第4章

程序员必学的拥塞控制算法

第7章

TCP前沿的研究动向

第 1 章

TCP 入门

确保传输可靠性

互联网是全世界拥有通信功能的设备互相联通所构成的网络。所有设备并非各行其是地随意运作，而是遵循着同一套规则。这套规则称为协议，是全世界通用的标准。

通信是靠多个协议分层运作而实现的。TCP 是其中一个协议，主要承担“确保传输可靠性”的重要职责。

本章将首先概述实现网络通信的各个协议，明确传输层的职责和特征，然后介绍本书的主题之一——TCP 的基本功能。

1.1 通信与协议

OSI参考模型、TCP/IP和RFC

所谓协议，其实是多种多样的。根据层级选择相应的协议，便能按照应用程序的要求实现通信。

本节将概述通信协议的总体情况。

OSI参考模型

设备间的通信方式，其实和人与人之间的交流方式基本一致。举例来说，听不懂方言，双方便无法沟通，但如果使用普通话，双方就可以沟通（图1.1）。只要全世界的交流语言互通，不因国家和地区而不同，那么全世界的人们就可以无障碍交流。

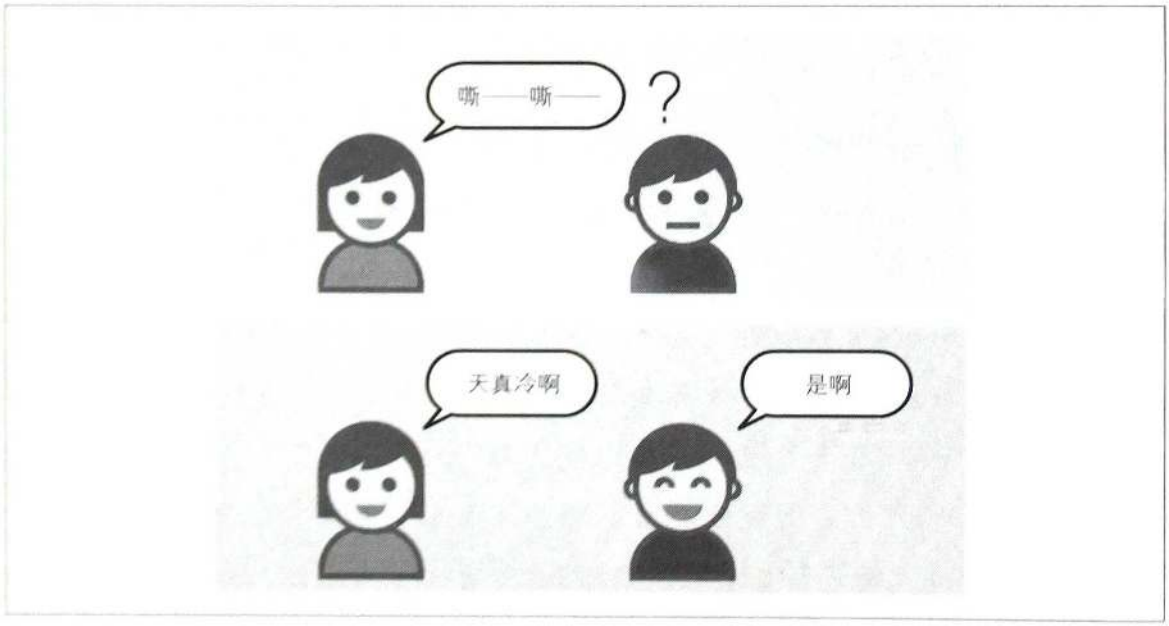

图1.1 使用互通的语言交流

语言互通靠普通话，而通信设备互通，靠的是**OSI参考模型**。OSI是Open System Interconnection（开放式系统互连）的简称。OSI参考模型由国际标准化组织（International Organization for Standardization，ISO）制定，

它支持互通的功能分层设计，能够使不同的设备具备相互通信的能力。

将协议分层后，软件开发者就只需针对具体的层所负责的功能，专注开发其专有逻辑即可。如此一来，不仅降低了实现难度，同时也将责任划分得更为具体。下面，我们先简单介绍一下各层的职能。

——[第 7 层] 应用层

应用层主要定义各个应用程序中使用的通信协议。例如，实现网页浏览的 HTTP（Hypertext Transfer Protocol，超文本传输协议）、实现文件下载的 FTP（File Transfer Protocol，文件传输协议）、实现 IP 地址自动分配的 DHCP（Dynamic Host Configuration Protocol，动态主机配置协议）、实现互联网域名与 IP 地址关系对应的 DNS（Domain Name System，域名系统）、实现网络设备时间同步的 NTP（Network Time Protocol，网络时间协议）、实现电子邮件收发的 SMTP（Simple Mail Transfer Protocol，简单邮件传输协议）和 POP（Post Office Protocol，邮局协议），以及实现远程计算机操作的 Telnet 等。

举例来说，在 HTTP 协议中，客户端计算机的 Web 浏览器为了获取 Web 服务器上的 HTML（HyperText Markup Language，超文本标记语言）文件，会发出请求（GET 请求），而服务器则会返回响应内容，最后客户端完成 HTML 文件、样式表和图像数据等的下载（图 1.2）。

图 1.2 HTTP 通信示例

[第6层]表示层

表示的英文是presentation，它的意思是表达、表现，指的是一种向对方传递信息的方法。字符的编码方式，图像或视频的压缩方式，以及数据的加密方式种类繁多，**表示层**所负责的正是将应用程序中这些特定的数据格式转换为通信设备间可以互相理解的、可在网络上互通的格式。

例如，不同的应用程序使用的字符编码方式各不相同，具体有UTF-8、UTF-16和GBK等。将它们转换为可在网络上互通的数据格式，便可实现编码方式各异的应用程序间的数据通信。

其他的协议包括图像压缩格式JPEG（Joint Photographic Experts Group，联合图像专家组）、视频压缩格式MPEG（Moving Picture Experts Group，动态图像专家组）和音乐文件格式MIDI（Musical Instrument Digital Interface，乐器数字接口）等。

[第5层]会话层

会话的英文是session，通常是指用于管理数据通信从开始到结束整个过程的一个基本单位。

会话层负责管理通信连接。通信连接是由各应用程序在收发数据时发出的请求（request）和响应（response）建立起来的。也就是说，会话层负责为各应用程序建立逻辑通信链路。以HTTP为例，在用户浏览一个Web页面的过程中，从发出获取HTTP文件的请求到HTTP响应的一系列数据收发过程称为一个会话。

如图1.3所示，Web浏览器、电子邮件收发系统和游戏程序的数据通信是分别在不同的会话中被管理的[1]。

1 在OSI参考模型中，虽然会话层定义了若干个协议，但这些协议并非独立运作，而是像HTTP协议一样，作为应用程序的功能之一直接实现在应用程序之中。

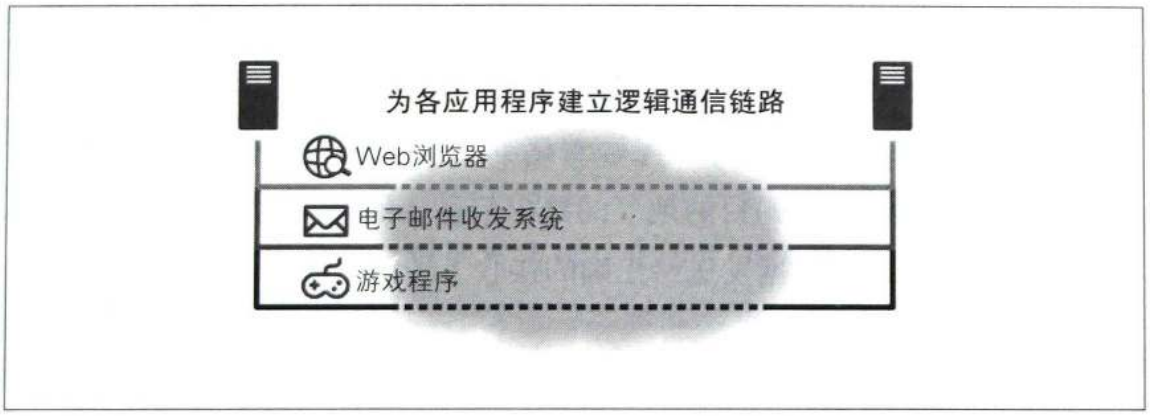

图 1.3 会话

[第 4 层] 传输层

传输层负责建立或断开**连接**（connection），并按照应用程序的要求使用不同的方法转发数据。所谓连接，指的是在会话中，为了进行数据转发而维持的端到端的逻辑通信链路[①]。

传输层主要有两种协议：确保可靠性的 **TCP 协议**和确保实时性的 **UDP 协议**。

各应用程序或会话中会建立一条或多条连接。一条也好，多条也好，都是作为逻辑通信链路执行处理的，而在其下层的数据传输中，可以使用任何介质或线路。

此外，传输层会忽略应用程序转发过来的数据长度，直接将数据分割为适合下层传输介质的长度并转发。分割数据的基本单位在 TCP 中称为报文段（segment），在 UDP 中称为数据报（datagram）。从 1.2 节开始，我们将对此进行详细介绍。

[第 3 层] 网络层

网络层负责管理地址、选择路由和把数据发送到目的地。网络层的代表协议是 IP 和 ICMP（Internet Control Message Protocol，互联网控制消息协议）。支持这两种协议的通信设备主要是路由器（router）和 3 层交换机（Layer 3 switch，也称为 L3 交换机）。

① “连接”和“会话”非常容易混淆，请务必注意两者的区别。

网络由多台路由器、交换机互相连接而成，数据经由这些设备被不断转发。网络上的所有通信设备都会被分配一个IP地址，它的功能相当于居住地址。此外，这些设备都保存着诸如“接下来将数据发送到哪台设备，数据才能到达最终的目的地”的路由信息。网络层的职责就是基于路由信息，将待发送的数据发送给正确的接收方（图1.4）。

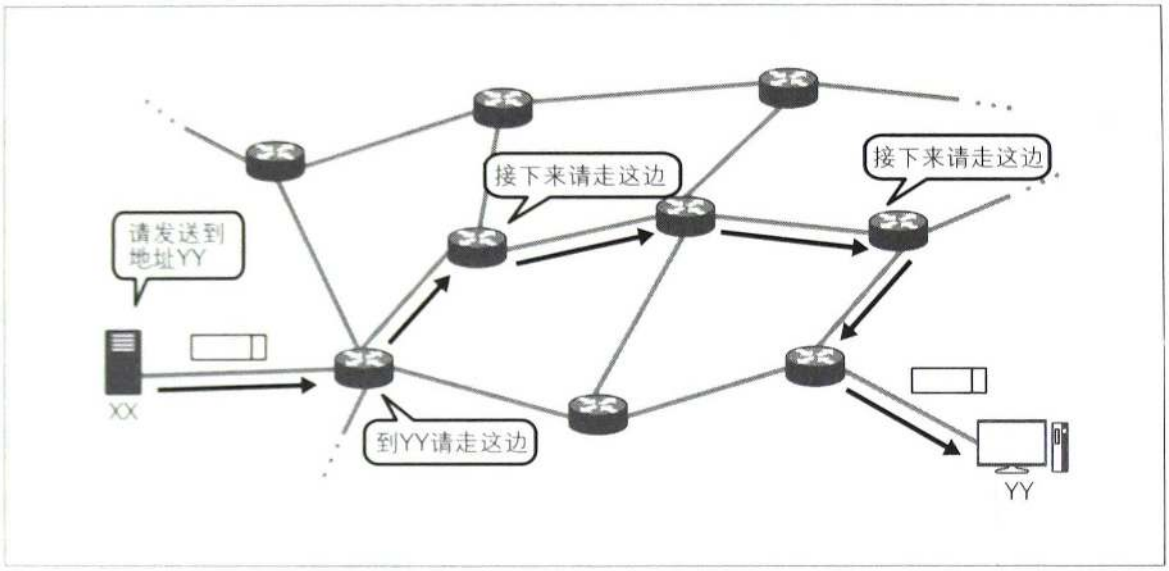

图1.4 数据分发

在发送数据时，基本单位是称为IP**包**（packet）的小数据块。使用IP包发送数据有很多优势。举例来说，当通信链路发生故障导致一部分IP包丢失时，只需重发这些丢失的IP包便可解决问题[1]，其他的IP包则可通过备用链路顺利完成传输。

但是，即使有数据在转发的过程中丢失，网络层也不会进行数据重发，因为这是上一层，即传输层的职责。

另外，为了完成数据分发，各路由器和通信设备需要提前获取网络路由信息。由于手动配置网络路由信息的工作过于烦琐，所以人们设计了应用层（第7层）协议RIP（Routing Information Protocol，路由信息协议）和BGP（Border Gateway Protocol，边界网关协议），以实现路由信息的自动配置。

1 严格来说，这里主要是通过TCP协议重发TCP报文段。

专 栏

IP 地址

IP 协议有 IPv4 和 IPv6 两个版本。两者最主要的差异是包含的地址数量不同。地址数量又称为地址空间（address space）。人们最初使用的 IPv4，其地址由 32 位二进制数表示。随着通信设备逐渐增加，可使用地址逐渐枯竭，所以出现了地址扩展到 128 位的 IPv6 协议。

IP 地址根据用途不同，有多种使用方法。我们先来看格式，IPv4 地址的格式是“xxx.xxx.xxx.xxx”，也就是由 4 个 3 位数加上用于分隔的“.”组成。这是我们常见的 IP 地址。在这种形式的 IP 地址中，高位部分称为网络地址，低位部分称为主机地址。网络地址主要用来标识网络（区域），而主机地址用来标识网络内部的具体设备。

网络地址和主机地址根据所在的网络规模不同，有 5 种划分方法（图 C1.1）。A 类地址的主机地址部分较长，通常适合大规模网络。与之相对，C 类地址主要适合小规模网络。以固定电话号码为例，网络地址和主机地址分别相当于区号和座机号。D 类地址专门用于 IP 多播，而 E 类地址则是预留给未来的。

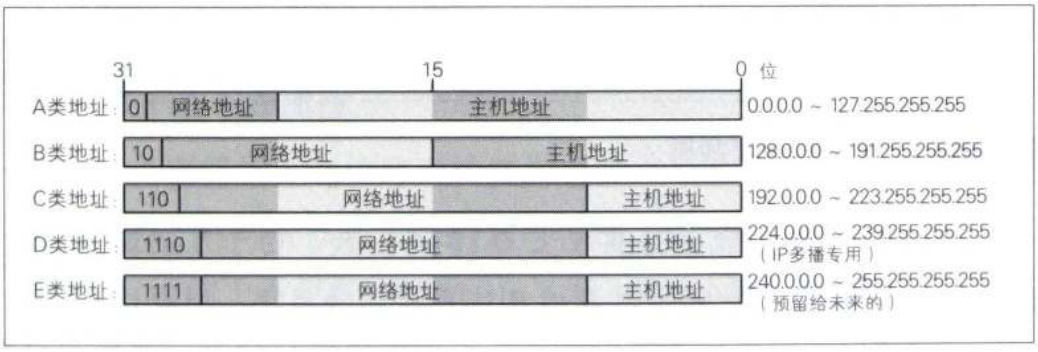

图 C1.1 IP 地址的分类

此外，每一类地址都有“公有地址”和“私有地址”之分。公有地址是分配给所有接入互联网的设备的 IP 地址，而私有地址是分配给办公室或者家庭内部的 LAN（Local Area Network，局域网）之类的，位于局部网络中的设备的 IP 地址。表 C1.1 中记述了几类