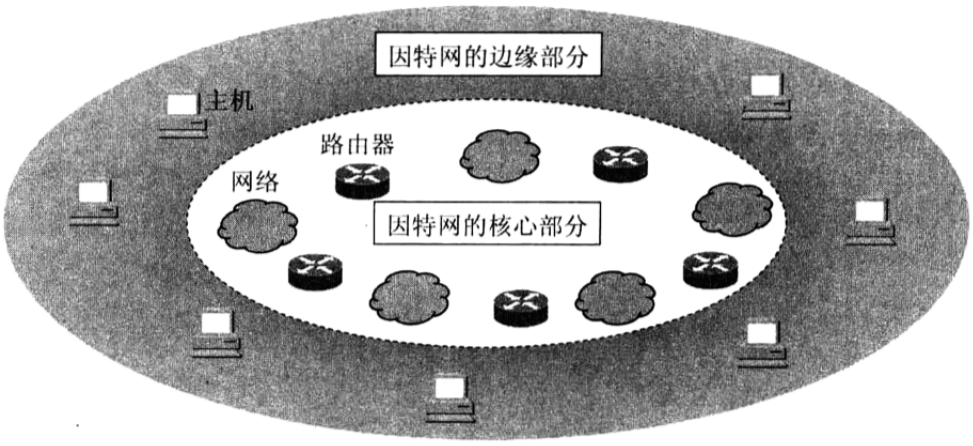

图 4: 因特网的边缘部分与核心部分

在网络边缘的端系统之间的通信方式通常可以划分为两大类：客户-服务器和对等方式。

1.2.1 客户-服务器方式

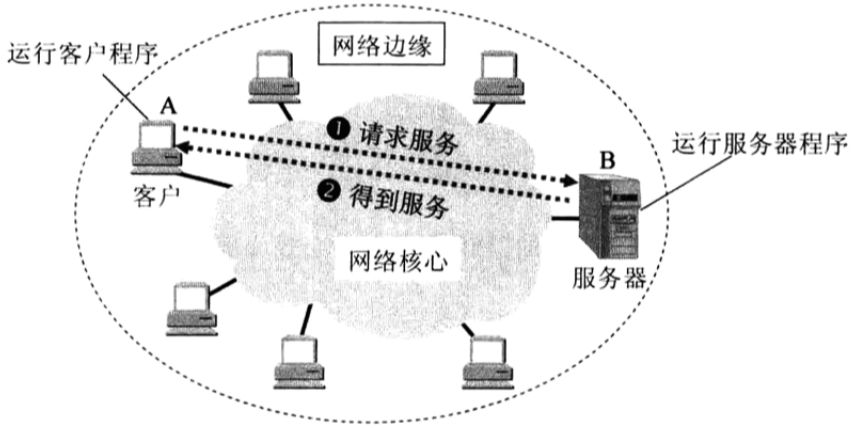

图 5: 客户-服务器方式

1.2.2 对等连接方式

对等连接 (peer-to-peer，简写为 P2P) 是指两个主机在通信时并不区分哪一个是服务请求方还是服务提供方。只要两个主机都运行了对等连接软件 (P2P 软件)，它们就可以进行平等的、对等连接通信。这时，双方都可以下载对方已经存储在硬盘中的共享文档。因此这种工作方式也称为 P2P 文件共享。

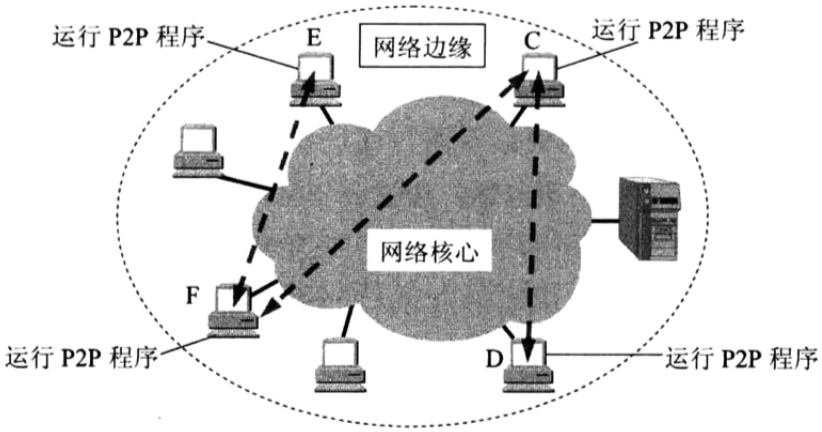

图 6: 对等连接方式

实际上，对等连接方式从本质上看仍然是使用客户-服务器方式，只是对等连接中的每一个主机既是客户又同时是服务器。例如主机 C，当 C 请求 D 的服务时，C 是客户，D 是服务器。但如果 C 又同时向 F 提供服务，那么 C 又同时起着服务器的作用。

在网络核心部分起特殊作用的是路由器 (router)，它是一种专用计算机 (但不是主机)。路由器是实现分组交换 (packet switching) 的关键构件，其任务是转发收到的分组，这是网络核心部分最重要的功能。

1.2.3 电路交换

这种必须经过“建立连接 (占用通信资源)→ 通话 (一直占用通信资源)→ 释放连接 (归还通信资源)”三个步骤的交换方式称为电路交换。

电路交换的一个重要特点就是在通话的全部时间内，通话的两个用户始终占用端到端的通信资源。

1.2.4 分组交换的主要特点

分组交换则采用存储转发技术。通常我们把要发送的整块数据称为一个报文 (message)。

电路交换：整个报文的比特流连续地从源点直达终点，好像在一个管道中传送。

报文交换：整个报文先传送到相邻结点，全部存储下来后查找转发表，转发到下一个结点。

分组交换：单个分组 (这只是整个报文的一. 部分) 传送到相邻结点，存储下来后查找转发表，转发到下一个结点。

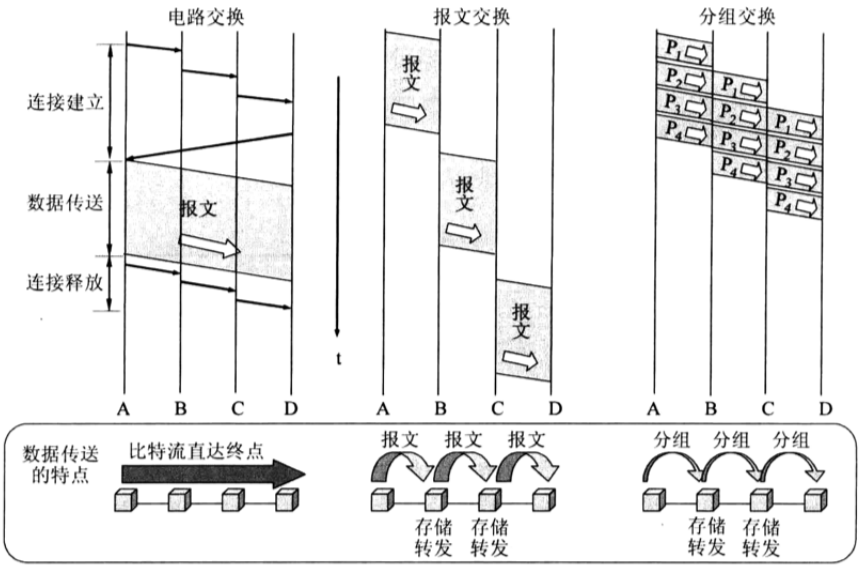

图 7: 三种交换的比较

1.3 计算机网络体系结构

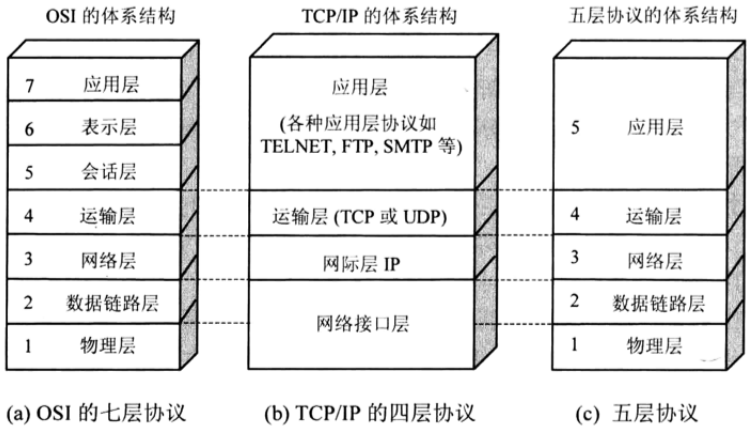

图 8: 计算机网络体系结构

协议栈：由千计算机网络的体系结构采用了分层结构，因此不论在主机中还是在路由器中的协议都有好几层。这些一层一层的协议画起来就很像堆栈的结构，因此就把这些协议层称为协议栈。

实体：表示任何可发送或接收信息的硬件或软件进程。在许多情况下，实体就是一个特定的软件模块。

对等层：在网络体系结构中，通信双方实现同样功能的层。例如，A 向 B 发送数据，那么 A 的第 n 层和 B 的第 n 层就构成了对等层。

协议数据单元：通常记为 PDU, 它是对等实体之间进行信息交换的数据单元。

服务访问点：通常记为 SAP, 在同一系统中相邻两层的实体进行交互（即交换信息）的地方，通常称为服务访问点。

客户：在计算机网络中进行通信的应用进程中的服务请求方。

服务器：在计算机网络中进行通信的应用进程中的服务提供方。但在很多情况下，服务器也常指运行服务器程序的机器。

1.4 问题解答

一个主机能否同时连接到两种不同的网络上，其中的一个网络采用面向连接的方式通信，而另一个网络采用无连接方式?

可以。一个主机可以使用两个不同的接口。一个接口连接到面向连接的分组交换网（例如 X.25 网），而另一个接口连接到分组交换网（如使用 IP 协议的互联网）。具有多个网络接口的主机叫做“多归属主机" (multi-homed host) 。

我们能否在同一时间，在不同的层次使用不同的连接方式（面向连接和无连接方式）?

当然可以。例如，当我们发送电子邮件时，电子邮件协议需要使用面向连接的 TCP 协议，但 TCP 协议要使用下面的无连接的 IP 协议。IP 协议又使用数据链路层的 PPP 协议，而 PPP 协议是面向连接的。

到商店购买了一个希捷公司生产的 4TB 的硬盘。当安装到电脑上以后，我们使用 Windows 资源管理器在该磁盘的“属性”中发现只有 3 .63 TB 。是什么地方出了差错吗?

解答：不是。这个因为希捷公司的硬盘标记中的 T 表示 10^{12} 而微软公司 Windows 软件中的 T 表示 2^{40} 。$3.63 \times 2^{40} = 4 * 10^{12}$ 即希捷公司的 4TB 和微软公司的 3.63 TB 相等。

为什么协议不能设计成 100% 可靠的?

解答：设想某一个要求达到 100% 可靠的协议需要 A 和 B 双方交换信息共 N 次，而这 N 次交换信息都是必不可少的。也就是说，在所交换的 N 次信息中是没有冗余的。假定第 N 次交换的信息是从 B 发送给 A 的。B 发送给 A 的这个信息显然是需要 A 加以确认的。这是因为：若不需要 A 的确认，则表示 B 发送这个信息丢失了或出现差错都不要紧。这就是说，B 发送的这个信息是可有可无的。如果 B 发送的这个信息是可有可无的，那么最启这次的信息交换就可以取消，因而这个协议就只需要 A 和 B 交换信息 N - 1 次而不是 N 次。这就和原有的假定不符。如果 B 发送的这个最后的信息是需要 A 加以确认的，那么这个协议需要 A 和 B 交换信息的次数就不是 N 次，而是还要增加一次确认 C A 向 B 发送的确认），即总共需要交换信息 N+ 1 次。但这就和原来假定的“双方交换信息共 N 次”相矛盾。显然，这个矛盾无法解决。这样就证明了协议不能设计成 100% 可靠的。然而在非常重要的任务中，协议可以设计成非常接近千 100% 可靠的。

2　物理层

3 数据链路层

3.1 使用点对点信道的数据链路层

4 网络层

4.1

5 传输层

TCP 是面向连接的，但 TCP 使用的 IP 却是无连接的。这两种协议都有哪些主要的区别?

TCP 是面向连接的，但 TCP 所使用的网络则可以是面向连接的 (如 X.25 网络，已经淘汰) 也可以是无连接的 (如现在大量使用的 IP 网络)。选择无连接网络就使得整个系统非常灵活，当然也带来了一些问题。显然，TCP 所提供的功能和服务要比 IP 所能提供的功能和服务多得多。这是因为 TCP 使用了诸如确认、滑动窗口、计时器等机制，因而可以检测出有差错的报文、重复的报文和失序的报文。

5.1 端口

前面讲过数据链路层按 MAC 地址寻址，网络层按 IP 地址寻址，而传输层是按端口号寻址的。端口就是传输层服务访问点，端口能够让应用层的各种应用进程将其数据通过端口向下交付给传输层以及让传输层知道应当将其报文段中的数据向上通过端口交付给应用层的相应进程。

由于同一时刻一台主机上会有大量的网络应用进程在运行，所以需要有大量的端口号来标识不同的进程。

1) 熟知端口 (保留端口): 数值- -般为 0 1023。当一种新的应用程序出现时，必须为它指派一个熟知端口，以便其他应用进程和其互。

2) 登记端口: 数值为 1024 49 151。它是为没有熟知端口号的应用程序使用的，使用这类端口号必须在 IANA 登记，以防止重复。

3) 客户端端口或短暂端口: 数值为 49152 65535。由于这类端口号仅在客户进程运行时才动态选择，所以称为短暂端口或临时端口。通信结束后，此端口就自动空闲出来，以供其他客户进程使用。

应用程序	FTP	TELNET	SMTP	DNS	TFTP	HTTP	SNMP
熟知端口	21，20	23	25	53	69	80	161

图 9: 几个常见的熟知端口

套接字 =(主机 IP 地址，端口号)

5.2 UDP

1) 发送数据之前不需要建立连接。

2) UDP 的主机不需要维持复杂的连接状态表。

3) UDP 用户数据报只有 8 个字节的首部开销。

4) 网络出现的拥塞不会使源主机的发送速率降低 (没有拥塞控制)。这对某些实时应用 (如 IP 电话、实时视频会议) 是很重要的。

5) UDP 支持一对一、一对多、多对一和多对多的交互通信。

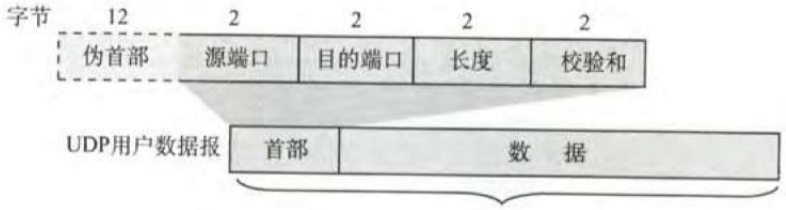

图 10: UDP 数据报的首部格式

其中，伪首部包括源 IP 地址字段、目的 IP 地址字段、全 0 字段、协议字段 (UDP 固定为 17)、UDP 长度字段 (图 5-5 假设用户数据报的长度是 15B)。- -定要记住伪首部只用于计算和验证校验和，其既不向下传送，也不向上递交。

如果 UDP 校验和校验出 UDP 数据报是错误的，可以丢弃，也可以交付给上层，但是需要附上错误报告，即告诉上层这是错误的数据报。

通过伪首部，不仅可以检查源端口号、且的端口号和 UDP 用户数据报的数据部分，还可以检查 IP 数据报的源 IP 地址和且的地址。

图 11: 计算 UDP 校验和的例子

5.3 TCP

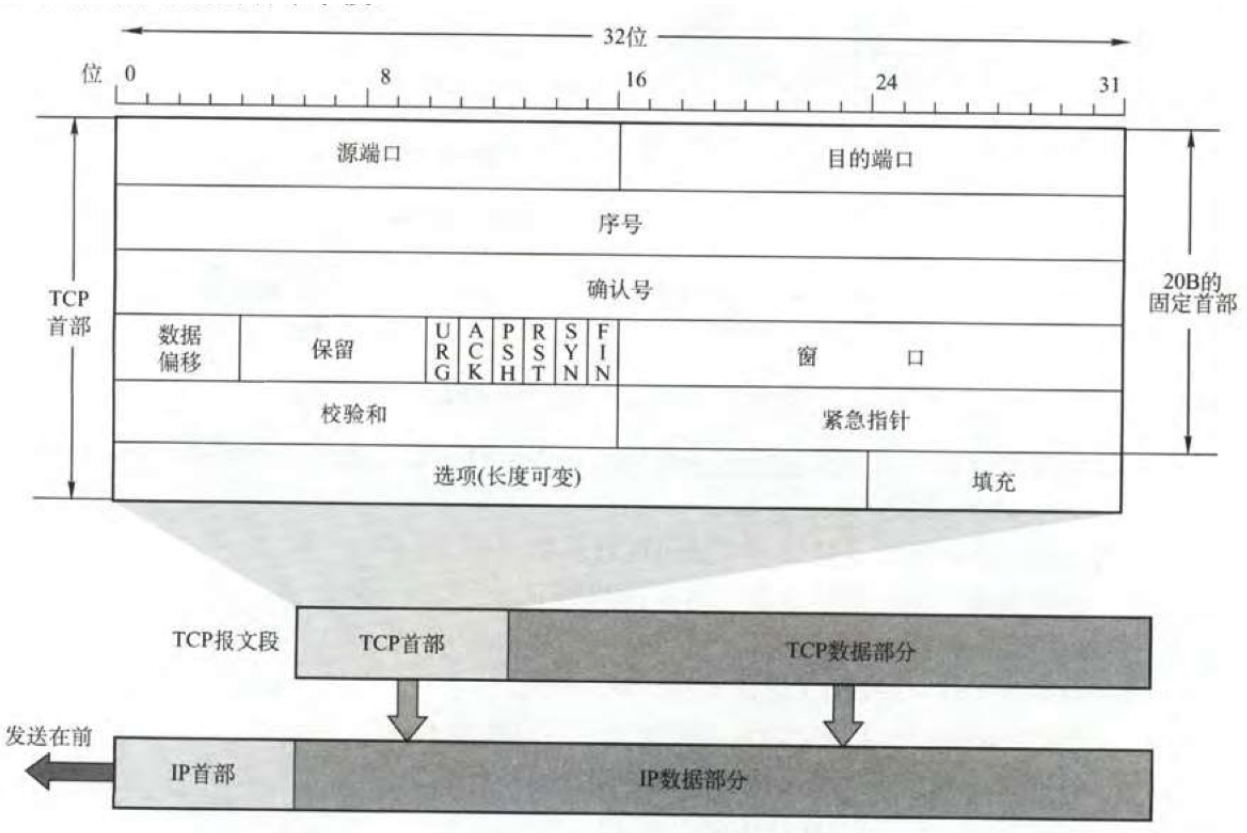

图 12: TCP 报文的首部格式

1) 源端口和目的端口: 各占 2B。与 UDP 一样，TCP 的首部当然也有源端口和目的端口。

2) 序号: 占 4B。尽管从应用层交付下来的是 TCP 报文段, 但是 TCP 是面向字节流的 (就是说 TCP 传送时是按照-一个个字节来传送的)，所以在一个 TCP 连接中传送的字节流需要编号，这样才能保证按序交付。例如，某报文段的序号从 301 开始，而携带的数据共有 100B。这就表明本报文段数据的第一个字节的序号是 301, 最后一个字节的序号是 400。显然，下一一个报文段 (如果还有) 的数据序号应当从 401 开始，即下一个报文段的序号字段应为 401, 这个字段名也称为 “报文段序号”。

3) 确认号: 占 4B。TCP 是含有确认机制的，所以接收端需要给发送端发送确认号，这个确认号只需记住一点: 若确认号等于 N, 则表明到序号 N-1 为止的所有数据都已经正确收到。例如，B 正确收到了 A 发送过来的一个报文段, 其序号字段值是 501, 而数据长度是 200B(序号 501 700)，这表明 B 正确收到了 A 发送的到序号 700 为止的数据。因此，B 期望收到 A 的下一个数据序号是 701, 于是 B 将发送给 A 的确认报文段中的确认号设置为 701。注意，现在的确认号不是 501，也不是 700，而是 701。

4) 数据偏移: 占 4 位。前面已经讲过，这里的数据偏移不是 IP 数据报中分片的那个数据偏移，而是表示首部长度，千万不要混淆。占 4 位可表示 0001 1111- -共 15 种状态，而基本单位是 4B，所以数据偏移确定了首部最长为 60B。

5) 保留字段: 占 6 位。保留为今后使用，但目前应置为 0，该字段可以忽略不计。

6) 紧急 URG: 当 URG=1 时，表明紧急指针字段有效。它告诉系统此报文段中有紧急数据，应尽快传送 (相当于高优先级的数据)。就好像有一-个等待红灯的超长车队，此时有一辆救护车过来，属于紧急事件，救护车就可以不用等红灯了，直接从边上绕过所有的车。但是紧急 URG 需要和紧急指针配套使用，比如说有很多救护车过来，现在就需要一个紧急指针指向最后一-辆救护车，一旦最后一辆救护车过去之后，TCP 就告诉应用程序恢复到正常操作，也就是说数据从第一字节到紧急指针所指字节就是紧急数据。

7) 确认比特 ACK: 只有当 ACK=1 时，确认号字段才有效; 当 ACK=0 时，确认号无效。TCP 规定，一旦连接建立了，所有传送的报文段都必须把 ACK 置 1。

8) 推送比特 PSH: TCP 收到推送比特置 1 的报文段，就尽快地交付给接收应用进程，而不再等到格个缓存都填满后再向上交付。

9) 复位比特 RST: 当 RST=1 时，表明 TCP 连接中出现严重差错 (如由于主机崩溃或其他原因)，必须释放连接，然后再重新建立传输连接。

10) 同步比特 SYN: 同步比特 SYN 置为 1, 表示这是一个连接请求或连接接受报文，后面 TCP 连接会详细讲到。

11) 终止比特 FIN: 释放-一个连接。当 FIN=1 时，表明此报文段的发送端的数据已发送完毕，并要求释放传输连接。

12) 窗口字段: 占 2B。窗口字段用来控制对方发送的数据量，单位为字节 (B)。记住一句话: 窗口字段明确指出了现在允许对方发送的数据量。例如，设确认号是 701，窗口字段是 1000。这就表明，从 701 号开始算起，发送此报文段的一方还有接收 1000B 数据的接收缓存空间。

13) 校验和字段: 占 2B。校验和字段检验的范围包括首部和数据两部分。在计算校验和时，和 UDP 一样，要在 TCP 报文段的前面加上 12B 的伪首部 (只需将 UDP 伪首部的第 4 个字段的 17 改为 6，其他和 UDP 一样)。

14) 紧急指针字段: 占 2B。前面已经讲过紧急指针指出在本报文段中的紧急数据的最后一个字节的序号。

15) 选项字段: 长度可变。TCP 最初只规定了一种选项，即最大报文段长度 MSS。MSS 告诉对方 TCP: “我的缓存所能接收的报文段的数据字段的最大长度是 MSS 字节。”

16) 填充字段: 为了使整个首部长度是 4B 的整数倍。

URG=1, 表示紧急指针指向报文内数据段的某个字节 (数据从第一-字节到指针所指字节就是

紧急数据)，不进入接收缓冲 (前面讲了待上交的数据要先进入接收缓存，然后再交付给应用层。而这里就直接交给上层进程，余下的数据都是要进入接收缓冲的)。一般来说，TCP 是要等到整个缓存都填满了后再向上交付，如 PSH=1，就不用等到整个缓存都填满，直按方付. 旧是这里的交付仍然是从缓冲区中交付的，URG 是不经过缓冲区的，千万记住!

三次握手

第一步: 客户机 A 的 TCP 向服务器 B 发出连接请求报文段，其首部中的同步位 SYN=1(TCP 规定，SYN 报文段不能携带数据，但要消耗-一个序号)，并选择序号 seq=x，表明传送数据时的第一个数据字节的序号是 x。

第二步: 服务器收到了数据报，并从 SYN 位为 1 知道这是一个建立连接的请求。如果同意，则发回确认。B 在确认报文段中应使 SYN=1，ACK=1，其确认号 ack= =x+1，自己选择的序号 seq=y。注意，此时该报文段也不能携带数据 (助记: 因为有 SYN=1, 所以不能带数据)。

第三步: A 收到此报文段后向 B 给出确认，其 ACK=1, 确认号 ack=y+1。A 的 TCP 通知上层应用进程，连接已经建立。B 的 TCP 收到主机 A 的确认后，也通知其上层应用进程，此时 TCP 连接已经建立，ACK 报文可以携带数据 (没有 SYN 字段)，如果不携带数据则不消耗序号。

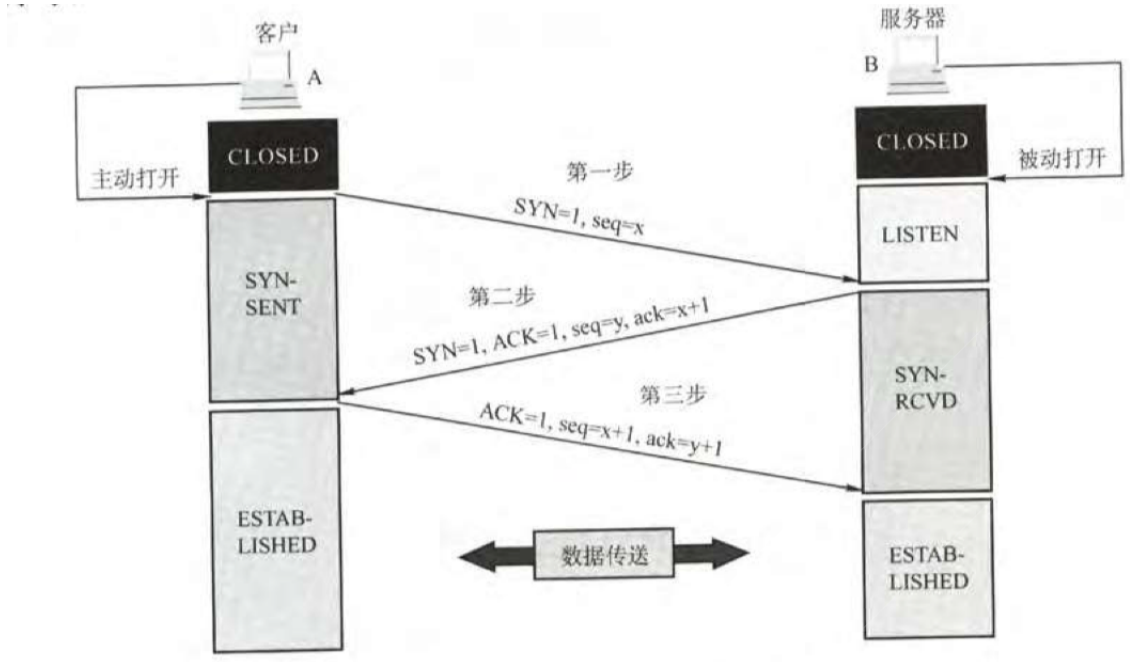

图 13: 三次握手方法建立 TCP 传输连接