

# 第 1 篇 计算机网络基础

---

第 1 章 计算机网络概述

第 2 章 OSI 参考模型与 TCP/IP 模型

第 3 章 局域网基本原理

第 4 章 IP 基本原理

第 5 章 TCP 和 UDP 基本原理

# 第1章 计算机网络概述

计算机网络已经广泛应用在我们的身边，正改变着人们工作和生活方式。

网络给社会带来的革新是深远的。传统各行各业之间信息的分隔局面，正在被信息化所革新，使得行业之间信息的共享，业务平台互通成为可能。另外，计算机软件已不再局限于过去的单机运行，形形色色的网络应用——如办公自动化系统、远程教学、应用于各行各业的管理软件等等，无不与计算机网络发生着紧密的联系。

计算机网络的迅速普及和企业的 IT 化发展导致了社会对网络工程师的大量需求。企业越来越需要大量的专业人才为他们设计、架构、管理并充分发挥计算机互联网络的作用。

对于初学者而言，首先建立对计算机网络的初步而轮廓性的认识是非常必要的。本章将会为你学习后续章节的知识打下良好的基础；对于已经学习过相关知识的学员，通读本章，将能够帮助您对网络的基础知识进行快速的回顾。

## 1.1 本章目标



紫光集团 H3C  
核心企业 | 数字科技领先品牌

### 课程目标

学习完本课程，您应该能够：

- 掌握计算机网络的定义和基本功能
- 了解计算机网络的演进过程
- 掌握计算机网络的类型和衡量计算机网络的性能指标
- 了解计算机网络的协议标准及其标准化组织

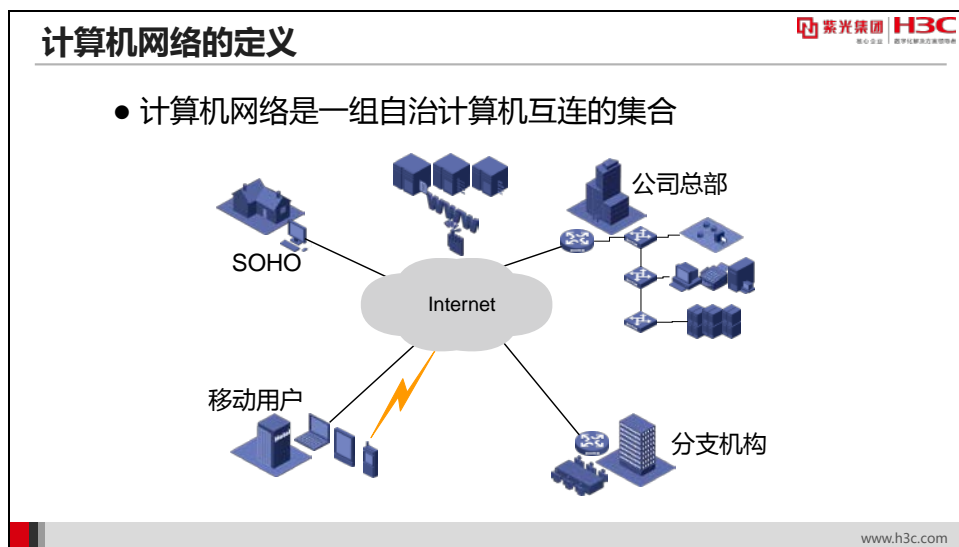


版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

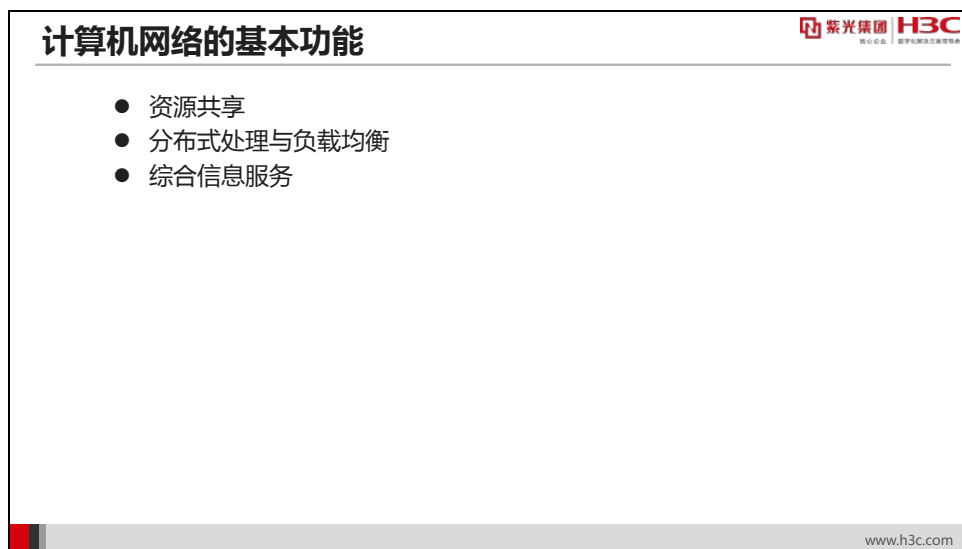
## 1.2 什么是计算机网络

### 1.2.1 计算机网络的定义



计算机网络，顾名思义是由计算机组成的网络系统。根据 IEEE（电子电器工程师协会，Institute of Electrical and Electronics Engineers）高级委员会坦尼鲍姆博士的定义：**计算机网络是一组自治计算机互连的集合**。自治是指每个计算机都有自主权，不受别人控制；互连则是指使用通信介质进行计算机连接，并达到相互通信的目的。这个定义过于专业化。通俗地讲，计算机网络就是把分布在不同地理区域的独立计算机以及专门的外部设备利用通信线路连成一个规模大、功能强的网络系统，从而使众多的计算机可以方便地互相传递信息，共享信息资源。

## 1.2.2 计算机网络的基本功能



归纳说来，计算机网络能为人们带来以下显而易见的益处：

- 资源共享

资源分为软件资源和硬件资源。软件资源包括形式多种多样的数据，如数字信息、消息、声音、图像等；硬件资源包括各种设备，如打印机、FAX、MODEM 等。网络的出现使资源共享变得简单，交流的双方可以跨越时空的障碍，随时随地传递信息、共享资源。

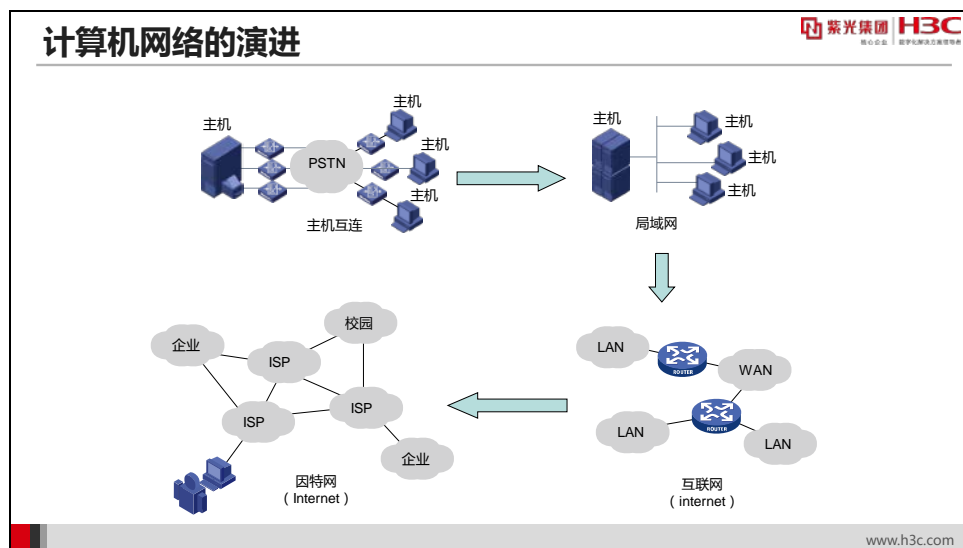
- 分布式处理（distributed processing）与负载均衡（load balancing）

通过计算机网络，海量的处理任务可以分配到分散在全球各地的计算机上。例如，一个大型 ICP（Internet Content Provider）网络访问量相当之大，为了支持更多的用户访问其网站，在全世界多个地方部署了相同内容的 WWW（World Wide Web）服务器；通过一定技术使不同地域的用户看到放置在离他最近的服务器上的相同页面，这样可以实现各服务器的负荷均衡，并使得通信距离缩短。

- 综合信息服务

网络发展的趋势是应用日益多元化，即在一套系统上提供集成的信息服务，如图像、语音、数据等。在多元化发展的趋势下，新形式的网络应用不断涌现，如电子邮件（E-mail）、IP 电话、视频点播（VOD-Video On Demand）、网上交易（E-marketing）、视频会议（Video Conferencing）等。

## 1.3 计算机网络的演进



计算机网络是计算机技术与通信技术两个领域的结合，一直以来它们紧密结合，相互促进，相互影响，共同推进了计算机网络的发展。计算机网络经历了以下几个主要发展阶段：

### ● 主机互连

这种产生于二十世纪 60 年代初期，基于主机（Host）之间的低速串行（Serial）连接的联机系统是计算机网络的最初雏形。在这种早期的网络中，终端借助电话线路访问计算机，由于计算机发送/接收的为数字信号，电话线传输的是模拟信号，这就要求在终端和主机间加入调制解调器（Modem，俗称“猫”），进行数/模间的转换。

在这种联机系统中，计算机是网络的中心，同时也是控制者。这是一种非常原始的计算机网络，它的主要任务是通过远程终端与计算机的连接，提供应用程序执行、远程打印和数据服务等功能。

### ● 局域网

二十世纪 70 年代，随着计算机体积、价格的下降，出现了以个人计算机为主的商业计算模式。商业计算的复杂性要求大量终端设备的资源共享和协同操作，导致了对本地大量计算机设备进行网络化连接的需求，局域网（LAN，Local Area Network）由此产生了。

当今主流局域网技术——以太网（Ethernet）就是在此时期产生的。1973 年，Xerox 公司的 Robert Metcalfe 博士（以太网之父）提出并实现了最初的以太网。后来 DEC、Intel 和 Xerox 合作制定了一个产品标准，该标准最初以这三家公司名称的首字母命名，称作 DIX 以太网。其它流行的 LAN 技术还有 IBM 的令牌环技术等。

### ● 互联网（internet）

由于单一的局域网无法满足对网络的多样性要求，二十世纪 70 年代后期，广域网技术逐渐发展起来，以便将分布在不同地域的局域网互相连接起来。1983 年，ARPANET 采纳 TCP

（传输控制协议，Transmission Control Protocol）和 IP（因特网协议，Internet Protocol）协议作为其主要的协议簇，使大范围地网络互联成为可能。

- 因特网（Internet）


20 世纪 80 年代到 90 年代是网络互联发展时期。在这一时期，ARPANET 网络的规模不断扩大，将全球无数的公司、校园、ISP（Internet Service Provider）和个人用户，最终演变成今天的延伸到全球每一个角落的 Internet。1990 年 ARPANET 正式被 Internet 取代，退出了历史舞台。越来越多的机构、个人参与到 Internet 中来，使得 Internet 获得了高速发展。

## 1.4 计算机网络中的基本概念

### 1.4.1 局域网、城域网和广域网

#### 局域网、城域网和广域网

- LAN ( Local Area Network )
  - 通常指几千米以内的，可以通过某种介质互联的计算机、打印机、modem或其他设备的集合
- MAN ( Metropolitan Area Network )
  - MAN覆盖范围为中等规模，介于局域网和广域网之间，通常是在一个城市内的网络连接（距离为几十公里左右）
- WAN ( Wide Area Network )
  - 分布距离远，它通过各种类型的串行连接以便在更大的地理区域内实现接入



紫光集团 H3C  
通信设备 网络解决方案提供商

www.h3c.com

按计算机网络覆盖范围的大小，可以将计算机网络分为局域网（Local Area Network, LAN）、城域网（Metropolitan Area Network, MAN）、广域网（Wide Area Network, WAN）。

局域网通常指几千米范围以内的，可以通过某种介质互联的计算机、打印机、Modem 或其他设备的集合。局域网连接的是小范围内的计算机，系统覆盖半径从几米到几千米，覆盖范围局限在房间、大楼或园区内。一个局域网通常为一个组织所有，常用于连接公司办公室或企业内的个人电脑和 workstation，以便共享资源（如打印机、数据库等）和交换信息。目前常见局域网的传输速度为 1Gbps、10Gbps 或 40Gbps，传输延迟低（几十微秒），出错率低。局域网与其它网络的区别主要体现在以下几个方面：

- 网络所覆盖的物理范围
- 网络的拓扑结构
- 网络所使用的传输技术

由于局域网分布范围极小，一方面容易管理与配置，另一方面容易构成简洁规整的拓扑结构，加上网络延迟小（一般在几十微秒以下）、数据传输速率高、传输可靠、拓扑结构灵活的优点，使之得到广泛的应用，成为了实现有限区域内信息交换与共享的典型有效的途径。

城域网覆盖范围为中等规模，介于局域网和广域网之间，通常是在一个城市内的网络连接（距离为几十公里左右）。目前城域网建设主要采用 IP 技术和 MPLS 或 SR 技术，宽带 IP 城域网是根据业务发展和竞争的需要而建设的城市范围内（可能包括所辖的县区等）的宽带多媒体通信网络，是宽带骨干网络（如中国电信 IP 骨干网络等）在城市范围内的延伸。城域网作为本地公共信息服务平台的组成部分，负责承载各种多媒体业务，为用户提供各种接入方式，满足

政府部门、企事业单位、个人用户对基于 IP 的各种多媒体业务的需求，因此，宽带 IP 城域网必须是可管理、可扩展的电信运营网络。

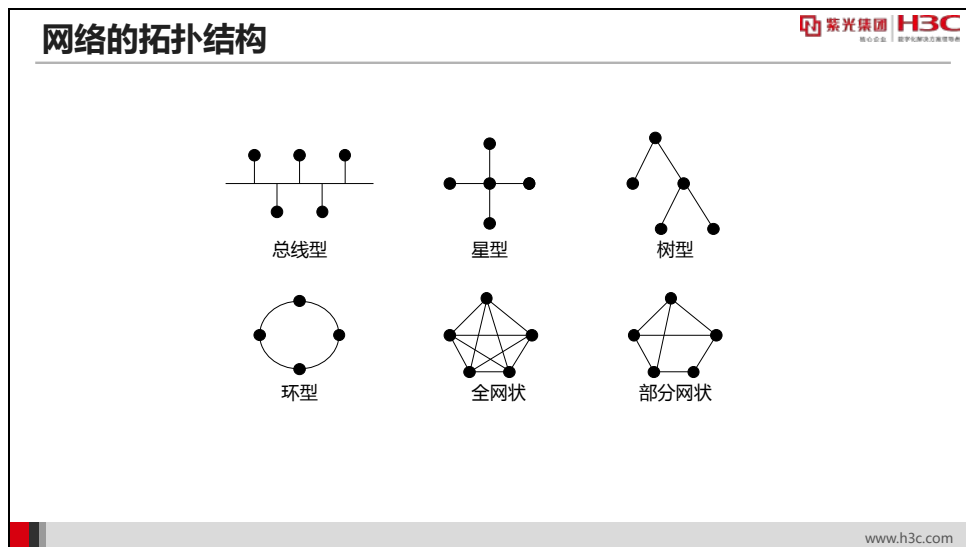
城域网划分为“城域网城域部分”和“城域网接入部分”。城域网城域部分为运营商网络，由运营商统一规划与建设，又可分为城域核心层和城域汇接层。城域核心层主要完成城域网内部信息的高速传送与交换，实现与其它网络的互联互通，而城域汇接层主要完成信息的汇聚与分发。

城域网接入部分可由运营商、企业、建筑商以及物业管理部门建设，其不仅仅提供传统意义上的接入功能，还可能需要向用户提供本地业务。城域网接入部分又分为接入汇接层和用户接入层，接入汇接层完成信息的汇接与分发，实现用户管理，城域网接入部分的业务提供、计费等功能，而用户接入层为用户提供具体的接入手段。

广域网在超过局域网的地理范围内运行，分布距离远，它通过各种类型的专线连接以便在更大的地理区域内实现接入。通常，企业网通过广域网线路接入到当地 ISP。广域网可以提供全部时间或部分时间的连接，允许通过不同类型的广域网接口在不同的速率工作。广域网本身往往不具备规则的拓扑结构。由于速度慢，延迟大，入网站点无法参与网络管理，所以，它要包含复杂的互连设备（如交换机、路由器）处理其中的管理工作，互连设备通过通信线路连接，构成网状结构（通信子网）。其中，入网站点只负责数据的收发工作；广域网中的互连设备负责数据包的路由等重要管理工作。广域网的特点是数据传输慢（相比局域网而言）、延迟比较大（几毫秒）、拓扑结构不灵活，广域网拓扑很难进行归类，一般多采用网状结构，网络连接往往要依赖运营商提供的电信数据网络。



## 1.4.2 网络的拓扑结构



网络拓扑（Network Topology）指的是计算机网络的物理布局。简单地说，就是指将一组设备以什么样的结构连接起来，通常也称为拓扑结构。基本的网络拓扑模型主要有总线型拓扑、环型拓扑、星型拓扑和网状拓扑，绝大部分网络都可以由这几种拓扑独立或混合构成。了解这些拓扑结构是设计网络 and 解决网络疑难问题的前提。

### ● 总线型（Bus）拓扑

总线型拓扑结构是将各个节点的设备用一根总线连接起来，所有的节点间通信都通过统一的总线完成。在早期的局域网中，这是一种应用很广的拓扑结构。其突出的特点是结构简单、成本低、安装使用方便，消耗的电缆长度短、便于维护。但它也具有固有的致命缺点——存在单点故障。总线如果出现故障，整个总线型网络都会瘫痪。由于共享总线带宽，当网络负载过重时，会导致总线型网络性能下降。为了克服这些问题，随后产生了星形的拓扑结构。

### ● 星型（Star）拓扑

星型拓扑结构，是一种以中央节点（如交换机）为中心，把若干个外围节点连接起来的辐射式互连结构，中央节点对各设备间的通信和信息交换进行集中控制和管理。它的主要特点是系统的可靠性较高，当某一线路发生故障时，不会影响网络中的其他主机；扩充或删除设备较容易，将设备直接连接到中央节点即可；中央节点可以方便地控制和管理网络，并及时发现和系统故障。其缺点是需要的连接线缆比总线型拓扑结构多；且一旦中央节点发生故障，网络将不能工作。星形拓扑结构是在当前的局域网中使用较为广泛的一种拓扑结构，它已基本代替了早期局域网采用的总线型拓扑结构。

### ● 环型（Ring）拓扑

环型拓扑结构是将各节点通过一条首尾相连的通信线路连接起来的一个封闭的环型网。每一台设备只能和它的一个或两个相邻节点直接通信，如果需要与其他节点通信，信息必须依次经过两者之间的每一个设备。环型网络可以是单向的，也可以是双向的。单向是指所有的传输

都是同方向的，此时每个设备只能直接与一个邻近节点通信；双向是指数据能在两个方向上进行传输，此时设备可以直接与两个邻近节点直接通信。


环型拓扑的结构简单，系统中各工作站地位相等；建网容易，增加或减少节点时仅需简单的连接操作；能实现数据传送的实时控制，可预知网络的性能。在单环型拓扑中，任何一节点发生故障，就会导致环中的所有节点无法正常通信，在实际应用中一般采用多环结构，这样在单点发生故障时可以形成新的环型，继续正常工作。环型拓扑的另一个缺点是当一个节点要往另一个节点发送数据时，它们之间的所有节点都得参与传输，这样，比起总线拓扑来，更多的时间被花在替别的节点转发数据上。

### ● 网状（Mesh）拓扑

网状（Mesh）拓扑可分为全网状（Full Mesh）和部分网状（Partial Mesh）。全网状拓扑是指参与通信的任意两个节点之间均通过传输线直接相互连接，所以这是一种极端安全可靠的方案。由于不再需要竞争公用线路，通信变得非常简单，任意两台设备可以直接通信，而不用涉及其他设备。然而，对  $N$  个节点构建全网状拓扑需要  $N(N-1)/2$  个连接，这使得在大量节点之间建立全网状拓扑的费用变得极其昂贵。而且，如果两台设备间通信流量很小，那么它们之间的线路利用率就很低，几乎肯定有很多连接得不到充分利用。由于全网状拓扑实现起来费用高、代价大、结构复杂、不易管理和维护，在局域网中很少采用。实际应用中常常采用部分网状拓扑替代全网状拓扑，即在重要节点之间采用全网状拓扑，对相对非重要的节点则省略一些连接。

## 1.4.3 电路交换与分组交换

### 电路交换与分组交换



- 电路交换：基于电话网的电路交换
  - 优点：延迟小、透明传输
  - 缺点：带宽固定，网络资源利用率低，初始连接建立慢
- 分组交换：以分组为单位存储转发
  - 优点：多路复用，网络资源利用率高
  - 缺点：延迟大，实时性差，设备功能复杂

[www.h3c.com](http://www.h3c.com)

电路交换（Circuit Switching）和分组交换（Packet Switching）是通信中的一对重要概念。

### ● 电路交换

交换的概念最早来自于电话系统。电话交换机采用的就是电路交换技术。通信网络中的电路交换与拨打电话的原理类似，当端节点要求发送数据时，交换机就在发送节点和接收节点之

间创建一条独占的数据传输通道。这条通道既可能是一条物理线路，也可能是经过多路复用得到的逻辑通道。这条通道具备固定的带宽，由通信双方独占，一直到通信结束。除非两个节点断开连接，否则传输信道便一直处于服务状态。

电路交换的优点是传输延迟小，由于一旦建立了线路，便不再需要交换，因此保证了较低的延迟；其次一旦线路建立，便不会发生资源的抢占和冲突；电路交换能实现数据的透明传输（即传输通路对用户数据不进行任何修正或解释）、信息传输的吞吐量大。

电路交换的缺点是所占带宽固定，网络资源利用率低。在电路交换系统中，物理线路的带宽是预先分配好的。对于已经预先分配好的线路，即使通信双方没有数据要交换，线路带宽也不能为其他用户所使用，从而造成带宽的浪费。此外，电路交换建立连接所需的时间比较长。电路交换本来是为打电话而设计的。每次需建立连接时，呼叫信号必须经过若干个交换机，得到各交换机的认可，并最终传到被呼叫方。在 PSTN 电话网中，这个过程常常需要 10 秒甚至更长的时间（呼叫市内电话、国内长途和国际长途，需要的时间是不同的）。另外比较重要的一点是，如果使用电路交换技术，网络中每台计算机都必须能建立到所有其他计算机的直接电路连接，而这在大规模网络中几乎是不可能实现的。

由于计算机通信具有频繁、快速、小量、流量峰谷差距大、同时对多点通信等特点，过长的电路建立时间、每个计算机固定电路带宽是不合适的，因此电路交换并不适用于大规模计算机网络中的终端直接通信。

### ● 分组交换

分组交换技术将需要传输的信息划分为具有一定长度的分组（Packet，也称为包），以分组为单位进行存储转发。每个分组都载有接收方地址和发送方地址的标识，便于在网络中寻址；网络中的传送设备则根据这些地址进行分组转发，使信息最终传递到目的节点。

由于采用动态复用的技术来传送各个分组，虽然任意时刻线路总是被某个分组独占，但线路的带宽在统计上得到复用，从而提高了线路的利用率。


分组交换能够保证任何用户都不能长时间独占某传输线路，因而它可以较充分地利用信道带宽，并且可以达到处理并行交互式通信的能力。IP 电话就是使用分组交换技术的一种新型电话，它的通话费远远低于传统电话，原因就在这里。

但是在分组交换中，数据要被分割成分组，而网络设备也需要逐一对分组实行转发，这使得分组交换引入了更大的端到端延迟。由于每个分组都要载有额外的地址信息，因此同样的有效数据实际上需要占用更多的带宽资源。另外由于来自多对通信节点的数据复用同一个信道，突发的数据可能造成信道的拥塞。所有这些使得分组交换网络设备和协议需要具备处理寻址、转发、拥塞等的的能力，这也加大了对分组交换网络设备处理能力和复杂程度的要求。

## 1.5 衡量计算机网络的主要指标

### 衡量计算机网络的主要指标

- 带宽 (bandwidth)
  - 描述在一定时间范围内能够从一个节点传送到另一个节点的数据量
  - 通常以bps为单位
  - 例如目前常见以太网带宽为1Gbps
- 延迟 (delay)
  - 描述网络上数据从一个节点传送到另一个节点所经历的时间



紫光集团 H3C  
网络产品 数字化转型解决方案提供商

www.h3c.com

影响网络性能的因素有很多，传输的距离、使用的线路、传输技术、带宽（bandwidth）、网络设备性能等都会对网络的性能产生影响。带宽和延迟（delay）是衡量网络性能的两个主要指标。

LAN 和 WAN 都使用带宽（bandwidth）来描述在一定时间范围内能够从一个节点传送到另一个节点的数据量。带宽分为模拟带宽和数字带宽，本书所述的带宽指数字带宽。带宽的单位是位每秒（bps, bit per second），代表每秒钟某条链路能发送的数据位数。

目前常见的网络带宽有：

- 目前常用以太网技术的带宽可以为 1000Mbps、10Gbps、40Gbps 等；
- 家庭宽带接入上网带宽可以为 100Mbps、500Mbps、1000Mbps 等；

网络的延迟（delay）又称时延，定义了网络把数据从一个网络节点传送到另一个网络节点所需要的时间。网络延迟主要由传播延迟（propagation delay）、交换延迟（switching delay）、介质访问延迟（access delay）和队列延迟（queuing delay）等组成。总之，网络中产生延迟的因素很多，既受网络设备的影响，也受传输介质、网络协议标准的影响；既受硬件制约，也受软件制约。由于物理规律的限制，延迟是不可能完全消除的。


## 1.6 网络标准化组织



在计算机网络的发展过程中有许多国际标准化组织做出了重大的贡献，他们统一了网络的标准，使各个厂商生产的网络产品可以相互兼容。这些组织主要有：

- **国际标准化组织 (ISO, International Organization for Standardization):** 该组织负责制定大型网络的标准，包括与 Internet 相关的标准。ISO 提出了 OSI 参考模型。OSI 参考模型描述了网络的工作机理，为计算机网络构建了一个易于理解的、清晰的层次模型。
- **电子电器工程师协会 (IEEE, Institute of Electrical and Electronics Engineers):** 主要提供了网络硬件的标准，使各厂商生产的硬件设备能相互连通。IEEE LAN 标准是当今居于主导地位的 LAN 标准。它主要定义了 802.X 协议族，其中 802.3 为以太网标准、802.4 为令牌总线网 (Token Bus) 标准、802.5 为令牌环网 (Token Ring) 标准、802.11 为无线局域网 (WLAN) 标准。
- **美国国家标准局 (ANSI, American National Standards Institute):** 是由公司、政府和其他组织成员组成的自愿组织，光纤分布式数据接口 (FDDI) 即由其定义。
- **国际电信联盟 (ITU, International Telecomm Union):** 定义了作为广域连接的电信网络的标准。
- **Internet 架构委员会 (IAB, Internet Architecture Board):** 下设工程任务委员会 (IETF)、研究任务委员会 (IRTF)、号码分配委员会 (IANA) 等，负责各种 Internet 标准的定义，是目前最具影响力的国际标准化组织。

## 1.7 本章总结



紫光集团 H3C  
核心企业 | 新华三集团之重要成员

### 课程总结

- 计算机网络可以实现资源共享、综合信息服务、负载均衡与分布式处理等基本功能
- 计算机网络的类型可以按照地域、拓扑结构、数据交换的形式及网络组件等不同类型进行分类
- 衡量计算机网络的性能指标有很多种，其中带宽和延迟最为重要

版权所有 2003-2021 新华三技术有限公司.保留一切权利

[www.h3c.com](http://www.h3c.com)

## 第2章 OSI 参考模型与 TCP/IP 模型

在网络发展的早期时代，网络技术的发展变化速度非常快，计算机网络变得越来越复杂，新的协议和应用不断产生，而网络设备大部分都是按厂商自己的标准生产，不能兼容，很难相互间进行通信。

为了解决网络之间的兼容性问题，实现网络设备间的相互通讯，国际标准化组织 ISO 于 1984 年提出了 OSIRM（Open System Interconnection Reference Model，开放系统互连参考模型）。OSI 参考模型很快成为计算机网络通信的基础模型。

由于种种原因，并没有一种完全忠实于 OSI 参考模型的协议族流行开来。相反，源于美国国防部高级研究项目机构（DARPA，Defense Advanced Research Project Agency）六十年代开发的 ARPANET 的 TCP/IP 协议得到了广泛应用，成为 Internet 的事实标准。

### 2.1 本章目标



### 课程目标

学习完本课程，您应该能够：

- 了解OSI参考模型和TCP/IP模型的产生背景
- 理解OSI参考模型和TCP/IP模型的层次结构及相关概念
- 理解OSI参考模型和TCP/IP模型各层的功能



版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

## 2.2 OSI参考模型

### 2.2.1 OSI 参考模型层次结构

### OSI参考模型

- OSI参考模型定义了网络中设备所遵守的层次结构
- 分层结构的优点：
  - 开放的标准化接口
  - 多厂商兼容性
  - 易于理解、学习和更新协议标准
  - 实现模块化工程，降低了开发实现的复杂度
  - 便于故障排除

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

如今，人们可以方便地使用不同厂家的设备构建计算机网络，而不需要过多考虑不同产品之间的兼容性问题。而在 OSI 模型出现（20 世纪 80 年代）之前，实现不同设备间的互通并不容易。这是因为在计算机网络发展的初期阶段，许多研究机构、计算机厂商和公司都推出了自己的网络系统，然而它们之间互不相容，没有兼容性可言。没有一种统一标准存在，就意味着这些不同厂家的网络系统之间无法相互连接。

国际上各大厂商为了在数据通信网络领域占据主导地位，纷纷推出了各自的网络架构体系和标准，例如 IBM 公司的 SNA，NOVELL 的 IPX/SPX 协议，APPLE 公司的 AppleTalk 协议，DEC 公司的 DECNET，以及目前应用最广泛的 TCP/IP 协议等等。同时，各大厂商针对自己的协议生产出了不同的硬件和软件。这些努力无疑促进了网络技术的快速发展和网络设备种类的迅速增加，但由于多种协议的并存，也使网络变得越来越复杂。而且厂商之间的网络设备大多不能兼容，很难进行通信。

为了解决网络之间兼容性的问题，帮助各个厂商生产出可兼容的网络设备，国际标准化组织 ISO（International Organization for Standardization）于 1984 年提出了开放系统互连参考模型（Open System Interconnection Reference Model, OSI/RM），它很快成为计算机网络通信的基础模型。

OSI 模型是对发生在网络设备间的信息传输过程的一种理论化描述，它仅仅是一种理论模型，并没有定义如何通过硬件和软件实现每一层功能，与实际使用的协议（如 TCP/IP 协议）是有一定区别的。虽然 OSI 仅是一种理论化的模型，但它是所有网络学习的基础，因此除了解各层的名称外，更应深入了解它们的功能及各层之间是如何工作的。



OSI 参考模型很重要的一个特性是其分层体系结构。分层设计方法可以将庞大而复杂的问题转化为若干较小且易于处理的子问题。将复杂的网络通信过程分解到各个功能层次，各个层次的设计和测试相对独立，并不依赖于操作系统或其它因素，层次间也无需了解其它层是如何实现的。OSI 七层参考模型具有以下优点：

- 开放的标准化接口：通过规范各个层次之间的标准化接口，使各个厂商可以自由地生产出网络产品，这种开放给网络产业的发展注入了活力。
- 多厂商兼容性：采用统一的标准的层次化模型后，各个设备生产厂商遵循标准进行产品的设计开发，有效地保证了产品间的兼容性。
- 易于理解、学习和更新协议标准：由于各层次之间相对独立，使得讨论、制定和学习协议标准变得比较容易，某一层次协议标准的改变也不会影响其他层次的协议。
- 实现模块化工程，降低了开发实现的复杂度：每个厂商都可以专注于某一个层次或某一模块，独立开发自己的产品，这样的模块化开发降低了单一产品或模块的复杂度，提高了开发效率，降低了开发费用。
- 便于故障排除：一旦发生网络故障，可以比较容易地将故障定位于某一层次，进而快速找出故障根源。

## 2.2.2 OSI 参考模型层次间的关系以及数据封装



OSI 参考模型的每一层都定义了所实现的功能，完成某些特定的通信任务，并只与紧邻的上层和下层进行数据的交换。

物理层涉及到在通信信道（Channel）上传输的原始比特流，它定义了传输数据所需要的机械、电气、功能及规程的特性等，包括电压、电缆线、数据传输速率、接口的定义等。

数据链路层的主要任务是提供对物理层的控制，检测并纠正可能出现的错误，并且进行流量控制。数据链路层与物理地址、网络拓扑、线缆规划、错误校验和流量控制等有关。

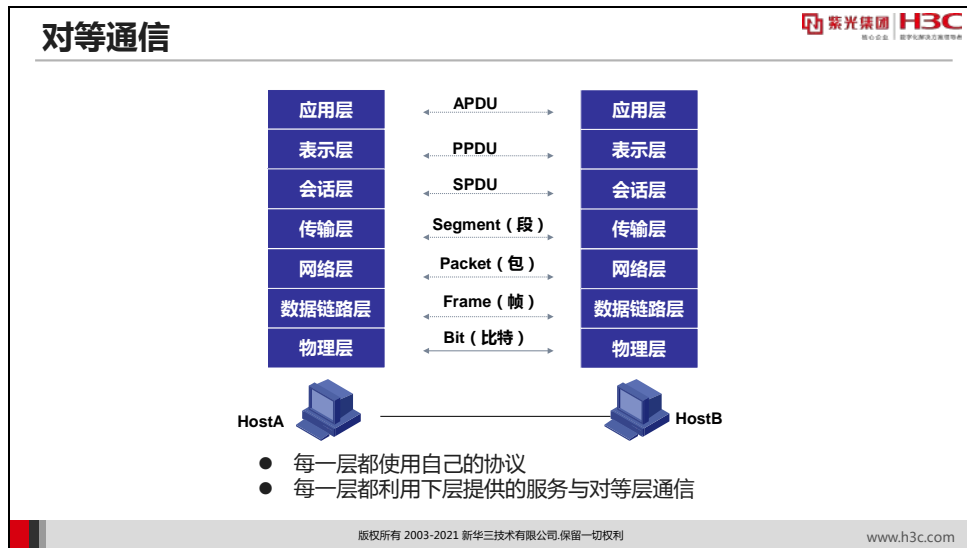
网络层决定传输包的最佳路由，其关键问题是确定数据包从源端到目的端如何选择路由。网络层通过路由选择协议来计算路由。

传输层的基本功能是从会话层接受数据，并且在必要的时候把它分成较小的单元，传递给网络层，并确保到达对方的各段信息正确无误，传输层建立、维护虚电路、进行差错校验和流量控制。

会话层允许不同机器上的用户建立、管理和终止应用程序间的会话关系，在协调不同应用程序之间的通信时要涉及会话层，该层使每个应用程序知道其它应用程序的状态。同时，会话层也提供双工（Duplex）协商、会话同步等。

表示层关注于所传输的信息的语法和语义，它把来自应用层与计算机有关的数据格式处理成与计算机无关的格式，以保证对端设备能够准确无误地理解发送端数据。同时，表示层也负责数据加密等。

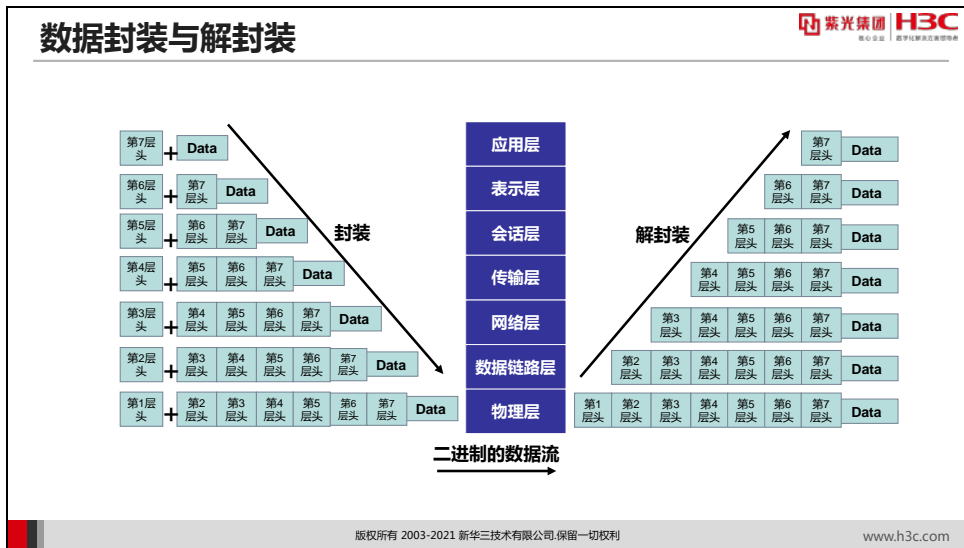
应用层是 OSI 参考模型最接近用户的一层，负责为应用程序提供网络服务。这里的网络服务包括文件传输、文件管理和电子邮件的消息处理等。



应用层数据称为 APDU (Application Protocol Data Unit, 应用层协议数据单元), 表示层数据称为 PPDU (Presentation Protocol Data Unit, 表示层协议数据单元), 会话层数据称为 SPDU (Session Protocol Data Unit, 会话层协议数据单元); 传输层数据称为段 (segment), 网络层数据称为数据包 (packet), 数据链路层数据称为帧 (frame), 物理层数据称为比特 (bit)。

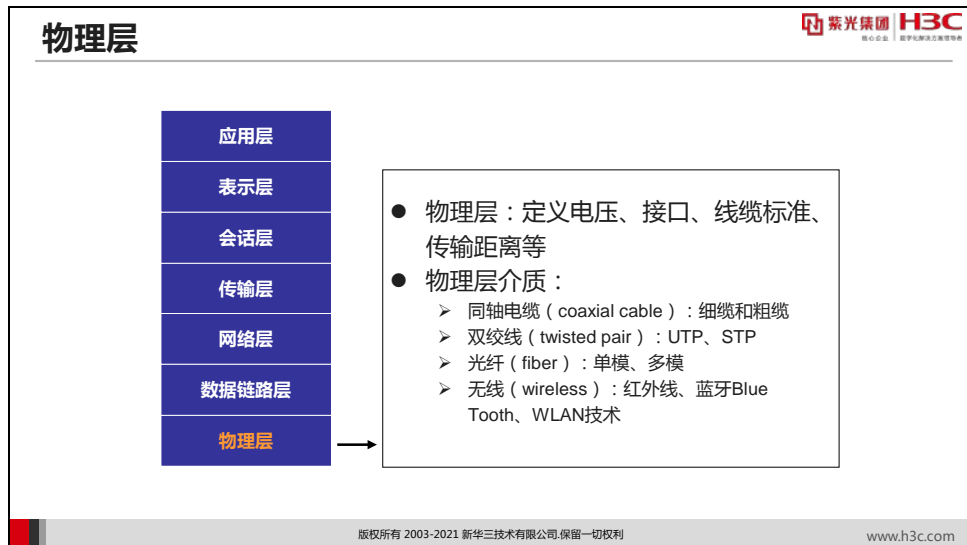
在 OSI 参考模型中, 终端主机的每一层都与另一方的对等层次进行通信, 但这种通信并非直接进行的, 而是通过下一层为其提供的服务来间接与对端的对等层交换数据。下一层通过服务访问点 (SAP, Service Access Point) 为上一层提供服务。例如, 一个终端设备的传输层和另一个终端设备的传输层利用数据段进行通信。传输层的段成为网络层数据包的一部分, 网络层数据包又成为数据链路层帧的一部分, 最后转换成比特流传送到对端物理层, 又依次到达对端数据链路层、网络层、传输层, 实现了对等层之间的通信。

为了保证对等层之间能够准确无误地传递数据, 对等层间应运行相同的网络协议。例如, 应用层的 E-mail 程序不会与对端应用层 Telnet 程序通信, 但可以与对端 E-mail 应用程序通信。



封装（encapsulation）是指网络节点将要传送的数据用特定的协议打包后传送。多数协议是通过在原有数据之前加上封装头（header）来实现封装的，一些协议还要在数据之后加上封装尾（trailer），而原有数据此时便成为载荷（payload）。在发送方，OSI 七层模型的每一层都对上层数据进行封装，以保证数据能够正确无误地到达目的地；而在接收方，每一层又对本层的封装数据进行解封装，并传送给上层，以便数据被上层所理解。

### 2.2.3 物理层



物理层（Physical Layer）是 OSI 参考模型的最低层或称为第一层，其功能是在终端设备间传输比特流。

物理层并不是指物理设备或物理媒介，而是有关物理设备通过物理媒体进行互连的描述和规定。物理层协议定义了通信传输介质的物理特性：

- 机械特性：说明了接口所用接线器的形状和尺寸、引线数目和排列等，例如我们见到的各种规格的电源插头的尺寸都有严格的规定。
- 电气特性：说明在接口电缆的每根线上出现的电压、电流等的范围。
- 功能特性：说明某根线上出现的某一电平的电压表示何种意义。
- 规程特性：说明对不同功能的各种可能事件的出现顺序。

物理层以比特流的方式传送来自数据链路层的数据，而不理会数据的含义或格式。同样，它接收数据后直接传给数据链路层。也就是说，物理层只能看到 0 和 1，它不能理解所处理的比特流的具体意义。

## 典型物理层标准和设备

- 物理层介质
  - 双绞线、同轴电缆、光纤、无线信号等
- 局域网物理层
  - 常见标准：1000Base-T、1000Base-SX/LX、10GBase-LR、10GBase-ER
  - 常见设备：交换机
- 广域网物理层
  - 常见标准：POS、CPOS接口等
  - 常见设备：Modem

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com



紫光集团 H3C  
核心企业 | 新华三集团成员企业

常见的物理层传输介质主要有同轴电缆(coaxial cable)、双绞线(twisted pair)、光纤(fiber)和无线电波等。

双绞线是一种在局域网上最为常用的电缆线。每一对双绞线由一对直径约 1mm 的绝缘铜线缠绕而成，这样可以有效抗干扰。双绞线分为屏蔽双绞线(shielded twisted pair, STP)和非屏蔽双绞线(unshielded twisted pair, UTP)。屏蔽双绞线具有很强的抗电磁干扰和无线电干扰能力，但是价格相对昂贵；非屏蔽双绞线易于安装，价格便宜，但是抗干扰能力相对较弱，传输距离较短。

光纤是另外一种网络传输介质，不受电磁信号的干扰。光纤由玻璃纤维和屏蔽层组成，传输速率高，传输距离远。但是光纤比其它介质更昂贵。

IEEE 802.3 标准定义了以太网物理层常用的线缆标准。其中常用的接口线缆标准有：1000BASE-T、1000BASE-SX/LX、10GBase-LR、10GBase-ER 等。典型的局域网物理层设备是交换机。

广域网物理层规定了以下常用接口：

- POS(Packet Over SONET/SDH)端口支持光纤介质，它是一种高速广域网连接技术。在路由器上插入一块 POS 模块，路由器就可以提供 POS 接口。POS 常用接口速率为 155M，622M，2.4G，10G，40G。POS 端口采用 PPP 或 HDLC 的二层封装来承载 IP 报文。
- CPOS 接口是指支持通道化的 POS 接口(Channelized POS)。它充分利用 SDH 体制的特点，提供对带宽精细划分的能力，可减少组网中对路由器低速物理端口的数量要求，增强路由器的低速端口汇聚能力，并提高路由器的专线接入能力。对于 CPOS 接口，其本身的物理端口不再作为业务口使用，称为控制口，也叫做 Controller。而通道化成的 E1/T1 通道作为同步串口使用

调制解调器(Modem)就是一种常见的广域网物理层设备。

## 2.2.4 数据链路层



数据链路层的目的是负责在某一特定的介质或链路上传递数据。因此数据链路层协议与链路介质有较强的相关性，不同的传输介质需要不同的数据链路层协议给以支持。

数据链路层的主要功能包括：

- 编帧和识别帧：由于物理层只发送和接收比特流，而并不关心这些比特的次序、结构和含义，因此需要链路层将比特编成帧，从一系列比特流中识别帧，并将帧解开传递给网络层。
- 数据链路的建立、维持和释放：当网络中的设备要进行通信时，通信双方有时必须先建立一条数据链路，在建立链路时需要保证安全性，在传输过程中要维持数据链路，而在通信结束后要释放数据链路。
- 传输资源控制：在一些共享介质上，多个终端设备可能同时需要发送数据，此时必须由数据链路层协议对资源的分配加以裁决。
- 流量控制：为了确保正常地收发数据，防止发送数据过快，导致接收方的缓存空间溢出，网络出现拥塞，就必须及时控制发送方发送数据的速率。
- 差错控制：由于比特流传输时可能产生差错，而物理层无法辨别错误，所以数据链路层协议需要以帧为单位实施差错检测。
- 寻址：数据链路层协议应该能够标识介质上的所有节点，并且能寻找到目的节点，以便将数据发送到正确的目的。
- 标识上层数据：数据链路层采用透明传输的方法传送网络层包（**packet**），它对网络层呈现为一条无错的线路。为了在同一链路上支持多种网络层协议，发送方必须在帧的控制信息中标识载荷（即包）所属的网络层协议，这样接收方才能将载荷提交给正确的上层协议来处理。

为了在对网络层协议提供统一的接口的同时对下层的各种介质进行管理控制，局域网的数据链路层又被划分为 LLC(Logic Link Control, 逻辑链路控制)和 MAC(Media Access Control, 介质访问控制)两个子层。



IEEE 的数据链路层标准是当今最为流行的 LAN 标准。这些标准统称为 IEEE802 标准。

- 802.1 描述了基本的局域网需要解决的问题，例如 802.1d 描述了生成树协议。
- 802.2 小组负责 LLC 子层标准的制定。
- 802.3 小组负责 MAC 子层标准的制定，典型技术如 CSMA/CD (Carrier Sense Multiple Access with Collision Detection)。
- 802.4 小组负责令牌总线标准的制定。
- 802.5 小组负责令牌环网络标准的制定，IBM 的令牌环小组和 IEEE802.5 小组建立的标准是基本相同的。

目前，我国应用最为广泛的 LAN 标准是基于 IEEE802.3 的以太网标准。以太网交换机就是一种典型的数据链路层设备。

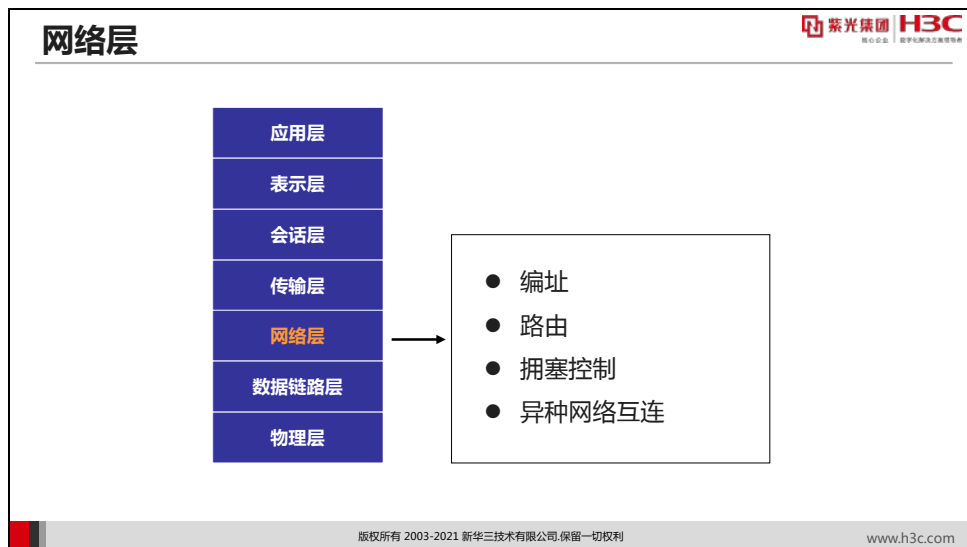
广域网常见的数据链路层标准有 HDLC (High-level Data Link Control, 高级数据链路控制)、PPP (Point-to-Point Protocol, 点到点协议) 等。

HDLC 是 ISO 开发的一种面向位同步的数据链路层协议，它规定了使用帧字符和校验和的同步串行链路的数据封装方法。

PPP 由 RFC (Request For Comment) 1661 描述。PPP 协议由 LCP (Link Control Protocol)、NCP (Network Control Protocol) 以及 PPP 扩展协议族组成。LCP 规定了链路建立、维护以及拆除。PPP 协议支持同步和异步连接，支持多种网络层协议。



## 2.2.5 网络层



在网络层，数据的传送单位是包（**packet**，也称为分组或报文）。网络层的任务就是要选择合适的路径并转发数据包，使数据包能够正确无误地从发送方传递到接收方。

网络层的主要功能包括：


- **编址**：网络层为每个节点分配标识，这就是网络层的地址（**address**）。地址的分配也为从源到目的的路径选择提供了基础。
- **路由选择**：网络层的一个关键作用是要确定从源到目的的数据传递应该如何选择路由，网络层设备在计算路由之后，按照路由信息对数据包进行转发。执行网络层路由选择的设备称为路由器（**router**）。
- **拥塞控制**：如果网络同时传送过多的数据包，可能会产生拥塞，导致数据丢失或延迟，网络层也负责对网络上的拥塞进行控制。
- **异种网络互连**：通信链路和介质类型是多种多样的，每一种链路都有其特殊的通信规定，网络层必须能够工作在多种多样的链路和介质类型上，以便能够跨越多个网段提供通信服务。

网络层处于传输层和数据链路层之间，它负责向传输层提供服务，同时负责将网络地址翻译成对应的物理地址。网络层协议还能协调发送、传输及接收设备的处理能力的不平衡性，如网络层可以对数据进行分段和重组，以使得数据包的长度能够满足该链路的数据链路层协议所支持的最大数据帧长度。

### 注意：

由于早期对英文名词的中文翻译缺乏标准，在通信领域中，**packet** 一词被习惯性地翻译成“包”、“分组”、“报文”等多种形式。本书将根据特定场合不加区分地使用这些中文名称。

## 网络层地址



IP 地址	网络地址	主机地址
	10.	8.2.48

- 网络层地址通常由两部分组成
  - 网络地址
  - 主机地址
- 网络层地址是全局唯一的


版权所有 2003-2021 新华三技术有限公司.保留一切权利
www.h3c.com

网络层地址存在于 OSI 参考模型的第三层，是对通信节点的标识，也是数据在网络中进行转发的依据。不同的网络层协议具有不同的地址格式。其中目前应用最广泛的 IP 地址由四个字节组成，通常用点分十进制数字表示。

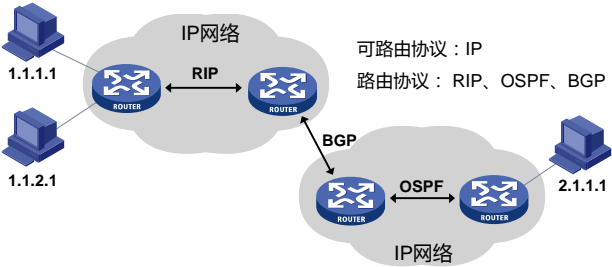
网络层地址通常具有层次化结构，以便将一个巨大的网络区分成若干小块，以便寻址和管理。一种常见的方法是将网络层地址分为“网络地址”和“主机地址”，这样在转发数据包时就可以先将其发送到网络地址所标识的网络，再由所在网络上的网关将其发给主机地址所标识的目的主机。

网络层地址通常是由管理员从逻辑上分配的，因此也称为逻辑地址。为了唯一地标识通信节点，任何一个网络层地址在网络中应该是惟一的。

## 路由协议与可路由协议



- 可路由协议（routed protocol）定义数据包内各个字段的格式和用途，对数据进行网络层封装
- 路由协议（routing protocol）在路由器之间传递信息，计算路由并形成路由表，为可路由协议选择路径



版权所有 2003-2021 新华三技术有限公司.保留一切权利
www.h3c.com

可路由协议（**routed protocol**）是定义数据包内各个字段的格式和用途的网络层封装协议，该网络层协议允许将数据包从一个网络设备转发到另外一个网络设备。目前最常用的可路由协议有 **TCP/IP** 协议族中的 **IP** 协议。

路由协议（**routing protocol**）运行于路由器上，在路由器之间传递信息，计算用于转发的路由并形成路由表（**routing table**），以便为可路由协议提供路由选择服务。路由协议使路由信息能够在相邻路由器之间传递，确保所有路由器了解到达各个目的的路径。

对于一种可路由协议可以设计出多种路由协议为其服务。例如对于 **IP** 协议而言，其常见的路由协议有 **RIP**（**Routing Information Protocol**，路由信息协议）协议、**OSPF**（**Open Shortest Path First**，开放式最短路径优先）、**IS-IS**（**Intermediate System to Intermediate System**）等等。

紫光集团 H3C  
核心企业 数字化转型服务商

## 面向连接和无连接的服务

- 面向连接的服务
  - 通信之前先建立连接，通信完成后断开连接
  - 有序传递
  - 应答确认
  - 差错重传
  - 适合于对可靠性要求高的应用
- 无连接的服务
  - 尽力而为的服务
  - 无需建立连接
  - 无序列号机制，无确认机制，无重传机制
  - 适合于对延迟敏感的应用

版权所有 2003-2021 新华三技术有限公司.保留一切权利 [www.h3c.com](http://www.h3c.com)

在计算机通信中，面向连接的服务（**Connect-oriented Service**）和无连接服务（**Connectionless Service**）是一对重要的概念。

使用面向连接的服务进行通信时，两个实体在通信前首先要建立连接，而在通信完成后释放连接。当被叫用户拒绝连接时，连接宣告失败。

在建立连接阶段，有关的服务原语以及协议数据单元中，必须给出源主机和目的主机的地址，建立虚链路连接；在数据传输阶段，可以使用一个连接标识符来表示上述这种连接关系。

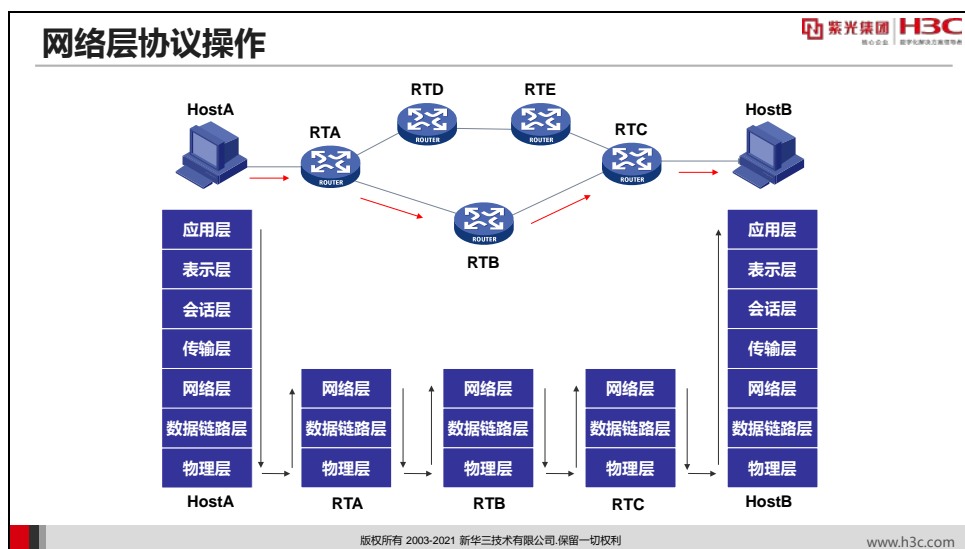
通常面向连接的服务提供可靠的报文序列服务。接收方确认收到的每一份报文，使发送方确信它发送的报文已经到达目的地。确认过程增加了额外的开销和延迟，但如果报文丢失，发送方可以重新发送。在建立连接之后，每个用户可以发送可变长度（在某一限度之内）的报文，这些报文按顺序发送给远端的实体。在正常情况下，当两个报文发往同一目的地时，先发的先收到，但是先发的报文在途中有可能被延误，造成后发的报文反而先收到。接收方利用序列号判断接收的报文是否乱序，并对其按正确的顺序进行排列。面向连接的服务比较适用于在一定时间内向同一个目的地发送很多报文的情况，对于短报文数据的发送而言，面向连接的服务显得开销过大。

在无连接服务中，两个实体之间的通信不需要先建立好一个连接，因此其下层的有关资源不需要事先进行预定保留，这些资源是在数据传输时动态地进行分配的。无连接服务是以邮政系统为模型的，每个报文（信件）带有完整的目的地址，并且每一个报文都独立于其它报文，经由系统选定的路线传递。无连接服务提供尽力而为（**best-effort**）服务，即网络以当前拥有的资源尽力转发报文，但并不保证确切的服务质量。

无连接服务的特征是它不需要通信的两个实体同时处于激活状态，而只需要正在工作的实体处于激活状态。它的优点是灵活方便和比较迅速，但无连接服务不能防止报文的丢失、重复或失序。因此它比较适合传送少量的零星的报文。

并不是所有的应用程序都需要连接。对于某些应用而言，百分之百的可靠性没有必要；对另一些应用而言，其上层应用已经实现了可靠应答机制，所以其本身也不必再确保可靠性。

OSI 参考模型的网络层协议通常提供无连接的服务，不保证数据包的有序可靠传输。数据可靠传输功能通常在传输层实现。



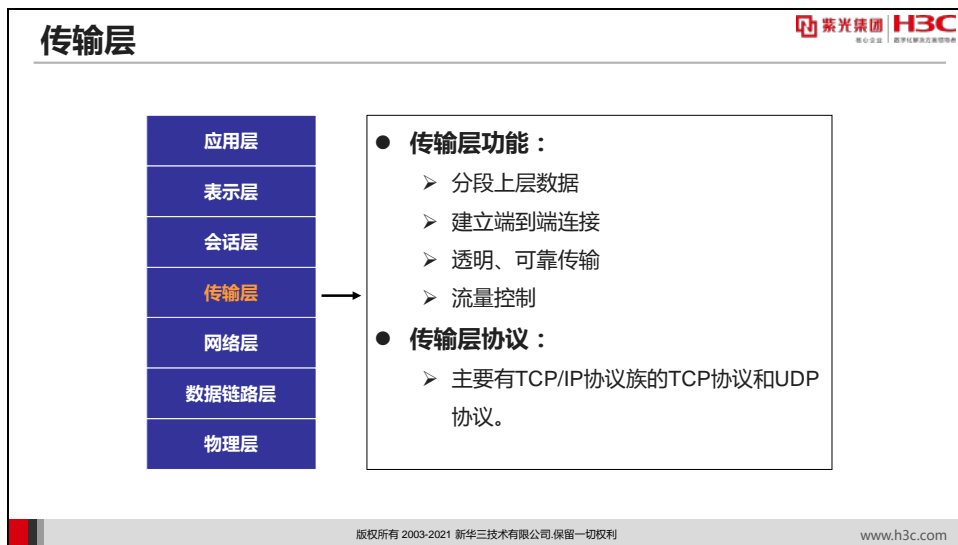
上图演示了数据从主机到服务器的发送过程。

当主机 HostA 上的应用程序需要发送数据到位于另一个网络的 HostB 时，首先将应用层信息转化为能够在网络中传播的数据；随后，在表示层给数据加上表示层报头，协商数据格式，是否加密，转化成对端能够理解的数据格式；然后，数据在会话层又加上会话层报头；以此类推，传输层加上传输层报头成为段（**segment**），网络层将段封装成包（**packet**），数据链路层加上数据链路层头封装为帧（**frame**），最终在物理层转换为比特流。HostA 将比特流发送给网络中距自己最近的网关（**gateway**）——路由器 RTA。

RTA 接收到比特流后，辨认出数据帧并检查该帧，确定被携带的网络层数据类型，然后去掉链路层帧头，得到网络层包。网络层路由转发进程检查包头以决定目的地址所在网段，然后通过查找路由转发信息获取相应输出接口及下一跳的路由器 RTB。输出接口的链路层为该包加上链路层帧头，封装成数据帧并发送到 RTB。

在随后的转发过程中，包在每一跳路由器都经历这一过程，直至包到达路由器 RTC。RTC 在查找路由转发信息时发现目的主机 HostB 与自己处于同一链路上，随即将包封装成目地网络的链路层数据帧，发送给相应的目的主机。目的主机 HostB 接收到该包后，由下而上经过各层的处理，最终送达相应的应用程序。

## 2.2.6 传输层



传输层（Transport Layer）的功能是为会话层提供无差错的传送链路，保证两台设备间传递信息的正确无误。传输层传送的数据单位是段（segment）。

传输层从会话层接收数据，并传递给网络层，如果会话层数据过大，传输层将其切割成较小的数据单元——段进行传送。

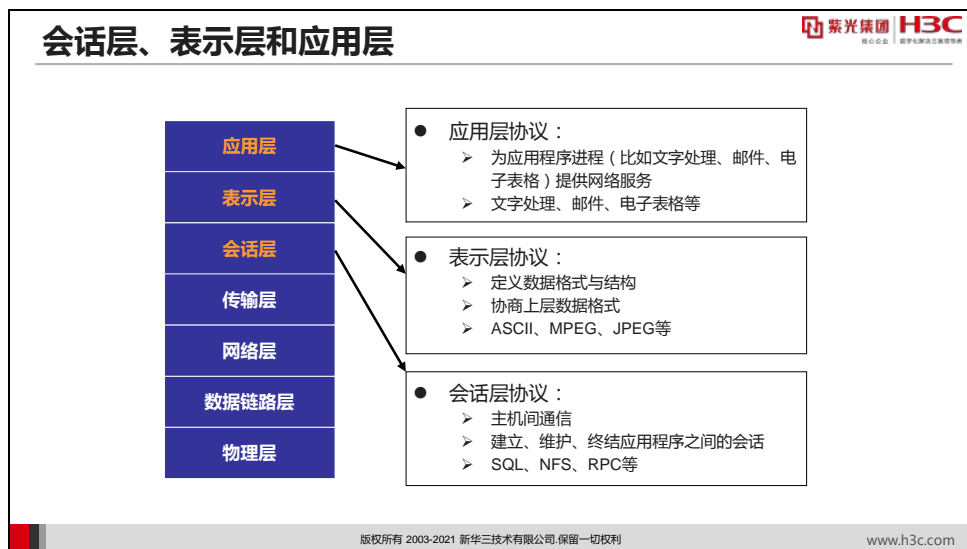
传输层负责创建端到端的通信连接。通过这一层，通信双方主机上的应用程序之间通过对方的地址信息直接进行对话，而不用考虑其间的网络上有多少个中间结点。

传输层既可以为每个会话层请求建立一个单独的连接；也可以根据连接的使用情况为多个会话层请求建立一个单独的连接，这称为多路复用（Multiplexing）。但不论如何，这种传输层服务对会话层都是透明的。

传输层的一个重要工作是差错校验和重传。包在网络传输中可能出现错误，也可能出现乱序、丢失等情况，传输层必须能检测并更正这些错误。一个数据流中的包在网络中传递时如果通过不同的路径到达目的，就可能造成到达顺序的改变。接收方的传输层应该可以识别出包的顺序，并且在将这些包的内容传递给会话层之前将它们恢复成发送时的顺序。接收方传输层不仅要和数据包重新排序，还需验证所有的包是否都已被收到。如果出现错误和丢失，接收方必须请求对方重新传送丢失的包。

为了避免发送速度超出网络或接收方的处理能力，传输层还负责执行流量控制（flow control），在资源不足时降低流量，而在资源充足时提高流量。

### 2.2.7 会话层、表示层和应用层



会话层（**Session Layer**）是利用传输层提供的端到端服务，向表示层或会话用户提供会话服务。就像它的名字一样，会话层建立会话关系，并保持会话过的畅通，决定通信是否被中断以及下次通信从何处重新开始发送。例如，某个用户登录到一个远程系统，并与之交换信息。会话层管理这一进程，控制哪一方有权发送信息，哪一方必须接收信息，这其实是一种同步机制。

会话层也处理差错恢复。例如，若一个用户正在网络上发送一个大文件的内容，而网络忽然发生故障，当网络恢复工作时，用户是否必须从该文件的起始处开始重传呢？回答是否定的，因为会话层允许用户在一个长的信息流中插入检查点，只需将最后一个检查点以后丢弃的数据重传。

如果传输在低层偶尔中断，会话层将努力重新建立通信。例如当用户通过拨号向 **ISP**（因特网服务提供商）请求连接到因特网时，**ISP** 服务器上的会话层向用户的 **PC** 客户机上的会话层进行协商连接。若用户的电话线偶然从墙上插孔脱落，终端机上的会话层将检测到连接中断并重新发起连接。

表示层（**Presentation Layer**）负责将应用层的信息“表示”成一种格式，让对端设备能够正确识别，它主要关注传输信息的语义和语法。在表示层，数据将按照某种一致同意的方法对数据进行编码，以便使用相同表示层协议的计算机能互相识别数据。例如，一幅图像可以表示为 **JPEG** 格式，也可以表示为 **BMP** 格式，如果对方程序不识别本方的表示方法，就无法正确显示这副图片。

表示层还负责数据的加密和压缩。加密（**encryption**）是对数据编码进行一定的转换，让未授权的用户不能截取或阅读的过程。如有人未授权时就截取了数据，看到的将是加过密的数据。压缩（**compression**）是指在保持数据原意的基础上减少信息的比特数。如果传输很昂贵的话，压缩将显著地降低费用，并提高单位时间发送的信息量。

应用层（**Application Layer**）是 OSI 的最高层，它直接与用户和应用程序打交道，负责对软件提供接口以使程序能使用网络服务。这里的网络服务包括文件传输、文件管理、电子邮件的消息处理等。必须强调的是应用层并不等同于一个应用程序。例如，在网络上发送电子邮件，你的请求就是通过应用层传输到网络的。

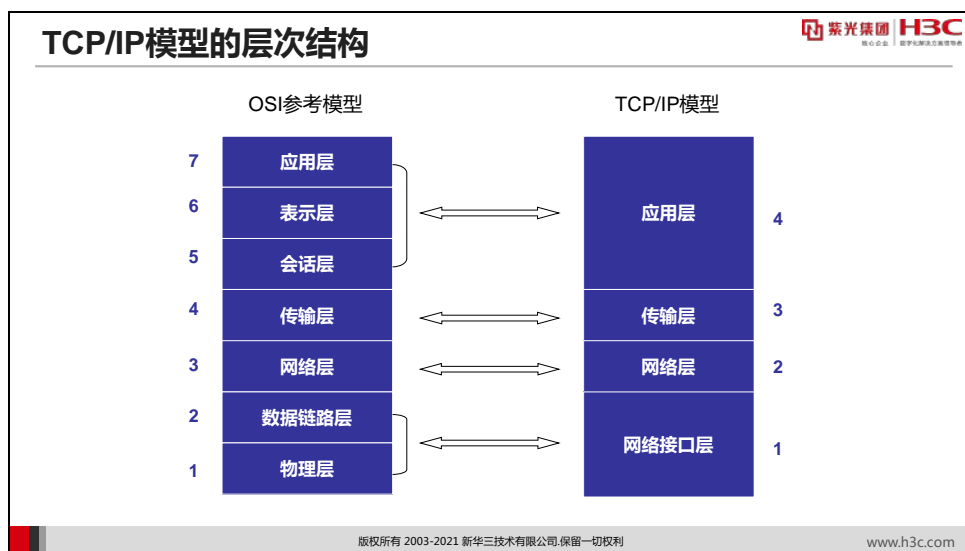


## 2.3 TCP/IP模型

OSI 参考模型的诞生为清晰地理解互联网络、开发网络产品和网络设计等带来了极大的方便。但是 OSI 过于复杂，难以完全实现；OSI 各层功能具有一定的重复性，效率较低；再加上 OSI 参考模型提出时，TCP/IP 协议已逐渐占据主导地位，因此 OSI 参考模型并没有流行开来，也从来没有存在一种完全遵守 OSI 参考模型的协议族。

TCP/IP 起源于 60 年代末美国政府资助的一个分组交换网络研究项目，到 90 年代已发展成为计算机之间最常用的网络协议。它是一个真正的开放系统，因为协议族的定义及其多种实现可以免费或花很少的钱获得。它已成为“全球互联网”或“因特网”（Internet）的基础协议族。

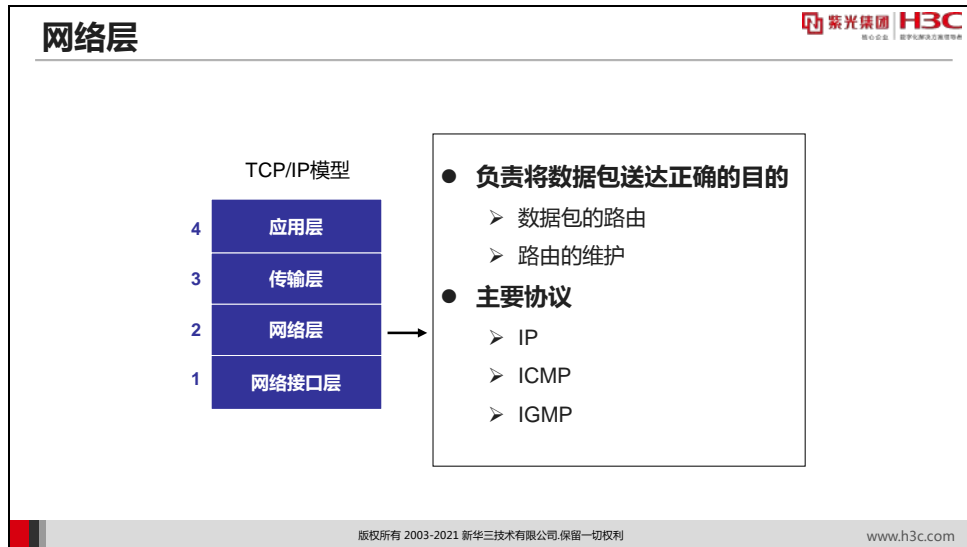
### 2.3.1 TCP/IP 模型的层次结构



与 OSI 参考模型一样，TCP/IP（Transfer Control Protocol / Internet Protocol，传输控制协议/网际协议）也采用层次化结构，每一层负责不同的通信功能。但是 TCP/IP 协议简化了层次设计，只分为 4 层——应用层、传输层、网络层和网络接口层。



## 2.3.2 网络层

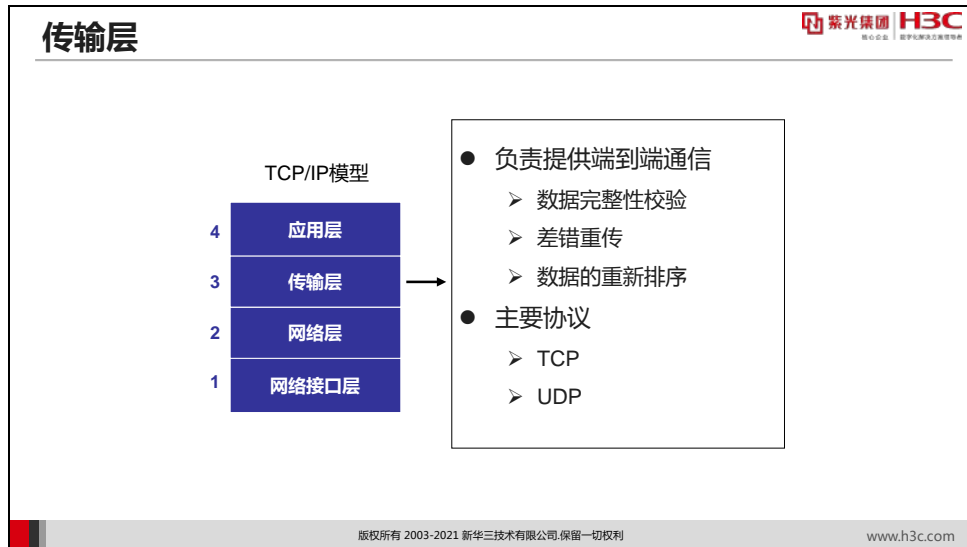


网络层是 TCP/IP 体系的关键部分。它的主要功能是使主机能够将信息发往任何网络并传送到正确的目标。

基于这些要求，网络层定义了包格式及其协议——IP（Internet Protocol，互联网协议）。网络层使用 IP 地址（IP address）标识网络节点；使用路由协议（routing protocol）生成路由信息，并且根据这些路由信息实现包的转发，使包能够准确地传送到目的地；使用 ICMP、IGMP 这样的协议协助管理网络。TCP/IP 网络层在功能上与 OSI 网络层极为相似。

ICMP（Internet Control Message Protocol，互联网控制消息协议）通常也被当作一个网络层协议。ICMP 通过一套预定义的消息在互联网上传递 IP 协议的相关信息，从而对 IP 网络提供管理控制功能。ICMP 的一个典型应用是探测 IP 网络的可达性。

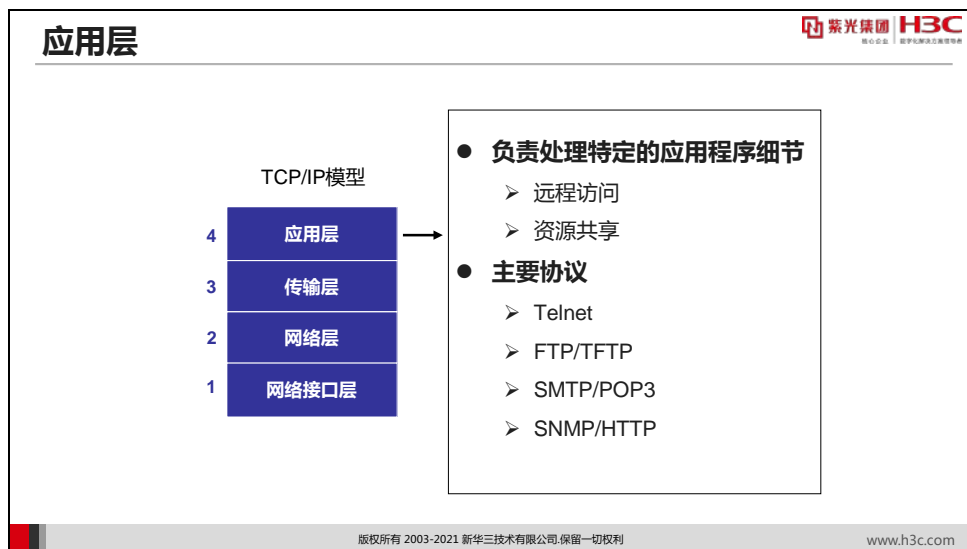
### 2.3.3 传输层



传输层主要为两台主机上的应用程序提供端到端的连接，使源、目的端主机上的对等实体可以进行会话。

在 TCP/IP 协议族的传输层协议主要包括 TCP（Transmission Control Protocol）和 UDP（User Datagram Protocol）。其中 TCP 是面向连接的，可以保证通信两端的可靠传递，支持乱序恢复、差错重传和流量控制。而 UDP 是无连接的，它提供非可靠性数据传输，数据传输的可靠性由应用层保证。

### 2.3.4 应用层



TCP/IP 模型没有单独的会话层和表示层，其功能融合在 TCP/IP 应用层中。应用层它直接与用户和应用程序打交道，负责对软件提供接口以使程序能使用网络服务。这里的网络服务包括文件传输、文件管理、电子邮件的消息处理等。典型的应用层协议包括 Telnet、FTP、SMTP、SNMP 等。

Telnet（TELEcommunications NETwork）的名字具有双重含义，既指这种应用也指协议自身。Telnet 给用户提供了一种通过连网的终端登录远程服务器的方式。

FTP（File Transfer Protocol，文件传输协议）是用于文件传输的 Internet 标准。FTP 支持文本文件（例如 ASCII、二进制等等）和面向字节流的文件结构。FTP 使用传输层协议 TCP 在支持 FTP 的终端系统间执行文件传输，因此，FTP 被认为提供了可靠的面向连接的文件传输能力，适合于远距离、可靠性较差的线路上的文件传输。

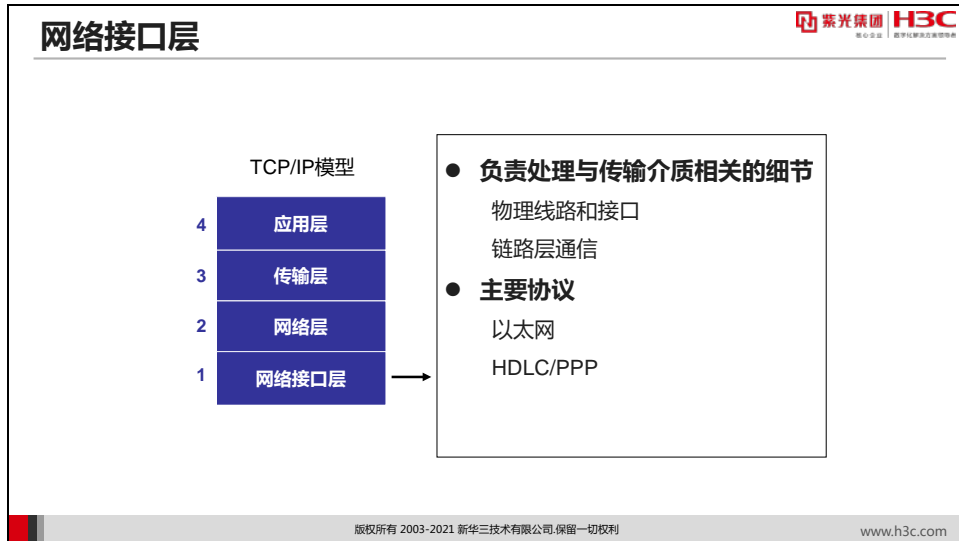
TFTP（Trivial File Transfer Protocol，简单文件传输协议）也用于文件传输，但 TFTP 使用 UDP 提供服务，被认为是不可靠的、无连接的。TFTP 通常用于可靠的局域网内部的文件传输。

SMTP（Simple Mail Transfer Protocol，简单邮件传输协议）支持文本邮件的 Internet 传输。所有的操作系统具有使用 SMTP 收发电子邮件的客户端程序，绝大多数 Internet 服务提供者使用 SMTP 作为其输出邮件服务的协议。SMTP 被设计成在各种网络环境下进行电子邮件信息的传输，实际上，SMTP 真正关心的不是邮件如何被传送，而只关心邮件顺利到达目的地。SMTP 具有健壮的邮件处理特性，这种特性允许邮件依据一定标准自动路由。SMTP 具有当邮件地址不存在时立即通知用户的能力，并且具有把在一定时间内不可传输的邮件返回发送方的特点。

SNMP（Simple Network Management Protocol，简单网络管理协议）负责网络设备监控和维护，支持安全管理、性能管理等。

HTTP（Hypertext Transfer Protocol，超文本传输协议）是 WWW（World Wide Web，万维网）的基础，Internet 上的网页主要通过 HTTP 进行传输。

### 2.3.5 网络接口层




TCP/IP 本身对网络层之下并没有严格的描述。但是 TCP/IP 主机必须使用某种下层协议连接到网络，以便进行通信。而且，TCP/IP 必须能运行在多种下层协议上，以便实现端到端、与链路无关的网络通信。TCP/IP 的网络接口层正是负责处理与传输介质相关的细节，为上层提供一致的网络接口。因此，TCP/IP 模型的网络接口层大体对应于 OSI 模型的数据链路层和物理层，通常包括计算机和网络设备的接口驱动程序和网络接口卡等。

TCP/IP 可以基于大部分局域网或广域网技术运行，这些协议便可以划分到网络接口层中。

典型的网络接口层技术包括常见的以太网局域网技术，用于串行连接的 HDLC（High-level Data Link Control，高级数据链路控制）和 PPP（Point-to-Point Protocol，点到点协议）等技术。

## 2.4 本章总结



紫光集团 H3C  
核心企业 | 新华三集团及新华三集团

### 课程总结

- OSI参考模型和TCP/IP的出现，为清晰地理解互联网络、开发网络产品和网络设计等带来了极大的方便，推动了计算机网络的飞速发展
- OSI参考模型分为七层结构，而TCP/IP模型分为四层结构

版权所有 2003-2021 新华三技术有限公司.保留一切权利

[www.h3c.com](http://www.h3c.com)

## 第3章 局域网基本原理

局域网（Local Area Network，LAN）覆盖的范围较小，通常是处于同一楼层、同一建筑方圆几千米以内的专用网络。本章介绍了几种常见局域网类型，并重点介绍以太网的介质、协商、仲裁机制等主要的工作原理。

### 3.1 本章目标



## 课程目标

学习完本课程，您应该能够：

- 了解局域网类型
- 掌握主要以太网类型及其主要特性
- 了解以太网中传输介质相关知识
- 了解WLAN技术基本原理

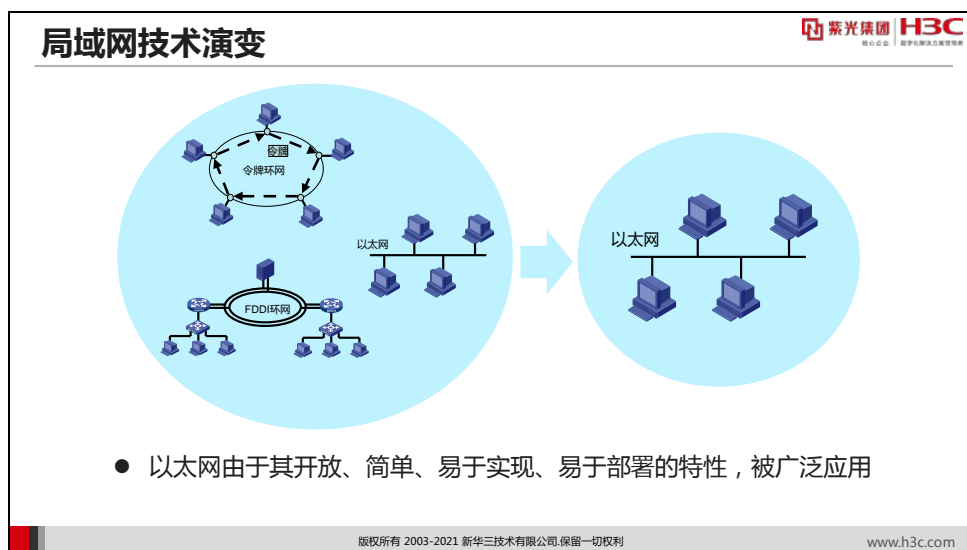


版权所有 2003-2021 新华三技术有限公司,保留一切权利

www.h3c.com

## 3.2 局域网概述

### 3.2.1 局域网技术演变



局域网最主要的功能是在一个较小的物理范围内为计算机提高资源共享和通信服务。局域网内的大量资源共享需求决定了局域网应该是速度较高的，而局域网范围内的众多计算机数量决定了局域网应该是多路访问（Multiple-access）的。

早期常见的局域网技术包括以太网（Ethernet）、令牌环（Token Ring）、FDDI（Fiber Distributed Data Interface，光纤分布式数据接口）等。

令牌环（Token Ring）最早由 IBM 公司设计开发，IEEE 802.5 标准就是在 IBM 公司的 Token Ring 协议的基础上发展和形成的。

在令牌环网中，节点通过环接口连接成环形拓扑。一个节点要想发送数据，首先必须获取令牌。令牌是一种特殊的 MAC 控制帧，令牌环帧中有一位标志令牌的“忙/闲”。令牌总是沿着环单向逐站传送，传送顺序与节点在环中排列顺序相同。

如果某节点有数据帧要发送，它必须等待空闲令牌的到来。令牌在工作中有“闲”和“忙”两种状态。“闲”表示令牌没有被占用，即网中没有计算机在传送信息；“忙”表示令牌已被占用，即有信息正在传送。希望传送数据的计算机必须首先检测到“闲”令牌，将它置为“忙”的状态，然后在该令牌后面传送数据。当所传数据被目的节点计算机接收后，数据被从网中除去，令牌被重新置为“闲”。

令牌环网在物理上采用了星型拓扑结构，但逻辑上是环形拓扑结构。令牌环网的缺点是机制比较复杂。网络中的节点需要维护令牌，一旦失去令牌就无法工作，需要选择专门的节点监视和管理令牌。令牌环技术的保守，设备的昂贵，技术本身的难以理解和实现，都影响了令牌环网的普及。

FDDI（Fiber Distributed Data Interface，光纤分布式数据接口）也是一种利用了环形拓扑的局域网技术。其主要特点包括：

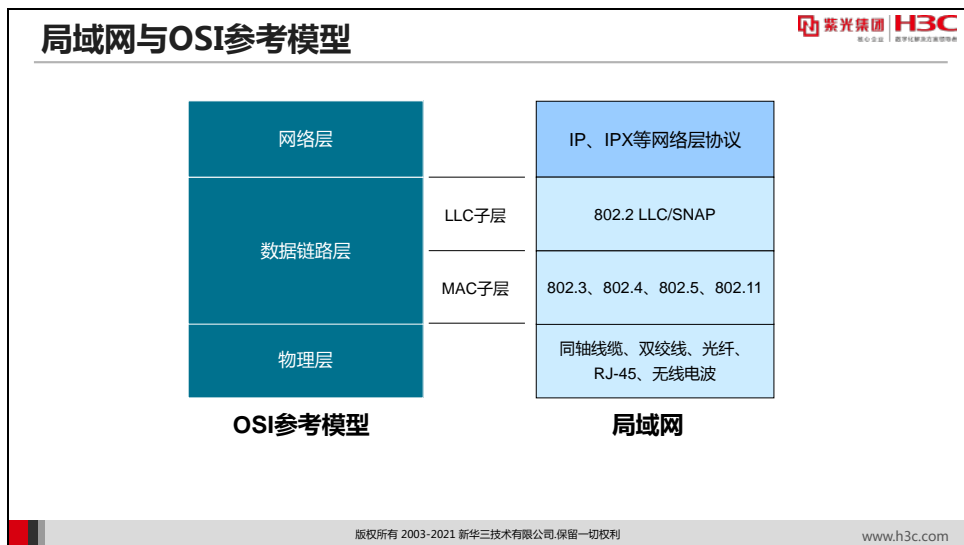
- 使用基于 IEEE 802.4 的令牌总线介质访问控制协议；
- 使用 IEEE 802.2 协议，与符合 IEEE 802 标准的局域网兼容；
- 数据传输速率为 100Mbps，连网节点数最大为 1000，环路长度为 100km；
- 可以使用双环结构，具有容错能力；
- 可以使用多模或单模光纤；
- 具有动态分配带宽的能力，能支持同步和异步数据传输。

由于 FDDI 在早期局域网环境中具有带宽和可靠性优势，其主要应用于核心机房、办公室或建筑物群的主干网、校园网主干等。

以太网自 Xerox、DEC 和 Intel 公司推出以来获得了巨大的成功。最初的以太网使用同轴电缆形成总线拓扑，随即又出现了用集线器（Hub）实现的星型结构，以及通过以太网交换机（switch）实现的交换式以太网。

随着以太网带宽的不断提高和可靠性的不断提升，令牌环和 FDDI 的优势已不复存在。由于其开放、简单、易于实现、易于部署的特性，以太网被广泛应用，迅速成为局域网中占统治地位的技术。

### 3.2.2 局域网与 OSI 参考模型



局域网技术主要对应于 OSI 参考模型的物理层和数据链路层。也即 TCP/IP 模型的网络接口层。

局域网的物理层规定了向局域网提供服务的设备、线缆和接口的物理电气特性、机械特性、连接标准等。常见的此类标准有：



- 用于 10BASE-T、100BASE-TX 和 1000BASE-T 的双绞线和 RJ-45 接头；
- 用于各种以太网传输的光纤；
- 用于 WLAN（Wireless LAN，无线局域网）的无线电波。

IEEE 将局域网的数据链路层划分为 LLC（Logic Link Control，逻辑链路控制）和 MAC（Media Access Control，介质访问控制）两个子层。上面的 LLC 子层实现数据链路层与硬件无关的功能，比如流量控制，差错恢复等；较低的 MAC 层提供 LLC 和物理层之间的接口。不同局域网 MAC 层不同，LLC 层相同。

数据链路层的主要功能之一是封装和标识上层数据，在局域网中这个功能由 LLC 子层实现。IEEE 802.2 定义了 LLC 子层，为 802 系列标准共用。

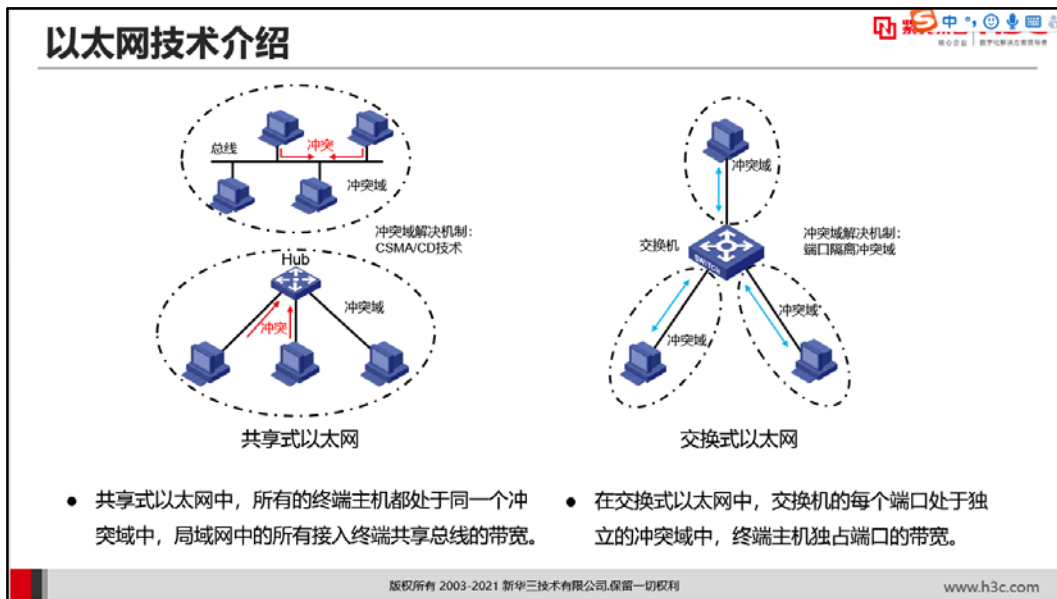
LLC 子层对网络层数据添加 802.2 LLC 头进行封装。为了区别网络层数据类型，实现多种协议复用链路，LLC 用 SAP（Service Access Point，服务访问点）标志上层协议。LLC 标准包括两个服务访问点——SSAP（Source Service Access Point，源服务访问点）和 DSAP（Destination Service Access Point，目的服务访问点），用以分别标识发送方和接收方的网络层协议。SAP 长度为 1 字节，且仅保留其中 6 位用于标识上层协议，因此其能够标识的协议数不超过 32 种。为保证在 802.2 LLC 上支持更多的上层协议，IEEE 发布了 802.2 SNAP（SubNetwork Access Protocol）标准。802.2 SNAP 也用 LLC 头封装上层数据，但其扩展了 LLC 属性，将 SAP 的值置为 AA，而新添加了一个 2 字节长的协议类型（Type）字段，从而可以标识更多的上层协议。

数据链路层的另一个主要功能是适应种类多样的传输介质，并且在任何一种特定的介质上处理信道的占用、站点的标识和寻址问题。在局域网中这个功能由 MAC 子层实现。由于 MAC 子层因不同的物理层介质而不同，它分别由多个标准分别定义。例如 802.3 定义了以太网（Ethernet）的 MAC 子层，802.4 定义了令牌总线网（Token Bus）的 MAC 子层，而 802.5 定义了令牌环网（Token Ring）的 MAC 子层。此外，MAC 层还负责对入站数据帧进行完整性校验。

## 3.3 以太网技术基础

为了便于理解以太网技术的发展情况和基本原理，本书将从早期的共享式以太网技术到现代的交换式以太网技术开始讲解。

### 3.3.1 以太网技术介绍



早期的共享式以太网的典型代表是使用 10Base2/10Base5 的总线型网络和以集线器为核心的星型网络。IEEE 802.3 规定了 10Mbps 的以太网标准。在使用集线器的以太网中，集线器将很多以太网设备集中到一台中心设备上，这些设备都连接到集线器中的同一物理总线结构中，因此实际上以集线器为核心的以太网与总线型以太网并无本质区别。

冲突域是指连接在同一共享介质上的所有节点的集合，冲突域内所有节点竞争同一带宽，一个节点发出的报文（无论是单播、组播、广播），其余节点都可以收到。在早期共享式以太网中，同一介质上多个节点在一个冲突域，不管一个帧从哪里来或到哪里去，所有的节点都能接收到这个数据帧，随着节点的增加，大量的冲突将导致网络性能急剧下降。为了避免节点冲突带来的影响，以太网中使用 CSMA/CD（载波监听多路访问/冲突检测）技术在一定程度上缓解了该问题。CSMA/CD 的基本工作过程如下：

终端设备不停的检测共享线路的状态。如果线路空闲则发送数据，如果线路不空闲则一直等待。如果有另外一个设备同时发送数据，两个设备发送的数据必然产生冲突，导致线路上的信号不稳定。终端设备检测到这种不稳定之后，马上停止发送自己的数据。终端设备发送一连串干扰脉冲，然后等待一段时间之后再进行发送数据。发送干扰脉冲的目的是为了通知其他设备，特别是跟自己在同一个时刻发送数据的设备，线路上已经产生了冲突。

共享式以太网存在的主要问题是所有节点共享带宽，每个节点的实际可用带宽随网络节点数的增加而递减。这是因为当信息繁忙时，多个节点都可能同时争用一个信道，而一个信道在

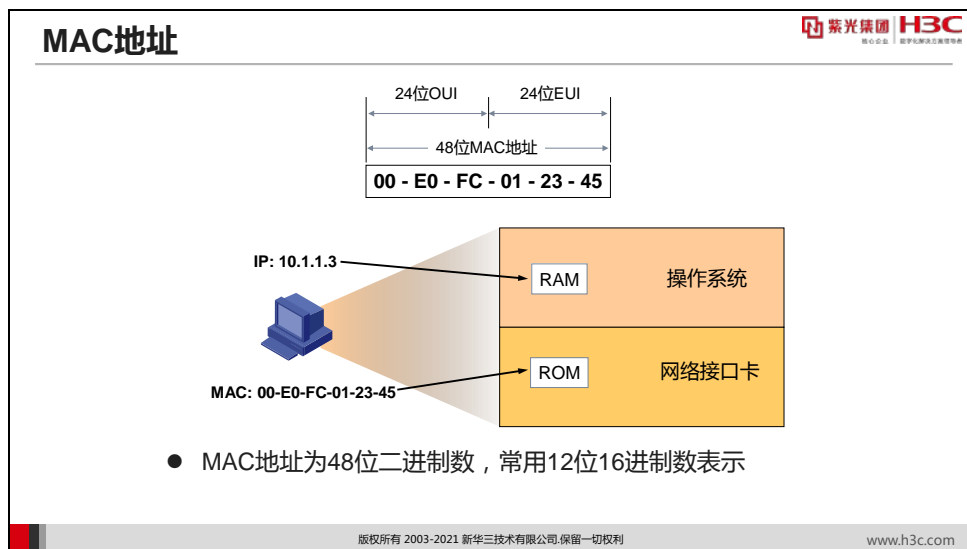
某一时刻只能被一个节点占用，因此会出现大量节点经常处于监测等待状态，使得信号在传送时发生抖动、停滞或失真，进而严重影响了网络的性能。

交换式以太网是以交换式集线器或者交换机为中心构建的星形拓扑结构网络。在交换式以太网中，交换机根据接收到的数据帧中的 **MAC** 地址决定数据帧应发往交换机的哪个端口。因为端口间的帧传输彼此屏蔽，因此节点就不必担心自己发送的数据帧在通过交换机时是否会与其他节点发送的帧发生冲突，即各个接口属于不同的冲突域中。

在交换式以太网中，交换机为每个节点提供专用的信息通道，除非两个源端口企图将信息同时发往同一目的端口，否则各个源端口与各自的目的端口之间可同时进行通信而不发生冲突，这样每个节点都可以使用全部的带宽，而不是各个节点共享带宽。

由于能减少冲突和提升带宽的优势，交换式以太网逐步替代了共享式以太网，且交换机只是在工作方式上与集线器不同，其它的连接方式、速度选择等则与集线器基本相同，所以在使用交换式以太网替换共享式以太网时不需要改变原有网络的其它硬件，包括电缆和用户网卡，仅需要用交换式集线器或交换机替换传统的集线器，因此可以节省用户网络升级的费用。

### 3.3.2 MAC 地址

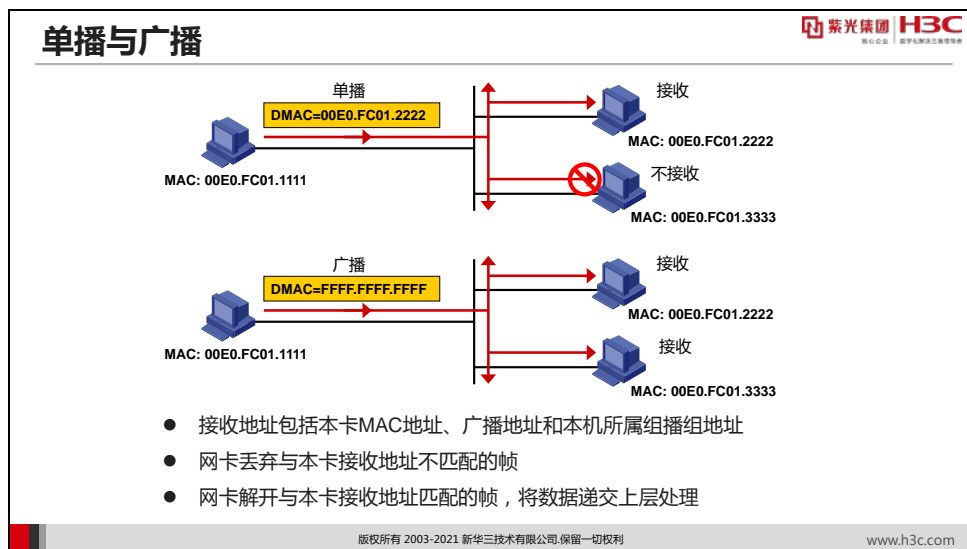


以太网上的计算机用 MAC 地址（Medium Access Control Address，介质访问控制地址）作为自己的唯一标识。MAC 地址为二进制 48 位，常用 12 位十六进制数表示。

MAC 地址分为 24 位的 OUI（Organizationally Unique Identifier，组织唯一标识符）和 24 位的 EUI（Extended Unique Identifier，扩展唯一标识符）两部分。IEEE RA（Registration Authority）是 MAC 地址的法定管理机构，负责分配 OUI；组织自行分配其 EUI。

MAC 地址固化在网卡的 ROM（Read Only Memory，只读存储器）中，每次启动时由计算机读取出来，因此也称为硬件地址（Hardware Address）。每块网卡的 MAC 地址是全球唯一的，也即全网唯一的。一台计算机可能有多个网卡，因此也可能同时具有多个 MAC 地址。

## 3.3.3 以太网单播和广播



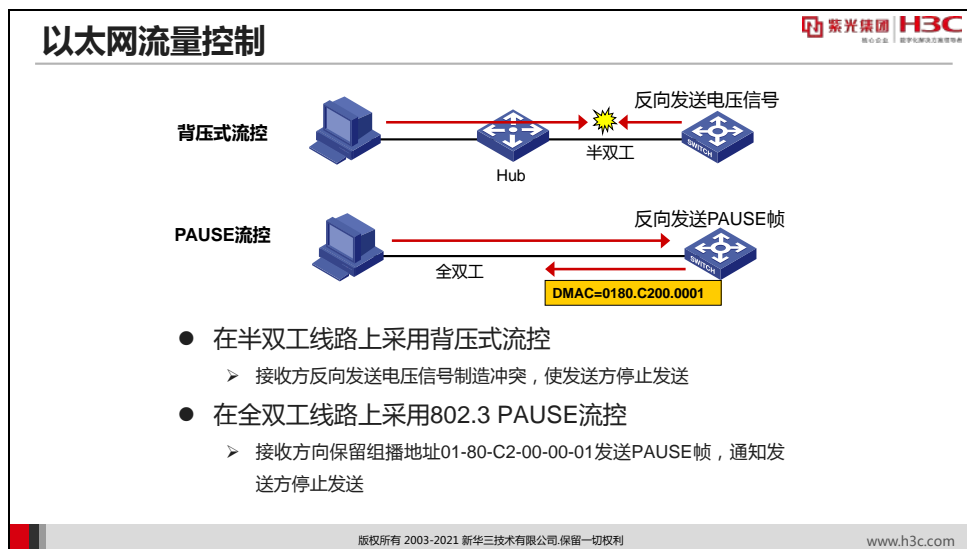
以太网中包含两个 MAC 地址，一个是发送者的 MAC 地址，称为源 MAC 地址，另一个是帧接收者的地址，称为目的 MAC 地址。目的为单一站点的发送称为单播（Unicast）；目的为全部站点的发送称为广播（Broadcast）；目的为某一组特定站点的发送称为组播（Multicast）。

发送单播时，帧的目的 MAC 地址填写为目的站点的 MAC 地址；发送广播时，目的 MAC 地址填写为以太网广播地址 FFFF.FFFF.FFFF，表示发送给全体站点；发送组播时，目的 MAC 地址填写为某一相应的组播 MAC 地址。

以太网卡具有过滤（filtering）功能。网卡只将发送给自己的帧接收、解封装并提交给上层协议处理；对于不是发送给自己的帧则一律丢弃。为了实现这个功能，网卡维护一个接收地址表，表中存储有自己的 MAC 地址、广播地址以及自己所属的组播组 MAC 地址。收到一个帧时，网卡首先将其目的地址与此接收地址表中的地址加以比较，若发现匹配则说明此帧是发给自己的。因此，虽然每个帧的物理信号能到达所有的站点，但只有正确的站点才能收到。

不过，还有一些网卡可以工作于混杂模式（promiscuous mode），即可以接收任意帧，而不考虑这些帧是否发送给自己。这类网卡通常用于 Sniffer 等网络监视工具中。

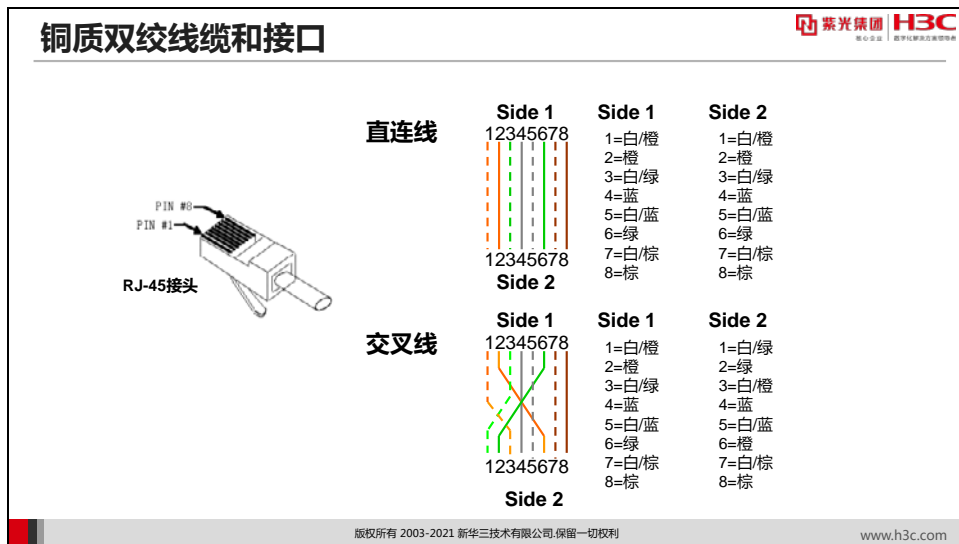
## 3.3.4 以太网流量控制



IEEE 802.3 还规定了在全双工环境中用 **PAUSE** 操作控制流量的方法。在全双工环境中，当接收方来不及处理数据时，可以向保留组播地址 **0180.C200.0001** 发送 **64** 字节的 **PAUSE** 帧，告诉发送方暂停发送。

在半双工以太网上则利用背压式（**back pressure**）方法进行流量控制。当接收方来不及处理数据时，可以向线路上发送一个电压信号，强行制造冲突，使得发送方暂时退避，从而允许接收方去处理积聚在其缓冲区中的数据。

## 3.3.5 铜质双绞线缆和接口



一对双绞线（**twisted pair**）由两根具有绝缘保护层的铜导线组成。每根铜导线都包覆有绝缘材料，两根线再按一定密度相互绞在一起，就可改变导线的电气特性，降低信号干扰的程度。双绞线通常为 8 芯（4 对）构成一根电缆。

双绞线分为 UTP（**Unshielded Twisted Pair**，无屏蔽双绞线）和 STP（**Shielded Twisted Pair**，屏蔽双绞线）。STP 配有类似于同轴电缆的屏蔽功能，价格相对高一些，安装比无屏蔽双绞线电缆困难。UTP 没有屏蔽功能，成本低，应用十分普及，在大多数中小企业内部局域网、网吧、家庭中均使用 UTP。

根据单位长度内绞环数的不同，常见的双绞线可分为 3 类 UTP、4 类 UTP、5 类 UTP、超 5 类 UTP 和 6 类 UTP 等。其中 5 类 UTP 应用尤为广泛。

双绞线的直连线和交叉线线序如上图所示。早期以太网中，各种设备网卡接口类型不同，互相连接时可能需要不同的线序。但目前绝大多数网络设备都支持网卡接口类型自适应，在连接时不必考虑所用网线为直连线还是交叉线。

### 3.3.6 常用光介质和连接器

#### 单模光纤与多模光纤

- **多模光纤**
  - 较粗的纤芯，传输多种不同波长不同角度的光
  - 衰耗大，传输距离通常在千米以内
  - 成本低
- **单模光纤**
  - 纤芯与光波长相同，传送单一波长的激光
  - 衰耗小，传输距离可达数十千米
  - 成本高

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

光导纤维是一种传输光束的介质。光导纤维线缆由一捆纤维组成，简称为光缆。光纤中传输的信号为光脉冲信号，光源通常为发光二极管或半导体激光器。

光纤用于数据传输有以下几个优点：

- **带宽高**：由于光的传输频率非常高，所以光纤传输的带宽非常高，最高可达几十 Gbps。
- **距离远**：同轴电缆和双绞线每隔几千米就需要接一个中继器，而光信号在光纤中衰减较小，可以传输很远的距离，可以减少整个通道中继器的数目，降低成本。
- **可靠性高**：由于光纤中传输的是光束，不但衰减小，而且不受外界电磁干扰影响，传输时可靠性很高。
- **安全性好**：光纤传输信号时本身不向外产生辐射，加上光纤的切割和连接比较困难，很难从中途进行窃听。因此它适用于要求高度安全的场合。
- **频带较宽**：同一根光纤可以同时传输各种频率的光信号，因而采用波分复用（Wavelength Division Multiplexing, WDM）技术可以大幅度提高光纤的传输带宽。

多模光纤（multi-mode fiber）采用较粗的纤芯，以发光二极管作为光源。入射光线利用全反射原理在光纤内传递。由于其入射光线的角度分散，经过长距离传输时光脉冲峰谷会逐渐模糊，而造成失真，因此适用于近距离传输，距离通常在千米以内。但其光纤和光源制造工艺要求低，成本也低。

单模光纤（single-mode fiber）纤芯直径通常为微米级，等于光波的波长，此时光线可以沿光纤直接前进，而不会产生多次全反射，因此失真小，传输距离很远，可达几十千米。但单模光纤要求光源发送单一波长的激光，并且对光纤的制造工艺要求较高，因此成本也比较高。



## 常用光纤连接器



- ST：卡接式圆形光纤接头



- FC：带螺纹的圆形光纤接头



版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

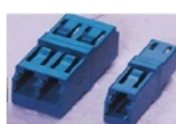
## 常用光纤连接器（续）



- SC：矩型光纤接头



- LC：一种Mini型连接器



版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

光纤需要通过光纤接头连接到设备上。常见光纤接头有：

- ST：圆形卡接式光纤接头；
- FC：圆形带螺纹光纤接头；
- SC：矩形插接式光纤接头；
- LC：由 LUCENT 开发的一种微型连接器；
- MT-RJ：一头双纤、收发一体，节约设备面板空间；
- MTP/MPO：一种特殊类型的多光纤连接器。

### 3.3.7 快速以太网和千兆以太网

快速以太网和千兆以太网				
名称	速度	介质类型	最大线缆长度	协议标准
100BASE-TX	100 Mbps	2对5类UTP	100m	802.3u
100BASE-FX	100 Mbps	多模光纤	2000m	
100BASE-T4	100 Mbps	4对3类UTP	100m	
1000BASE-SX	1 Gbps	多模光纤	275m / 550m	802.3z
1000BASE-LX	1 Gbps	单模光纤	550m / 5000m	
1000BASE-CX	1 Gbps	2对STP	25m	
1000BASE-T	1 Gbps	4对5类UTP	100m	802.3ab

新一代多媒体、影像和数据库应用很容易将早期 10Mbps 以太网的带宽吞没。诸如快速以太网、千兆以太网、万兆以太网等速度更快的以太网也早已出现。

802.3u 定义了速度为 100Mbps 的快速以太网（Fast Ethernet）系列标准。快速以太网除速度提高到 100Mbps 之外，使用与传统以太网相同的封装，使用同样的 CSMA/CD 算法。但快速以太网屏弃了同轴线缆，全部采用星型拓扑结构。

快速以太网中最流行的是 100BASE-TX。100BASE-TX 采用 2 对 5 类 UTP 线和 RJ-45 接头，同样采用以集线器为核心构建的星型拓扑，单条线缆长度可达 100m。10BASE-T 已经得到广泛应用，并且实际上大多已采用 5 类 UTP，而很多建筑也将 5 类双绞线作为布线标准，只要采购 100BASE-TX 集线器就可以方便地利用现有布线环境升级到 100Mbps 的速度，因此 100BASE-TX 很快流行起来。

在已经布设了 3 类 UTP 的场所，802.3u 提供了另一种解决方案——100BASE-T4。由于 3 类 UTP 信号衰减快，容易受到干扰，100BASE-T4 使用 4 对 3 类 UTP 提供 100Mbps 的带宽，单条线缆长度可达 100m。

另一种方案是使用光缆的 100BASE-FX，它使用 2 束多模光纤提供 100Mbps 的带宽，传送距离可达 2000m。

快速以太网的应用范围较广，已经成为接入设备的基本接入技术。相应地，在网络的汇聚点或服务器接入点等流量较大的位置就需要一种带宽更高的连接技术，千兆以太网（Gigabit Ethernet）应运而生。

千兆以太网仍然使用 IEEE 802.3 帧格式，在半双工方式下仍然使用 CSMA/CD 处理冲突，并且将以太网速率提升至 1Gbps。

IEEE 802.3z 定义的千兆以太网标准如下：

- 1000BASE-SX 主要适用于多模光纤传输线路。其使用 850nm 短波激光。在采用直径 50 $\mu$ m 的多模光纤时传输距离可达 275m, 采用直径 62.5 $\mu$ m 的多模光纤时传输距离可达 550m。
- 1000Base-LX 主要为适应单模光纤传输线路而设计。其使用 1310nm 长波激光。在采用直径 50 $\mu$ m/62.5 $\mu$ m 的多模光纤时传输距离可达 550m, 采用直径 10 $\mu$ m 的单模光纤时传输距离可达 5000m。
- 1000BASE-CX 使用 2 对 STP (Shielded Twisted-Pair), 最大传输距离 25m。

802.3ab 定义了基于铜线的千兆以太网——1000BASE-T。其采用 4 对 5 类 UTP, 最大传输距离 100m。

以太网技术发展到快速以太网和千兆以太网以后, 出现了与原 10M 以太网设备兼容的问题, 自协商技术就是为了解决这个问题而制定的。100BASE-TX 和 1000BASE-T 都定义了向下兼容到 10BASE-T 的自协商技术。

自协商功能允许一个网络设备将自己所支持的工作模式以自协商报文的方式传达给线缆上的对端, 并接收对方可能传递过来的相应信息。自协商功能完全由物理层芯片设计实现, 因此其速度很快, 且不带来任何高层协议开销。

如果对端设备不支持自协商, 默认假设其工作于 10M 半双工模式, 不使用显式的流量控制机制。自协商功能虽然方便易用, 但仍然存在一定的延迟, 也不能排除协商错误的可能性, 因此建议仅在普通端用户接入端口启动自协商, 而对服务器、路由器等连接端口使用固定配置参数。

## 3.3.8 万兆以太网

万兆以太网				
名称	速度	介质类型	最大线缆长度	协议标准
10GBase-SR	10 Gbps	多模光纤	300m	802.3ae
10GBase-LR	10 Gbps	单模光纤	10km	
10GBase-ER	10 Gbps	单模光纤	40km	
10GBase-ZR	10 Gbps	单模光纤	80km	厂商自有规范
10GBase-LRM	10 Gbps	多模光纤	260m	802.3aq
10GBase-LX4	10 Gbps	多模或单模	300m (多模) /10km (单模)	802.3ae
10GBase-CX4	10 Gbps	屏蔽双绞线	15m	802.3ak

版权所有 2003-2021 新华三技术有限公司,保留一切权利

www.h3c.com

万兆以太网 (续)				
名称	速度	介质类型	最大线缆长度	协议标准
10GBase-T	10 Gbps	6类、6a类双绞线	55米 (6类线) 100米 (6a类线)	802.3an
10GBase-KX4	10 Gbps	铜线 (并行接口)	1m	802.3ap
10GBase-KR	10 Gbps	铜线 (串行接口)	1m	
10GBase-SW	10 Gbps	多模光纤	300m	802.3ae
10GBase-LW	10 Gbps	单模光纤	10km	
10GBase-EW	10 Gbps	单模光纤	40km	

版权所有 2003-2021 新华三技术有限公司,保留一切权利

www.h3c.com

随着网络应用的发展,带宽的消耗也成倍增长。对于城域网或大型校园网来说,千兆数量级已满足不了核心设备间的互连带宽需求了。万兆以太网技术的出现将以太网性能提升到一个新的高度。

万兆以太网标准和规范都比较繁多,有2002年的IEEE 802.3ae,2004年的IEEE 802.3ak,2006年的IEEE 802.3an和IEEE 802.3aq,以及2007年的IEEE 802.3ap。

在万兆以太网规范方面,仅由上述IEEE标准中发布的规范就有10多个,如2002年在IEEE 802.3ae标准中发布的基于光纤的规范包括:10GBase-SR、10GBase-LR、10GBase-ER、10GBase-LX4、10GBase-SW、10GBase-LW、10GBase-EW;2004年在IEEE 802.3ak标准中发布的基于双绞线的10GBase-CX4;2006年在IEEE 802.3an标准发布的基

于双绞铜线的 10GBase-T；2006 年在 IEEE 802.3aq 标准中发布的基于光纤的 10GBase-LRM；2007 年在 IEEE 802.3ap 标准中发布的基于铜线的 10GBase-KR 和 10GBase-KX4。

以上这 10 多种万兆以太网规范可以分为三类：一是基于光纤的局域万兆以太网规范，二是基于双绞线（或铜线）的局域万兆以太网规范，三是基于光纤的广域万兆以太网规范。下面分别予以介绍。

### 1) 基于光纤的局域万兆以太网规范

用于局域网的光纤万兆以太网规范有：10GBase-SR、10GBase-LR、10GBase-LRM、10GBase-ER、10GBase-LX4。

**10GBase-SR:** 10GBase-SR 中的“SR”是“short range”（短距离）的缩写，表示仅用于短距离连接。该规范支持编码方式为 64B/66B 的短波（波长为 850nm）多模光纤（MMF），有效传输距离为 2 米到 300 米。但要支持 300 米传输需要采用经过优化的 50  $\mu\text{m}$  线径 OM3（Optimized Multimode 3，优化的多模 3）光纤（没有优化的线径 50  $\mu\text{m}$  光纤称之为 OM2 光纤，而线径为 62.5 $\mu\text{m}$  的光纤称之为 OM1 光纤）。

**10GBase-LR:** 10GBase-LR 中的“LR”是“Long Range”（长距离）的缩写，表示主要用于长距离连接。该规范支持编码方式为 64B/66B 的长波（1310nm）单模光纤（SMF），有效传输距离为 2 米到 10 公里，事实上最高可达到 25 公里。

**10GBase-ER:** 10GBase-ER 中的“ER”是“Extended Range”（超长距离）的缩写，表示连接距离可以非常长。该规范支持编码方式为 64B/66B 的超长波（1550nm）单模光纤（SMF），有效传输距离为 2 米到 40 公里。

**10GBase-LRM:** 10GBase-LRM 中的“LRM”是“Long Reach Multimode”（长距离延伸多点模式）的缩写，表示主要用于长距离的多点连接模式，对应的标准为 2006 年发布的 IEEE 802.3aq，采用 64B/66B 编码方式。采用该规范时，在 1990 年以前安装的 FDDI 62.5  $\mu\text{m}$  多模光纤 FDDI 网络和 100Base-FX 网络中的有效传输距离为 220 米，而在 OM3 光纤中可达 260 米，在连接长度方面，不如以前的 10GBase-LX4 规范，但是它的光纤模块比 10GBase-LX4 规范光纤模块具有更低的成本和更低的电源消耗。

**10GBase-LX4:** 10GBase-LX4 规范在 IEEE 802.3ae 标准中发布，设计通过波分复用技术采用 4 束光波通过单对光学电缆来发送信号，采用 8B/10B 编码方式。10GBase-LX4 工作波长为 1310nm，使用多模或单模暗光纤，主要适用于在需要在一个光纤模块中同时支持多模和单模光纤的环境。多模光纤传输距离为 240~300 米，单模光纤 10 公里以上，根据电缆类型和质量，还能达到更远的距离。

### 2) 基于双绞线（或铜线）的局域网万兆以太网规范

在 2002 年发布的几个万兆以太网规范中并没有支持铜线这种廉价传输介质的，但事实上，像双绞线这类铜线在局域网中的应用是最普遍的，不仅成本低，而且还容易维护，所以在近几年就相继推出了多个基于双绞线（6 类以上）的万兆以太网规范。它们包括：10GBase-CX4、10GBase-T、10GBase-KX4、10GBase-KR。下面分别予以简单介绍

**10GBase-CX4:** 它的有效传输距离仅 15 米。10GBase-CX4 规范不是利用单个铜线链路传送万兆数据, 而是使用 4 台发送器和 4 台接收器来传送万兆数据, 并以差分方式运行在同轴电缆上, 每台设备利用 8B/10B 编码, 以每信道 3.125GHz 的波特率传送 2.5Gbps 的数据。这需要在每条电缆组的总共 8 条双同轴信道的每个方向上有 4 组差分线缆对。另外, 与可在现场端接的 5 类、超五类双绞线不同, CX4 线缆需要在工厂端接, 因此客户必须指定线缆长度。线缆越长一般直径就越大。

**10GBase-T:** 10GBase-T 是基于屏蔽或非屏蔽双绞线, 主要用于局域网的万兆以太网规范, 最长传输距离为 100 米。这可以算是万兆以太网一项革命性的进步, 因为在此之前, 一直认为在双绞线上不可能实现这么高传输速率, 原因就是运行在这么高工作频率(至少为 500MHz)基础上的损耗太大。但标准制定者依靠损耗消除、模拟到数字转换、线缆增强和编码改进 4 项技术构件使 10GBase-T 变为现实。在连接器方面, 10GBase-T 使用已广泛应用于以太网的 650 MHz 版本 RJ-45 连接器。在 6 类线上最长有效传输距离为 55 米, 而在 6a 类类双线上可以达到 100 米。

**10GBase-KX4 和 10GBase-KR:** 10GBase-KX4 和 10GBase-KR 两个规范主要用于设备背板连接中, 如刀片服务器、路由器和交换机的集群线路卡, 所以又称之为“背板以太网”。万兆背板连接目前已经存在并行和串行两种版本: 并行版本(10Gbase-KX4 规范)是背板的通用设计, 将万兆信号拆分为四条通道(类似 XAUI), 每条通道的带宽都是 3.125Gbps。串行版本(10GBase-KR 规范)中只定义了一条通道, 采用 64/66B 编码方式实现 10Gbps 高速传输。10Gbase-KX4 使用与 10GBase-CX4 规范一样的物理层 8B/10B 编码, 10GBase-KR 使用与 10GBase-LR/ER/SR 这三个规范一样的物理层 64B/66B 编码。

### 3) 基于光纤的广域网万兆以太网规范

10G 以太网一个最大改变就是它不仅可以在局域网中使用, 还可应用于广域网中, 其对应的规范包括: 10GBase-SW、10GBase-LW、10GBase-EW。广域网 10G 广域以太网规范专为工作在 OC-192/STM-64 SDH/SONET 环境而设置, 使用 SDH (Synchronous Digital Hierarchy, 同步数字体系)/SONET (Synchronous optical networking, 同步光纤网络) 帧, 运行速率为 9.953 Gbps。它们所使用的光纤类型和有效传输距离分别对应于前面介绍的, 应用于局域网中的 10GBase-SR、10GBase-LR、10GBase-ER 规范。

## 3.3.9 超高带宽以太网

超高带宽以太网				
名称	速度	介质类型	最大线缆长度	协议标准
25GBASE-SR	25 Gbps	多模光纤	100m	802.3by
25GBASE-CR	25 Gbps	双轴铜缆	5m	
40GBASE-SR4	40 Gbps	多模光纤	100m	802.3ba
100GBASE-SR4	100 Gbps	多模光纤	100m	
200GBASE-SR4	200 Gbps	多模光纤	100m	802.3bs
400GBASE-SR16	400 Gbps	多模光纤	100m	

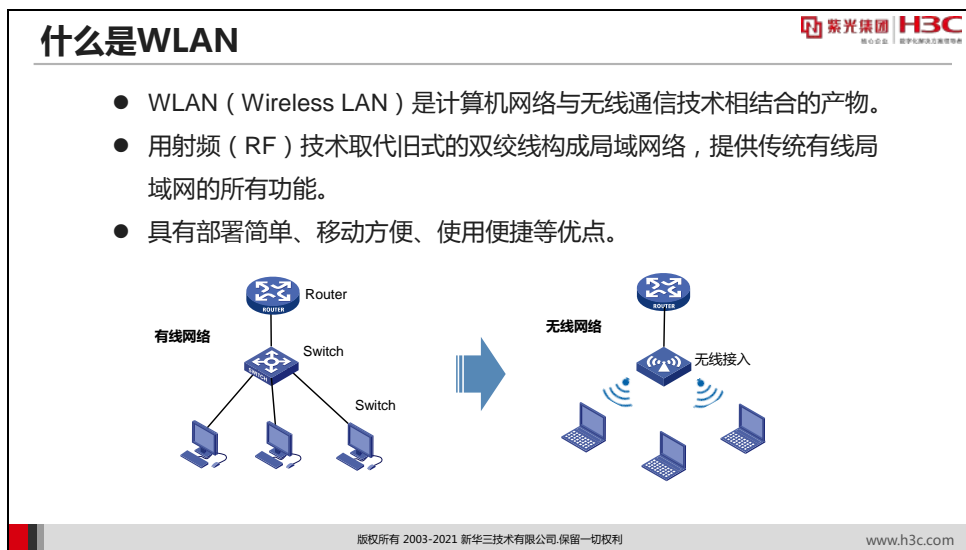
伴随着 5G、云计算、物联网的爆炸式增长，加上服务器和存储解决方案支持的高吞吐量，数据中心的带宽需要不断增长，以满足当前和未来云端的海量数据流需求。高带宽不可否认地正推动着数据中心朝着更高扩展性和灵活性的方向发展。25G、40G、100G、200G、400G 网络方案作为一种高带宽、高密度、低成本、低功耗的解决方案应运而生，推动着数据中心网络朝着更高性能和灵活性的方向发展。

25G 以太网根据 IEEE 802.3by 协议制定，使用四根光纤和成对铜缆并行的方法，通过 4 个 25-Gbit/s 通道实现 100G 的以太网传输速率，可实现 3 到 5 米的铜双轴线以及 100 米的多模光纤传输，旨在满足更多客户的需求，即 10G 以太网的速度标准满足不了大中企业网络流量高速增长的需求。

作为最新的网络应用，2010 年 IEEE 组织又发布了 IEEE 802.3ba 标准，该标准同时定义了两种传输速率 (40 GBE 和 100 GBE) 的基于光纤传输网络应用。2017 年 12 月 6 日，IEEE 802.3 以太网工作组正式批准了新的 IEEE 802.3bs 以太网定义标准，包括 200G 以太网 (200GbE) 和 400G 以太网 (400GbE)，200/400GbE 标准覆盖各种互连应用，超高带宽可完全满足云扩展数据中心、互联网交换、主机托管服务、服务供应商网络等各种带宽密集型应用的需求，并大大降低端口成本。

## 3.4 WLAN基础

### 3.4.1 WLAN 简介



WLAN 即 Wireless LAN（无线局域网），无线局域网是计算机网络与无线通信技术相结合的产物。它以射频（RF）技术取代旧式的双绞线构成局域网络，提供传统有线局域网的所有功能。无线网络所需的基础设施不需埋在地下或隐藏在墙里，并且可以随需移动或变化。这里指的无线技术不仅仅包含我们日常随处可见的 Wi-Fi，还有红外、蓝牙、ZigBee 等等。WLAN 已经成为宽带接入的有效手段之一，使用 WLAN 的区域及其承载的业务愈来愈多。

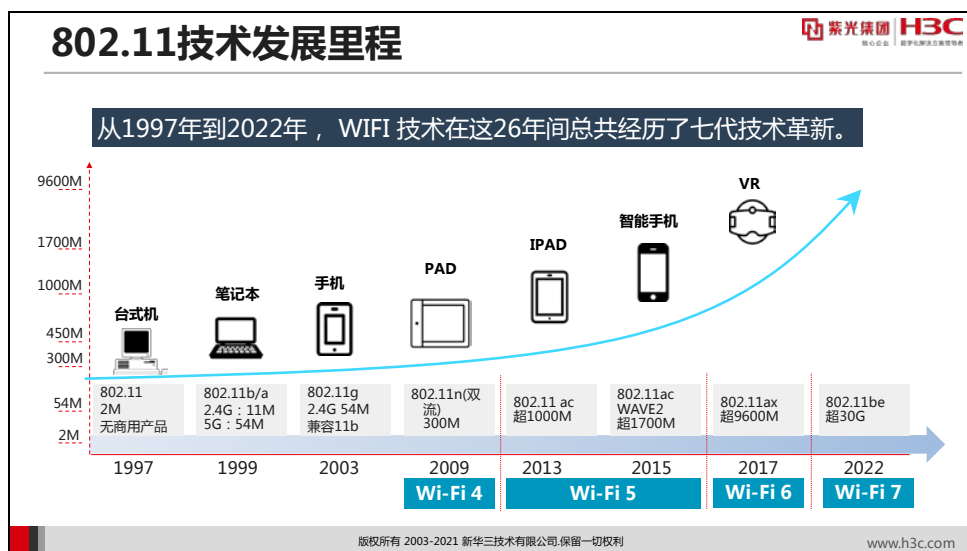
和传统的有线接入方式相比，WLAN 的优点：

- 网络使用自由：凡是自由空间均可连接网络，不受限于线缆和端口位置。在办公大楼、机场候机厅、度假村、商务酒店、体育场馆、咖啡店等场所尤为适用。
- 网络部署灵活：对于地铁、公路交通监控等难于布线的场所，采用 WLAN 进行无线网络覆盖，免去或减少了繁杂的网络布线，实施简单，成本低，扩展性好。

本课程介绍的 WLAN 特指通过 Wi-Fi 技术基于 802.11 标准系列，利用高频信号（例如 2.4GHz 或 5GHz）作为传输介质的无线局域网。



## 3.4.2 802.11 协议的发展进程



早在 20 世纪 80 年代中期，一些公司已经开始推出 WLAN 的雏形产品，随着 WLAN 技术的不断发展和国际标准的成熟，IEEE（The Institute of Electrical and Electronics Engineers）美国电气与电子工程师学会在 1997 年发布了第一个 WLAN 的 802.11 国际标准 IEEE 802.11，并在后续的几年中在 802.11 标准基础上相继衍生推出了包括 802.11b（2.4GHz）、802.11a（5GHz）、802.11g（2.4GHz）、802.11n（2.4G 和 5G）在内的多种 WLAN 物理层技术。

随着时间的推移，基于 802.11、802.11b、802.11a、802.11g 技术的产品已经逐渐退出市场，被速率超过百兆的 802.11n 所替代。目前在市场上，802.11n/802.11ac 的产品占据了市场的主流，且几乎成为所有智能手机、平板、笔记本的标配。目前 IEEE 定义了速率更高的 802.11ax 的标准，新的标准可以支持高达 9.6Gbps 的物理层发送速率，完美支持高带宽需求。可以预见，在较长一段时间内 802.11n、802.11ac、802.11ax 会是主流的 WiFi 协议标准。现在新出厂的旗舰手机、主流笔记本都已经预装 802.11ax 模块。而 802.11be 协议也就是 wifi7 也已经立项，最高速率可达 30Gbps。

## 3.4.3 802.11 协议概览

802.11协议体系概览						
	802.11 802.11b	802.11a	802.11g	802.11n	802.11ac	802.11ax
发布时间	1999	1999	2003	2009	2012	2019
合法频宽	83.5MHz	325MHz	83.5MHz	83.5MHz 325MHz	325MHz	83.5MHz 325MHz
频段范围 (中国)	2.4GHz	5GHz	2.4GHz	2.4GHz 5GHz	5GHz	2.4GHz 5GHz
非重叠信道	3个	13个	3	2.4G : 3个 5G : 13个	13个	2.4G : 3个 5G : 13个
理论最大物理发送速率	2Mbps/ 11Mbps	54Mbps	54Mbps	600Mbps	6933Mbps	9607.8Mbps
兼容性	11b	11a	11b/g	11a/b/g/n	11a/b/g/n/ac	11a/b/g/n/ac/ax

版权所有 2003-2021 新华三技术有限公司 保留一切权利

www.h3c.com

无线设备被限定在某个特定频段（frequency band）上操作。每个频段都有相应的频宽（bandwidth），亦即该频段可供使用的频率空间总和。频宽是评价链路（link）数据传输能力的基准。数学、信息以及信号处理理论均可证明，较大的频宽可以传输更多的信息。

WLAN 技术中常用的一些名词术语如下：

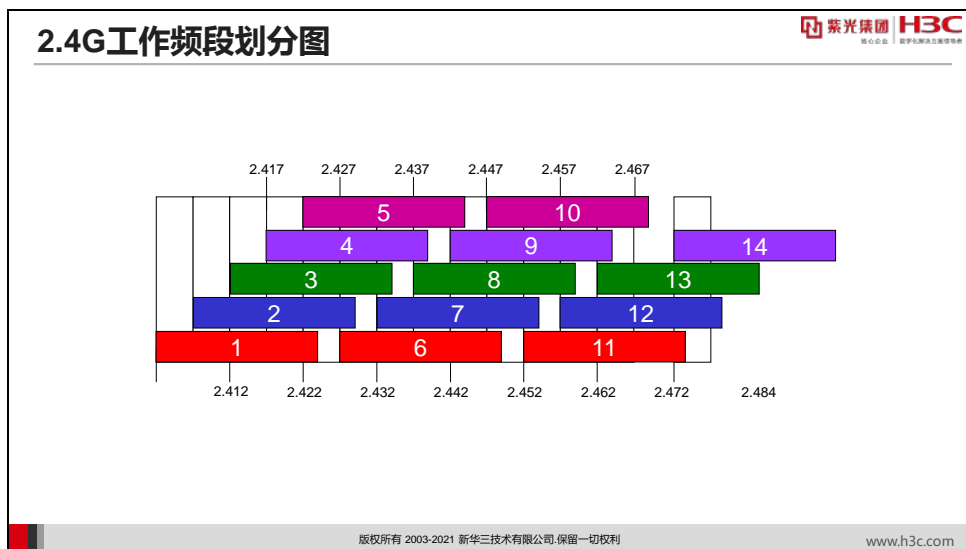
- 合法频宽\频率范围：国家或国际相关组织为特定无线设备规定的工作的频率范围。
- 非重叠信道：互相之间频段不交叠的信道。
- 调制技术：将数字基带信号变成模拟信号并通过电磁波发送出去的方法。
- 物理发送速率：物理层发送数据的速度，单位是 Mbps 或 Gbps，同时本速率也与终端协商、信道频宽强相关。
- 无线覆盖范围：无线设备发射能量所能到达的距离，一般以实际环境为准。

1990 年 IEEE 802 标准化委员会成立 IEEE 802.11 无线局域网标准工作组致力于 WLAN 相关领域的技术研究和标准定义，IEEE 802.11 无线局域网标准由物理层（PHY 层）和 MAC 层两部分的相关协议组成。PHY 层相关的 802.11 主要标准有：

- IEEE 802.11 标准：该标准定义物理层和媒体访问控制规范。这也是在无线局域网领域内的第一个国际上被认可的协议。在这个标准中，提供了 1Mbps 和 2Mbps 的数据传输速率以及一些基本的信令规范和服务规范。只支持 2.4G 频段。
- IEEE 802.11b 标准：1999 年 9 月被正式批准。该标准规定无线局域网工作频段在 2.4GHz~2.4835GHz，数据传输速率达到 11 Mbps。该标准是对 IEEE 802.11 的一个补充，引入 CCK 调制方式。在数据传输速率方面可以根据实际情况在 11 Mbps、5.5 Mbps、2 Mbps、1Mbps 的不同速率间自动切换。只支持 2.4G 频段。

- **IEEE 802.11a 标准：**1999 年制定完成。该标准规定无线局域网工作频段在 5.15~5.825GHz，数据传输速率达到 54 Mbps。802.11a 采用正交频分复用（OFDM）的独特扩频技术。只支持 5G 频段。
- **IEEE 802.11g 标准：**2003 年 6 月被正式批准。该标准可以视作对 802.11b 标准的提速（速率从 802.11b 的 11 Mbps 提高到 54Mbps），但仍然工作在 2.4G 频段。802.11g 采用两种调制方式，分别是 802.11a 的 OFDM 与 802.11b 的 CCK。故采用 802.11g 的终端可访问现有的 802.11b 接入点和新的 802.11g 接入点。只支持 2.4G 频段。
- **IEEE 802.11n 标准：**通过对 802.11 物理层和 MAC 层的技术改进，使得无线通信在吞吐量和可靠性方面都获得显著提高，速率可达到 600Mbps，其核心技术为 MIMO+OFDM。同时，802.11n 可以工作在双频模式，包含 2.4GHz 和 5GHz 两个工作频段，可以与 802.11a/b/g 标准兼容。
- **IEEE 802.11ac 标准：**在 802.11n 的基础上，通过引入 MU-MIMO、更宽的信道、更高阶的调制实现超过 1Gbps 的物理速率。因频频资源和干扰的原因，802.11ac 只支持 5G 频段，可以与 802.11a/an 标准兼容。
- **IEEE 802.11ax 标准：**802.11ax 是在 802.11ac 以后，无线局域网协议本身的进一步扩展，是第六代无线局域网标准，与 802.11ac 只能工作在 5G 频段相比，它可以同时工作在 2.4G 和 5G 频段。802.11ax 标准的首要目标之一是将独立网络客户端的无线速度提升 4 倍或者更高，802.11ax 标准在 5GHz 频段上可以带来高达 9.6Gbps 的 Wi-Fi 连接速度。

## 3.4.4 802.11 协议的频率划分



无线信道是对无线通信中发送端和接收端之间通路的一种形象比喻，对于无线电波而言，它从发送端传送到接收端，其间并没有一个有形的连接，它的传播路径也有可能不只一条，我们为了形象地描述发送端与接收端之间的工作，可以想象两者之间有一个看不见的道路衔接，把这条衔接通路称为信道，无线信道也就是常说的无线的“频段（Channel）”。

无线信号就是电磁波，无线电波无处不在，如果随意使用频谱资源，那将带来无穷无尽的干扰问题，所以无线通信协议除了要定义出允许使用的频段，还要精确划分出频率范围，每个频率范围就是信道。

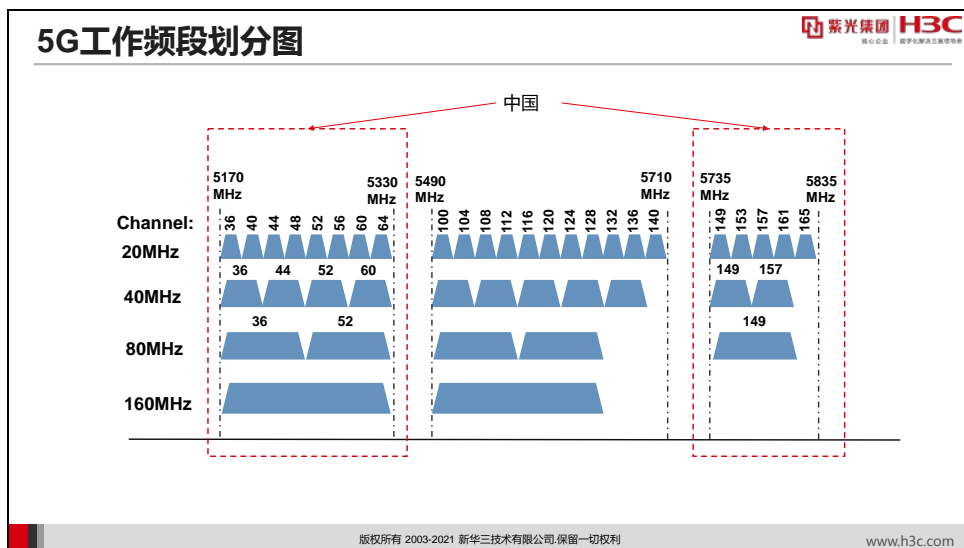
IEEE 802.11a 定义的频段包括 5.15~5.35GHz、5.50~5.70GHz 和 5.725~5.85GHz；而 IEEE 802.11b/g 定义的频段为 2.4~2.4835GHz。

802.11 协议在 2.4GHz 频段定义了 14 个信道，每个信道的频宽为 22MHz。两个信道中心频率之间间隔为 5MHz。信道 1 的中心频率为 2.412GHz，信道 2 的中心频率为 2.417GHz，依此类推至位于 2.472GHz 的信道 13。信道 14 是特别针对日本所定义的，其中心频率与信道 13 的中心频率相差 12MHz。

从上图可以看到，信道 1 在频谱上与信道 2、3、4、5 都有交叠的地方，这就意味着，如果有两个无线设备同时工作，且它们工作的信道分别为 1 和 3，则它们发送的信号会互相干扰。

为了最大程度地利用频段资源，可以使用 1、6、11，2、7、12，3、8、13，4、9、14 这四组互相不干扰的信道来进行无线覆盖。

由于只有部分国家开放了 12~14 信道频段，所以一般情况下都使用 1、6、11 三个信道。



目前在 5G 频段中国大陆主要使用 5.2G 和 5.8G 频段，在频宽为 20M 的情况下，5.8G 可用信道为：149、153、157、161、165，5.2G 的可用信道：36、40、44、48、52、56、60、64(由于国家使用雷达环境中会与 52、56、60、64 信道冲突，因此常规模式下建议避开这些雷达信道，以免出现无线终端接入问题)。

频宽可以设置为 20M、40M、80M、160M，如果是 40M，以 36 和 40 为例，就是 36 和 40 这两个 20M 的信道绑定成 40M 的，对外呈现配置使用的信道为 36。其他的以此类推。同时由于频宽的扩大，会导致不重叠的非干扰的可用信道也相应的变少。因此要根据实际业务需求，合理划分信道和频宽。

## 3.4.5 无线覆盖原则

### 无线覆盖原则 - 蜂窝式覆盖

- 任意相邻区域使用无频率交叉的频道，如1、6、11频道
- 适当调整发射功率，避免跨区域同频干扰
- 蜂窝式无线覆盖实现无交叉频率重复使用

紫光集团 H3C  
核心企业 | 数字基础设施领导者

版权所有 2003-2021 新华三技术有限公司. 保留一切权利 www.h3c.com

可以在二维平面上使用 1、6、11 三个信道实现任意区域无相同信道干扰的无线布网。当某个无线设备功率过大时，会出现部分区域有同频干扰，这时可以通过调整无线设备的发射功率来避免这种情况的发生。但是在三维平面上，要想在实际应用场景中实现任意区域无同频干扰几乎是不可能的。

### 多楼层信道立体覆盖

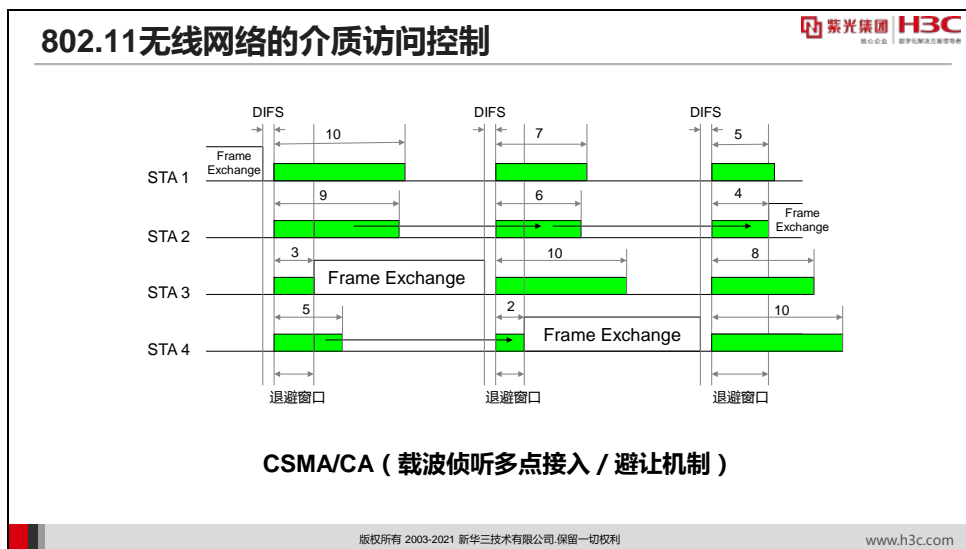
紫光集团 H3C  
核心企业 | 数字基础设施领导者

- 在多层无线覆盖时，信道设置要着眼三维空间的考量，依然采用蜂窝式进行立体频点规划，避免空间信号干扰。

www.h3c.com

在多层无线覆盖时，信道的设置要考虑三维空间的信号干扰。例如在 1 楼部署 3 个 AP，从左到右的信道分别是 1/6/11，此时在 2 楼部署的 3 个 AP 的信道就应该划分为 11/1/6，同理 3 楼为 6/11/1。这样就可最大程度地避免了楼层间的干扰，无论是水平方向还是垂直方向都按照无线蜂窝式覆盖原则进行部署。由于 2.4G 穿透能力较强所以本原则对于 2.4G 频段信道规划较为重要。

## 3.4.6 802.11 无线网络的介质访问控制



总线型局域网在 MAC 层的标准协议是 CSMA/CD (Carrier Sense Multiple Access with Collision Detection, 载波侦听多路访问/冲突检测)。但由于无线产品的适配器不易检测信道是否存在冲突, 因此 802.11 定义了一种全新的协议, 即 CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance, 载波侦听多路访问/冲突避免)。

CSMA/CA 一方面进行载波侦听, 以查看介质是否空闲; 另一方面通过随机的时间等待, 使信号冲突发生的概率减到最小, 以避免冲突。当侦听到介质空闲时, 优先发送。

不仅如此, 802.11 还在 MAC 层提供了确认机制。即采用 ACK 帧对数据帧进行确认。如果发送者没有收到 ACK 帧, 就会重传数据帧。所这种确认机制能够提高数据传输的可靠性, 并提供快速的恢复能力。


上图描述了一个 WLAN 网络里无线工作站的工作机制:

- 1) **侦听介质:** 在工作站要发送数据帧之前, 都会侦听介质是否空闲, 若检测到介质忙, 则继续侦听。
- 2) **固定帧间隔时长:** 当工作站检测到介质空闲时, 会继续侦听一个 DIFS (DCF interframe space, DCF 帧间隔) 时长, 以保证基本的空闲时间。
- 3) **启动定时器:** 当工作站检测到空闲时间达到了 DIFS 时长后, 会启动一个 backoff 计时器, 进行倒计时。该计时器的初始值为一个随机整数, 其取值处于 0 与 CW (contention window, 竞争窗口) 之间。计数器每过一个固定的 SlotTime 即减 1。
- 4) **发送与重传:** 工作站完成 backoff 倒计时后就会发送帧。如果发送失败需要重传, 工作站仍会重复上述过程, 且 CW 的尺寸会随着重传次数递增。如果发送成功或达到重传次数上限, 工作站会重置 CW, 将 CW 的尺寸恢复到初始值。这种机制的目的是保证各个工作站的转发机会平衡。

- 5) **其他终端状态：**在 **backoff** 计时器减到零之前，如果信道上有其他工作站开始发送数据，即本端检测到介质忙，则计时器暂停。如果工作站仍要发送数据，则在介质下次空闲后，仍需继续等待 **DIFS** 时长和 **backoff** 计时器时间，不过 **backoff** 计时器的值不再随机分配，而是继续上次的计数，直至零为止。



## 3.5 本章总结



紫光集团 H3C  
新华三集团 新华三集团

### 课程总结

- 局域网技术类型众多，其中以太网应用最广泛
- 局域网划分为LLC子层和MAC子层
- 千兆以太网，万兆以太网，超高速以太网大大提高了以太网速度
- 802.11规定的WLAN技术允许更便捷地部署局域网

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

## 第4章 IP 基本原理

TCP/IP 协议栈的网络层位于网络接口层和传输层之间，其主要协议包括 IP（Internet Protocol，互联网协议）、ARP（Address Resolution Protocol，地址解析协议）、RARP（Reverse Address Resolution Protocol，反向地址解析协议）、ICMP（Internet Control Message Protocol，互联网控制消息协议）、IGMP（Internet Group Management Protocol，互联网组管理协议）等。其中 IP 是整个网络层的核心协议。

### 4.1 本章目标

  
紫光集团 H3C  
新华三集团 中国领先的ICT解决方案提供商

### 课程目标

学习完本课程，您应该能够：

- 掌握IP地址的格式、分类和子网掩码
- 掌握路由基本概念和相关路由协议简介
- 掌握网络层ARP协议和RARP协议的工作原理
- 掌握IP寻址的基本原理

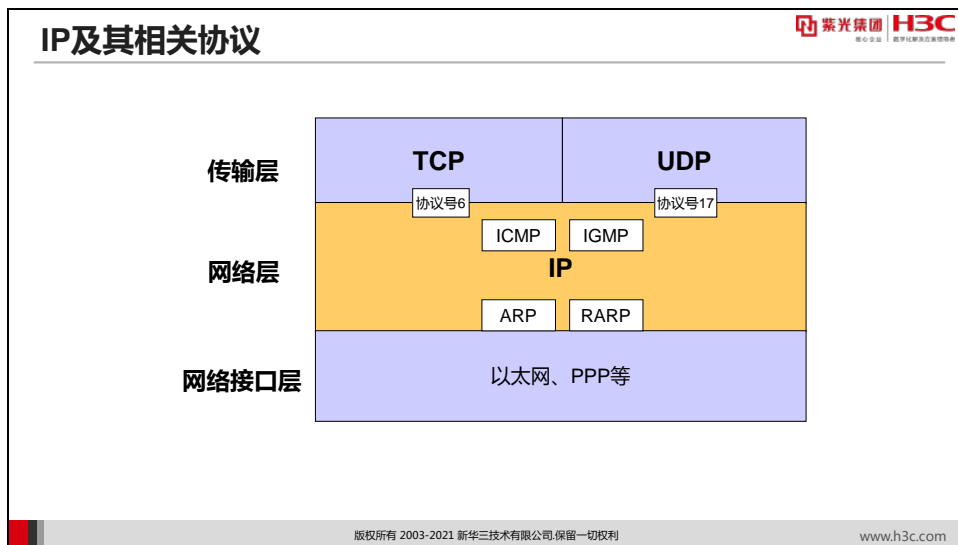


版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

## 4.2 IP协议概述

### 4.2.1 IP 及其相关协议




TCP/IP 协议栈的网络层位于网络接口层和传输层之间。网络层的主要功能是标识大规模网络中的每一个节点，并将数据包投递到正确的目的节点。

TCP/IP 的网络层主要定义了以下协议：

- IP (Internet Protocol, 互联网协议)：负责网络层寻址、路由选择、分段及包重组。
- ARP (Address Resolution Protocol, 地址解析协议)：负责把网络层地址解析成物理地址。
- RARP (Reverse Address Resolution Protocol, 反向地址解析协议)：负责把物理地址解析成网络层地址。
- ICMP (Internet Control Message Protocol, 互联网控制消息协议)：定义了网络层控制和传递消息的功能，可以报告 IP 数据包传递过程中发生的错误、失败等信息，提供网络诊断功能。
- IGMP (Internet Group Management Protocol, 互联网组管理协议)：负责管理 IP 组播组。

## IP的主要作用

- **标识节点和链路**
  - 用唯一的IP地址标识每一个节点
  - 用唯一的IP网络号标识每一个链路
- **寻址和转发**
  - 确定节点所在网络的位置，进而确定节点所在的位置
  - IP路由器选择适当的路径将IP包转发到目的节点
- **适应各种数据链路**
  - 根据链路的MTU对IP包进行分片和重组
  - 为了通过实际的数据链路传递信息，须建立IP地址到数据链路层地址的映射



紫光集团 H3C  
核心企业 新华三集团成员企业

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com

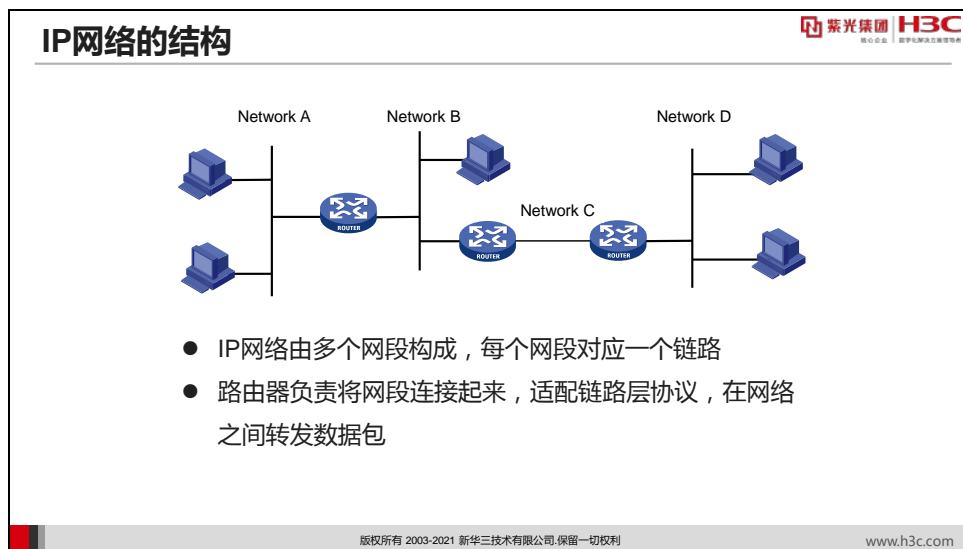
TCP/IP 网络层的核心协议是由 RFC 791 定义的 IP（Internet Protocol，互联网协议）。IP 是尽力传输的网络协议，其提供的数据传送服务是不可靠的、无连接的。IP 协议不关心数据包载荷的内容，不能保证数据包能成功地到达目的地，也不维护任何关于前后数据包的状态信息。面向连接的可靠服务由上层的 TCP 协议实现。

IP 将来自传输层的数据段封装成 IP 包并交给网络接口层进行发送，同时将来自网络接口层的帧解封装并根据 IP 协议号（Protocol Number）提交给相应的传输层协议进行处理。TCP（Transmission Control Protocol，传输控制协议）的 IP 协议号为 6，UDP（User Datagram Protocol，用户数据报协议）的 IP 协议号为 17。

IP 协议的主要作用包括：

- **标识节点和链路：**IP 为每个链路分配一个全局唯一的网络号（network-number）以标识每个网络；为节点分配一个全局唯一的 32 位 IP 地址，用以标识每一个节点。
- **寻址和转发：**IP 路由器（router）根据所掌握的路由信息，确定节点所在网络的位置，进而确定节点所在的位置，并选择适当的路径将 IP 包转发到目的节点。
- **适应各种数据链路：**为了工作于多样化的链路和介质上，IP 必须具备适应各种链路的能力，例如可以根据链路的 MTU（Maximum Transfer Unit，最大传输单元）对 IP 包进行分片和重组，可以建立 IP 地址到数据链路层地址的映射以通过实际的数据链路传递信息。

## 4.2.2 IP 网络结构



典型的 IP 互联网由众多的路由器和网段（network segment）构成。每个网段对应一个链路。路由器在这些网段之间执行数据转发服务。

路由器的主要功能包括：

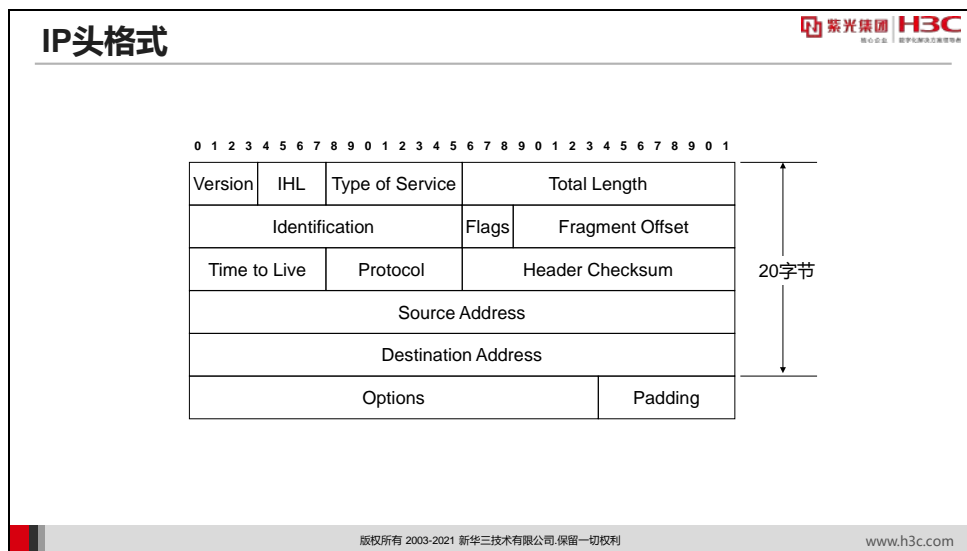
- 连接分离的网络：路由器的每个接口处于一个网络，将原本孤立的网络连接起来，实现大范围的网络通信。
- 链路层协议适配：由于链路层协议的多样性，不同类的链路之间不能直接通信。路由器可以适配各种数据链路的协议和速率，使其间的通信成为可能。
- 在网络之间转发数据包：为了实现这个功能，路由器之间需要运行网关到网关协议（Gateway to Gateway Protocol, GGP）交换路由信息和其他控制信息，以了解去往每个目的网络的正确路径，典型的 GGP 包括 RIP、OSPF、BGP 等路由协议（Routing Protocol）。

IP 网络的包转发是逐跳（hop-by-hop）进行的。即包括路由器在内的每一个节点要么将一个数据包直接发送给目的节点，要么将其发送给到目的节点路径上的下一跳（next hop）节点，由下一跳继续将数据包转发下去。数据包必须历经所有的中间节点之后才能到达目的。每一个路由器或主机的转发决策都是独立的，其依据是存储于自身路由表（routing table）中的路由（route）。

**注意：**

早期的 Internet 术语将路由器（router）称为网关（gateway），故而在探讨基本的 IP 通信时，这两个术语是不加区分的。

## 4.2.3 IP 封装

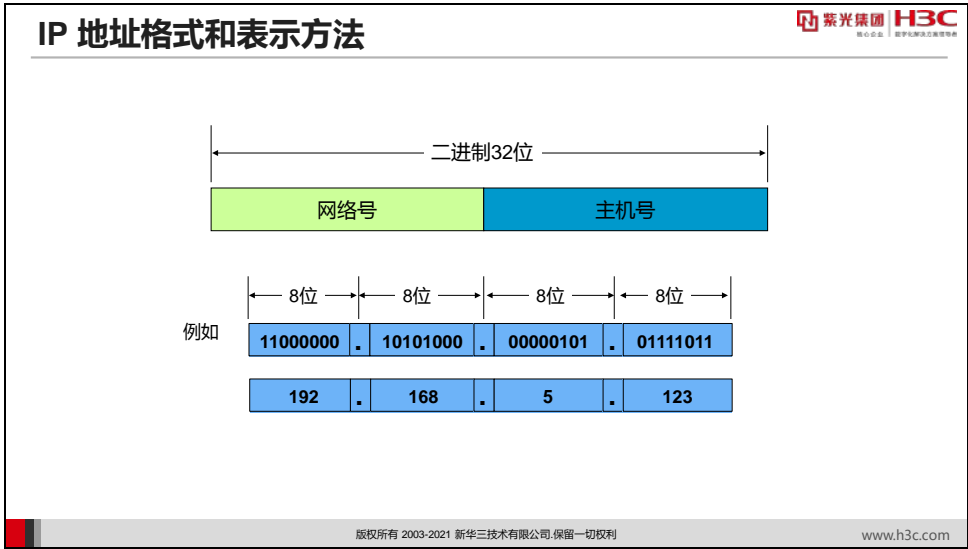


IP 头选项不经常使用，因此普通的 IP 头部长度的 20 字节。其中一些主要字段简介如下：

- **版本（Version）**：标明了 IP 协议的版本号，目前的协议版本号为 4。下一代 IP 协议的版本号为 6。
- **头长度（Internet Header Length, IHL）**：指 IP 包头部长度，占 4 位。
- **服务类型（Type of Service, ToS）**：用于标志 IP 包期望获得的服务等级，常用于 QoS（Quality of Service，服务质量）中。
- **总长度（Total Length）**：整个 IP 包的长度，包括数据部分。
- **标识（Identification）**：唯一地标识主机发送的每一个 IP 包。通常每发送一个包其值就会加 1。
- **生存时间（Time to Live, TTL）**：设置了数据包可以经过的路由器数目。一旦经过一个路由器，TTL 值就会减 1，当该字段值为 0 时，数据包将被丢弃。
- **协议（Protocol）**：标识数据包内传送的数据所属的上层协议，IP 用协议号区分上层协议。TCP 协议的协议号为 6，UDP 协议的协议号为 17。
- **头校验和（Head Checksum）**：IP 头部的校验和，用于检查包头的完整性。
- **源地址（Source Address）和目的地址（Destination Address）**：分别标识数据包的源节点和目的节点的 IP 地址。

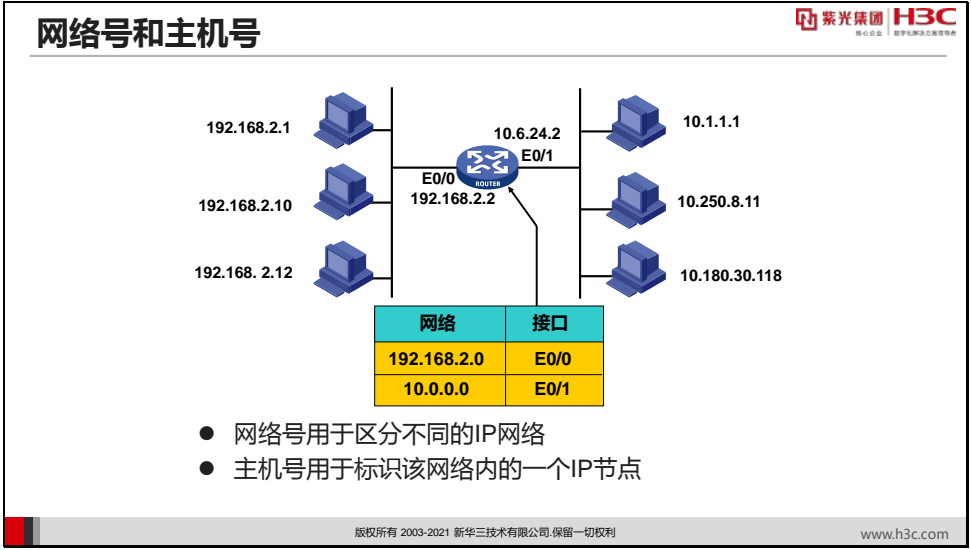
# 4.3 IP地址和地址映射

## 4.3.1 IP 地址格式和表示方法



连接到 Internet 上的设备必须有一个全球唯一的 IP 地址（IP Address）。IP 地址长度为二进制 32 位，通常采用点分十进制方式表示，即每个 IP 地址被表示为以小数点隔开的 4 个十进制整数，每个整数对应一个字节，如 192.168.5.123。

IP 地址与链路类型、设备硬件无关，而是由管理员分配指定的，因此也称为逻辑地址（Logical Address）。每台主机可以拥有多个网络接口卡，也可以同时拥有多个 IP 地址。路由器也可以看作这种主机，但其每个 IP 接口必须处于不同的 IP 网络，即各个接口的 IP 地址分别处于不同的 IP 网段。



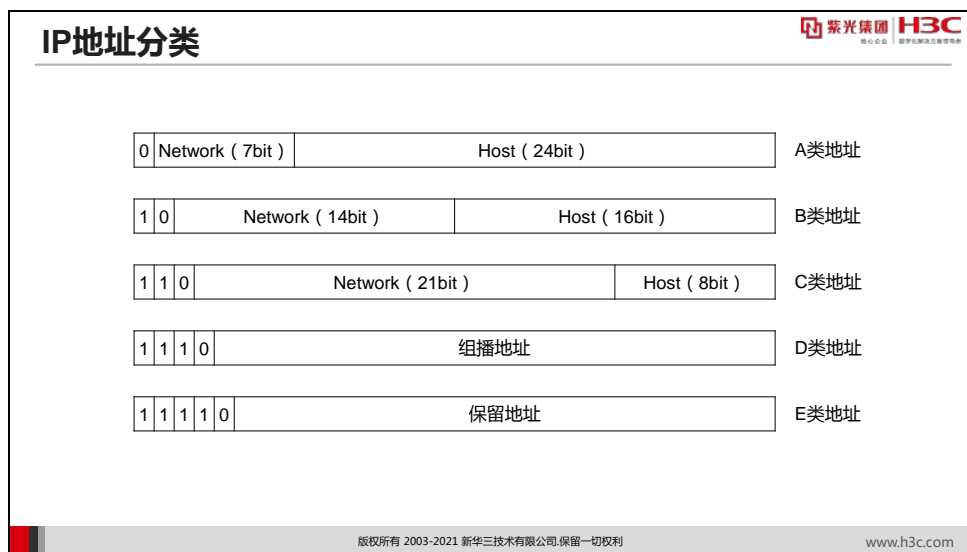
由于理论上总共有  $2^{32}$  个 IP 地址，也就是约 43 亿个 IP 地址，在互联网上，每一台路由器都储存每一个节点的路由信息几乎是不可能的。为便于实现路由选择、地址分配和管理维护，IP 地址采用二级结构，即 IP 地址由两个部分组成：

- 网络号（network-number）：用于区分不同的 IP 网络，即该 IP 地址所属的 IP 网段。一个网络中所有设备的 IP 地址具有相同的网络号。
- 主机号（host-number）：用于标识该网络内的一个 IP 节点。在一个网段内部，主机号是唯一的。

这样，路由器只需要储存每个网段的路由信息即可。



## 4.3.2 IP 地址分类



各个网段内具有的 IP 节点数各不相同，为了适应这种需求，IP 地址被分成五类：

- A 类 IP 地址的第一个八位段（octet）以 0 开始。A 类地址的网络号为第一个八位段，网络号取值范围为 1~126（127 留作它用）。A 类地址的主机号为后面的三个八位段，共 24 位。A 类地址的范围为 1.0.0.0~126.255.255.255，每个 A 类网络有  $2^{24}$  个 A 类 IP 地址。
- B 类 IP 地址的第一个八位段以 10 开始。B 类地址的网络号为前两个八位段，网络号的第一个八位段取值为 128~191。B 类地址的主机号为后面的二个八位段共 16 位。B 类地址的范围为 128.0.0.0~191.255.255.255，每个 B 类网络有  $2^{16}$  个 B 类 IP 地址。
- C 类 IP 地址的第一个八位段以 110 开始。C 类地址的网络号为前三个八位段，网络号的第一个八位段取值为 192~223。C 类地址的主机号为后面的一个八位段共 8 位。C 类地址的范围为 192.0.0.0~223.255.255.255，每个 C 类网络有  $2^8=256$  个 C 类 IP 地址。
- D 类地址第一个八位段以 1110 开头，因此 D 类地址的第一个八位段取值为 224~239。D 类地址通常为组播地址。
- E 类地址第一个八位段以 11110 开头，保留用于研究。

## 4.3.3 特殊的 IP 地址

特殊的IP地址		
网络号	主机号	地址类型和用途
Any	全0	网络地址，代表特定网段
Any	全1	网段广播地址，代表特定网段的所有节点
127	Any	环回地址，常用于环回测试
全0		代表所有网络，常用于指定默认路由
全1		全网广播地址，代表所有节点

IP 地址用于唯一的标识一台网络设备，但并不是每一个 IP 地址都用于这个目的。一些特殊的 IP 地址被用于各种各样的其他用途。

主机部分全为 0 的 IP 地址称为网络地址（Network Address）。网络地址用来标识一个网段。例如 1.0.0.0/8、10.0.0.0/8、192.168.1.0/24 等。

主机部分全为 1 的 IP 地址是网段广播地址。这种地址用于标识一个网络内的所有主机。例如，10.255.255.255 是网络 10.0.0.0 内的广播地址，表示网络 10.0.0.0 内的所有主机。一个发往 10.255.255.255 的 IP 包将会被该网段内的所有主机接收。

网络号为 127 的 IP 地址用于环路测试目的。例如 127.0.0.1 通常表示“本机”。

IP 地址 0.0.0.0 代表“所有的网络”，通常用于指定默认路由。而 IP 地址 255.255.255.255 是全网广播地址，代表“所有的主机”，用于向网络的所有节点发送数据包。

如上所述，每一个网段都会有一个网络地址和一个网段广播地址，因此实际可用于主机的地址数等于网段内的全部地址数减 2。例如 B 类网段 172.16.0.0 有 16 个主机位，因此有  $2^{16}$  个 IP 地址，去掉一个网络地址 172.16.0.0 和一个广播地址 172.16.255.255 不能用于标识主机，实际共有  $2^{16}-2$  个可用地址。

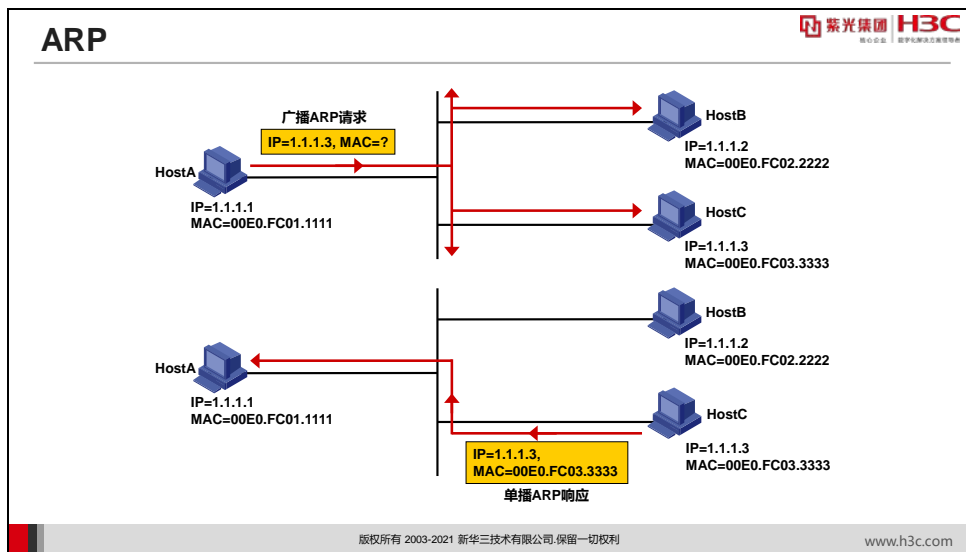
各类 IP 地址的实际可用地址范围如下所示：

- A 类：1.0.0.0～127.255.255.255
- B 类：128.0.0.0～191.255.255.255
- C 类：192.0.0.0～223.255.255.255
- D 类：224.0.0.0～239.255.255.255
- E 类：240.0.0.0～255.255.255.255

**注意：**

转发网段广播和全网广播会对网络性能造成严重的不利影响，因此几乎所有的路由器在默认情况下均不转发广播包。

## 4.3.4 ARP



IP 地址将物理地址对上层隐藏起来,使 Internet 表现出统一的地址格式。但在实际通讯时,IP 地址不能被物理网络所识别,物理网络所使用的依然是物理地址。因此,必须实现 IP 地址对物理地址的映射。

对于以太网而言,当 IP 数据包通过以太网发送时,以太网链路并不识别 32 位的 IP 地址,它们是以 48 位的 MAC 地址标识以太网节点的。因此,必须在 IP 地址与 MAC 地址之间建立映射 (map) 关系,建立这种映射的过程称为地址解析 (Resolution)。

ARP (Address Resolution Protocol, 地址解析协议) 就是用于动态地将 IP 地址解析为 MAC 地址的协议。主机通过 ARP 解析到目的 MAC 地址后,将在自己的 ARP 缓存表中增加相应的 IP 地址到 MAC 地址的映射表项,用于后续到同一目的地报文的转发。

假设 HostA 和 HostB 在同一个网段,HostA 要向 HostB 发送 IP 包,其地址解析过程如下:

- 1) HostA 首先查看自己的 ARP 表,确定其中是否包含有 HostB 的 IP 地址对应的 ARP 表项。如果找到了对应的表项,则 HostA 直接利用 ARP 表项中的 MAC 地址对 IP 数据包封装成帧,并将帧发送给 HostB。
- 2) 如果 HostA 在 ARP 表中找不到对应的表项,则暂时缓存该数据包,然后以广播方式发送一个 ARP 请求。ARP 请求报文中的发送端 IP 地址和发送端 MAC 地址为 HostA 的 IP 地址和 MAC 地址,目标 IP 地址为 HostB 的 IP 地址,目标 MAC 地址为全 0 的 MAC 地址。

- 3) 由于 ARP 请求报文以广播方式发送，该网段上的所有主机都可以接收到该请求。HostB 比较自己的 IP 地址和 ARP 请求报文中的目标 IP 地址，由于两者相同，HostB 将 ARP 请求报文中的发送端(即 HostA)IP 地址和 MAC 地址存入自己的 ARP 表中，并以单播方式向 HostA 发送 ARP 响应，其中包含了自己的 MAC 地址。其他主机发现请求的 IP 地址并非自己，于是都不做应答。
- 4) HostA 收到 ARP 响应报文后，将 HostB 的 MAC 地址加入到自己的 ARP 表中，同时将 IP 数据包用此 MAC 地址为目的地址封装成帧并发送给 HostB。

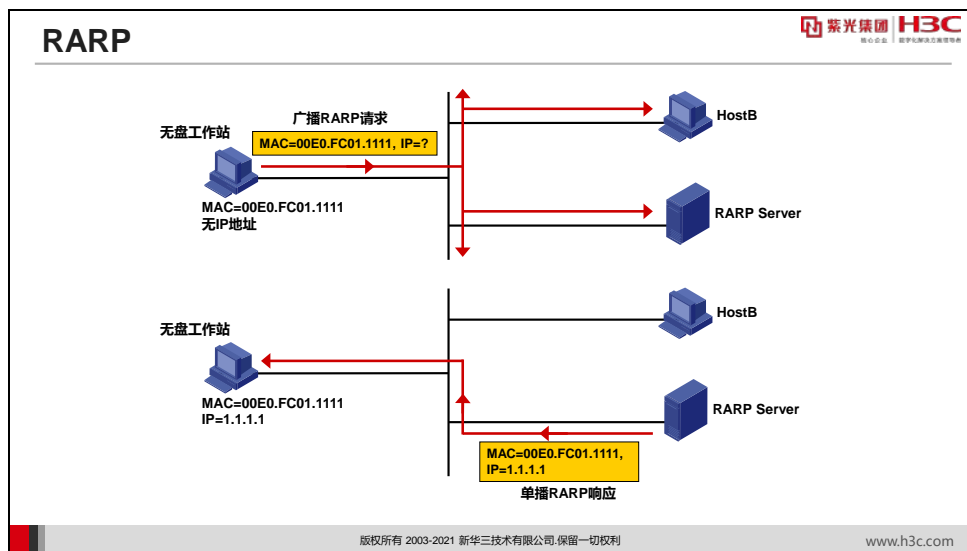
ARP 表项分为动态 ARP 表项和静态 ARP 表项：

- 动态 ARP 表项由 ARP 协议动态解析获得，如果超过一个老化时间（aging time）未被使用，则会被自动删除。
- 静态 ARP 表项通过管理员手工配置，不会被老化。静态 ARP 表项的优先级高于动态 ARP 表项，可以将相应的动态 ARP 表项覆盖。

此外，还有一种特殊的 ARP 应用——免费 ARP（Gratuitous ARP）。免费 ARP 协议包中携带的发送者 IP 地址和目标 IP 地址都是本机 IP 地址，发送者 MAC 地址是本机 MAC 地址，目标 MAC 地址是广播地址。对外发送免费 ARP 协议包可以实现以下功能：

- 确定其它设备的 IP 地址是否与本机 IP 地址冲突。
- 设备改变了硬件地址，通过发送免费 ARP 报文通知其他设备更新 ARP 表项。

## 4.3.5 RARP



主机只知道自己的硬件地址时，可以通过 RARP（Reverse Address Resolution Protocol，反向地址解析协议）解析自己的 IP 地址。RARP 常用于无盘工作站启动前获取自身 IP 地址。

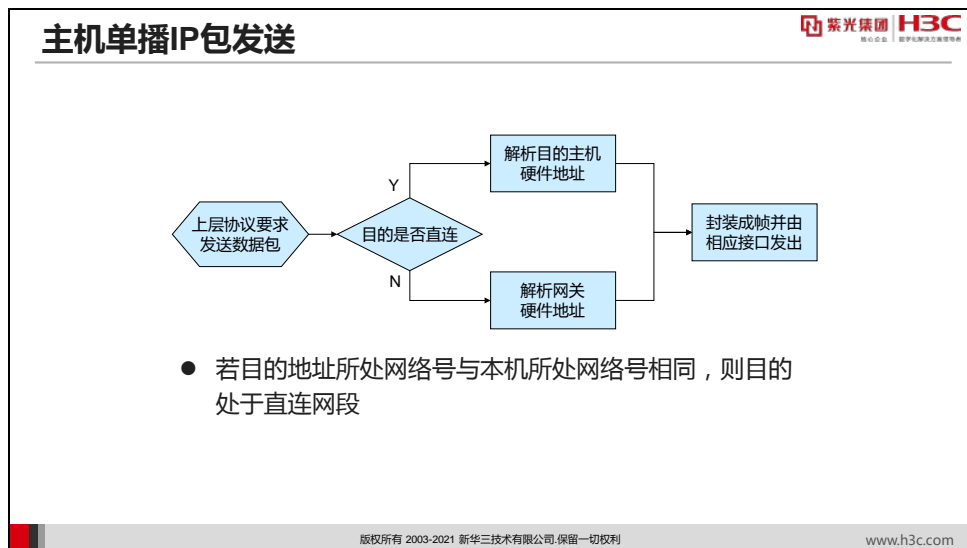
在刚刚启动时，无盘工作站只知道自己网卡的 MAC 地址，需要获得自己的 IP 地址，于是向网络中广播 RARP 请求。RARP 服务器接收广播请求后发送应答报文，无盘工作站随即获得 IP 地址。

RARP 服务器要响应请求，首先必须知道物理地址与 IP 地址的对应关系。为此，在 RARP 服务器中维持着一个本网段的“物理地址—IP 地址”映射表。当某无盘工作站发出 RARP 请求后，网上所有主机均收到该请求，但只有 RARP 服务器处理请求并根据请求者物理地址响应请求。无盘工作站发出的 RARP 请求中携带其物理地址，服务器根据此硬件地址查找其 IP 地址。由于服务器此时已经知道无盘工作站的物理地址，因此不再采用广播方式，而是直接向无盘工作站发送单播应答。

对应于 ARP、RARP 请求以广播方式发送，ARP、RARP 应答一般以单播方式发送，以节省网络资源。

## 4.4 IP包转发

### 4.4.1 主机单播 IP 包发送

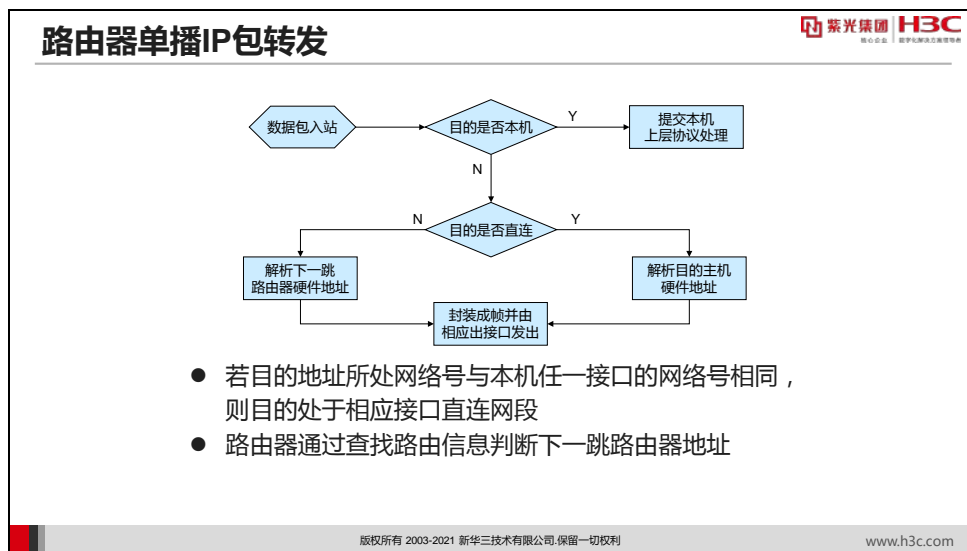


主机在发送 IP 包之前，首先需判断目的主机所处的位置。主机对比自身 IP 地址的网络地址与目的 IP 地址的网络地址，如果二者相等，则可知目的主机与自己处于同一网段；如果二者不相等，则目的主机与自己处于不同网段。

如果目的与本机处于同一网段，主机可以与其直接通信。此时主机首先解析目的主机 IP 地址对应的硬件地址，随即将 IP 包以此硬件地址为目的地址封装成帧，由直接连接此网段的接口发送给目的主机。

如果目的与本机处于不同网段，则主机需将 IP 包交给一台称为默认网关(default gateway)的路由器，由此路由器设法将 IP 包转发给目的主机。此时主机根据默认网关的 IP 地址解析出默认网关的硬件地址，随即将 IP 包以此硬件地址为目的地址封装成帧，由直接连接此网段的接口发送给默认网关。

## 4.4.2 路由器单播 IP 包转发

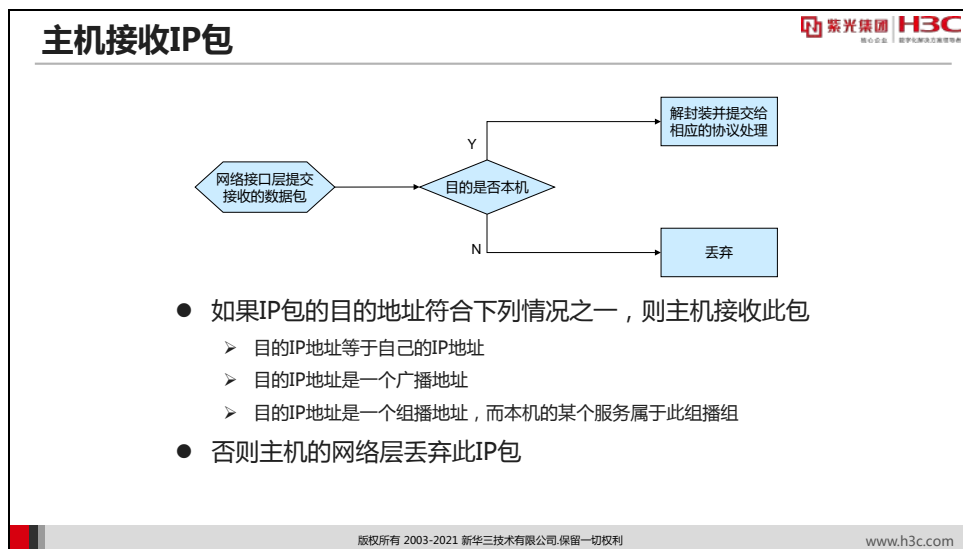


路由器收到一个 IP 包后，首先检测其目的地址。如果目的地址为本机，则接收此包并将其解封装，所得数据提交上层协议处理。

如果此 IP 包目的地址并非本机，而处于某个接口直连的网段，路由器可以与其直接通信。此时路由器首先解析目的主机 IP 地址对应的硬件地址，随即将 IP 包以此硬件地址为目的地址封装成帧，由直接连接此网段的接口发送给目的主机。

如果目的与路由器处于不同网段，则路由器需将 IP 包交给下一跳（next hop）路由器，由下一跳设法将 IP 包转发给目的主机。此时主机根据路由表中的路由信息查出下一跳的 IP 地址，解析出下一跳的硬件地址，随即将 IP 包以此硬件地址为目的地址封装成帧，由直接连接此网段的接口发送给下一跳路由器。

## 4.4.3 主机接收 IP 包



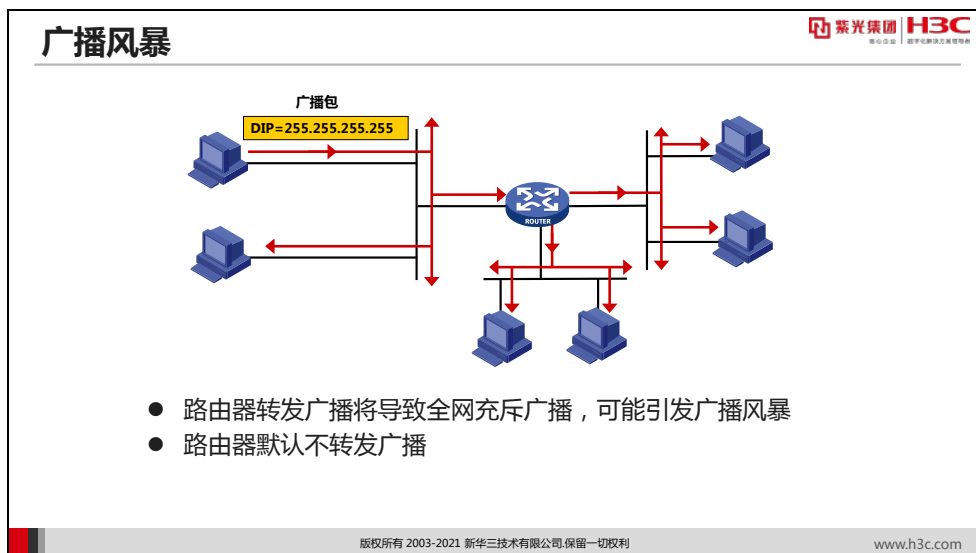
在收到网络接口层提交的 IP 数据包时，主机首先检查这个包的目的地地址是否等于自身 IP 地址。如果其目的地地址符合下列情况之一，则主机接收此包，并将其数据提交相应的上层协议处理：

- 这个包的目的 IP 地址等于自身 IP 地址；
- 这个包的目的 IP 地址是一个广播地址；
- 这个包的目的 IP 地址是一个组播地址，而本机的某个服务正好属于此组播组。

如果此 IP 包的目的地地址不符合上述任何一种情况，则主机的网络层丢弃此 IP 包。



## 4.4.4 广播风暴

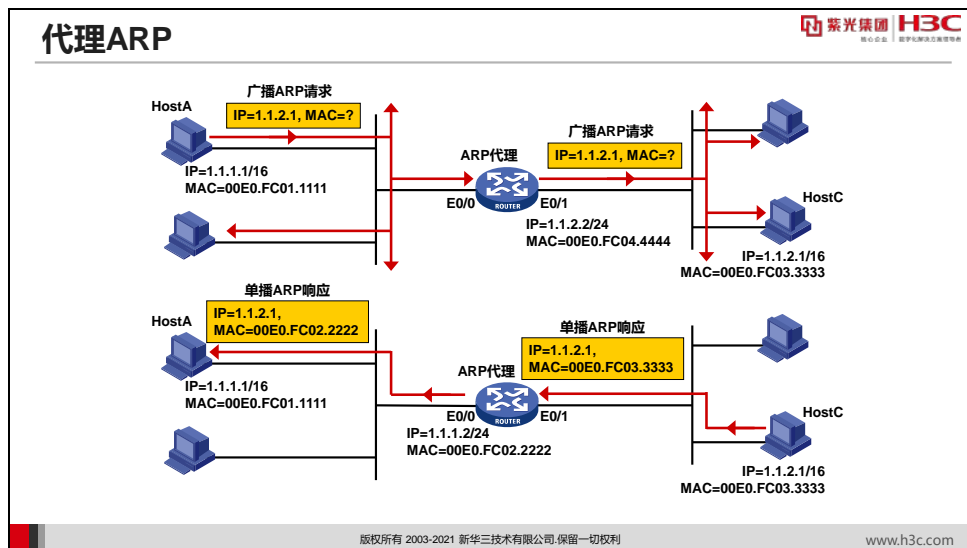


路由器转发广播将导致全网充斥广播。很多协议需要通过广播包完成公告、发现等任务。以 ARP 为例，每台主机对网段内其他任何主机通讯时都需要广播 ARP 请求。如果路由器转发广播包，每个广播包将传遍整个互联网，大大浪费了网络资源。并且由于广播包将会被提交到每一台主机的网络层进行处理，每一台主机的资源都会遭到无谓的浪费。这种情况发展到一定程度，整个网络会由于广播而瘫痪，这种情况称为广播风暴（Broadcast Storm）。

为了避免广播风暴的发生，路由器在默认情况下不转发广播包。

## 4.5 其他相关协议介绍

### 4.5.1 代理 ARP

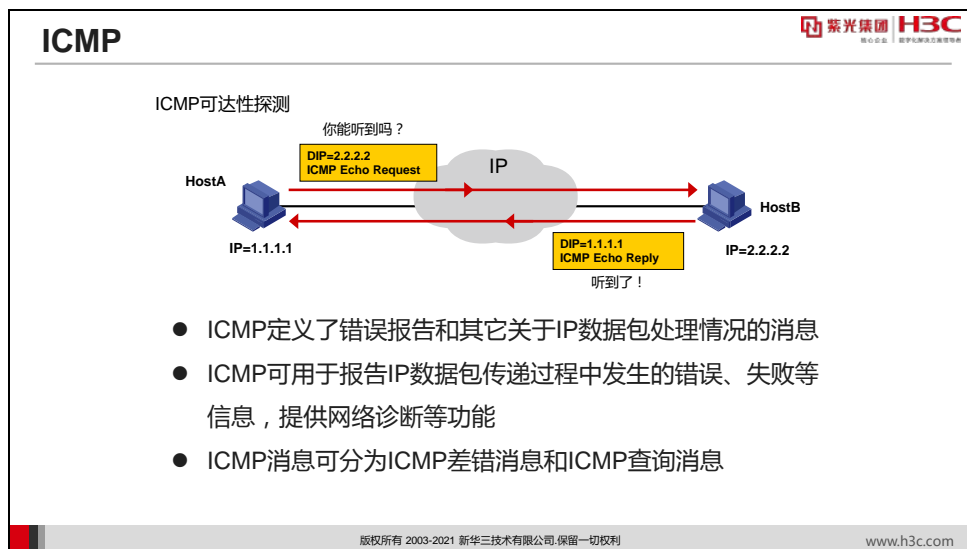


当主机不了解网关的信息，或主机无法判断目的是否处于本网段时，某些主机会对处于其他网段的目的主机 IP 地址直接进行 ARP 解析。此时，路由器可以运行代理 ARP (Proxy ARP) 协助主机实现通信。

如图所示，HostA 希望与 HostC 通信，但由于某种原因，HostA 直接发送了 ARP 请求，解析 HostC 的 MAC 地址。运行了代理 ARP 的路由器收到 ARP 请求后，代理 HostA 在 1.1.2.0 网段发出 ARP 请求，解析 HostC 的 MAC 地址。

HostC 认为路由器向其发出了 ARP 请求，遂回应以 ARP 响应，通告自己的 MAC 地址 00E0.FC03.3333。路由器收到 ARP 响应后，也向 HostA 发送 ARP 响应，但通告的 MAC 地址是其连接到 1.1.1.0 网络的以太网 E0/0 的 MAC 地址 00E0.FC02.2222。这样在 HostA 的 ARP 表中会形成 IP 地址 1.1.2.1 与 MAC 地址 00E0.FC02.2222 的映射项，因此 HostA 实际上会将所有要发给 HostC 的数据包发送到路由器上，再由路由器转发给 HostC。

## 4.5.2 ICMP



RFC 792 定义的 ICMP（Internet Control Message Protocol，互联网控制消息协议）是一个网络层协议，基于 IP 运行。ICMP 定义了错误报告和其它回送给源点的关于 IP 数据包处理情况的消息，可以用于报告 IP 数据包传递过程中发生的错误、失败等信息，提供网络诊断等功能。

ICMP 通常为 IP 层或者更高层协议使用。其中 ping 是一个最常见的应用，主机可通过它来测试网络的可达性。用户运行 ping 命令时，主机向目的主机发送 ICMP Echo Request 消息。Echo Request 消息封装在 IP 包内，其目的地址为目的主机的 IP 地址。目的主机收到 Echo Request 消息后，向源主机回送一个 ICMP Echo Reply 消息。源主机如果收到 Echo Reply 消息，即可获知该目的主机是可达的。假定某个中间路由器没有到达目的网络的路由，便会向源主机端返回一条 ICMP Destination Unreachable 消息，告知源主机目的不可达。

ICMP 消息可分为两种类型，即 ICMP 差错消息和 ICMP 查询消息。对于 ICMP 差错消息要作特殊处理，例如，在对 ICMP 差错消息进行响应时，永远不会生成另一份 ICMP 差错消息。（如果没有这个限制规则，可能会遇到一个差错产生另一个差错的情况，而差错再产生差错，这样会无休止地循环下去）。

类型字段的值	ICMP 消息的类型	差错消息	查询消息
0	Echo Reply		✓
3	Destination Unreachable	✓	
4	Source Quench	✓	
5	Redirect	✓	
8	Echo Request		✓
11	Time Exceeded	✓	

类型字段的值	ICMP 消息的类型	差错消息	查询消息
12	Parameter Problem	✓	
13	Timestamp Request		✓
14	Timestamp Reply		✓
15	Information Request		✓
16	Information Reply		✓

常用 ICMP 消息的含义如下：

- **目的不可达 (Destination Unreachable)**：目的主机可能不存在或已关机，可能发送者提供的源路由要求无法实现，或设定了不分段的包太大而不能封装于帧中。在这些情况下，路由器检测出错误，并向源发送者发送一个 **ICMP Destination Unreachable** 消息。它包含了不能到达目的地的数据包의完整 IP 头，以及其载荷数据的前 64 比特，这样发送者就能知道哪个包无法投递。
- **回波请求 (Echo Request)**：是由主机或路由器向一个特定的目的主机发出的询问。这种询问消息用来测试目的站是否可达。
- **回波响应 (Echo Reply)**：对回波请求作出响应时发送。收到 **Echo Request** 的主机对源主机发送 **ICMP Echo Reply** 消息作为响应。
- **参数问题 (Parameter Problem)**：假设一个 IP 包的头中产生错误或非法值，路由器发现问题后向源发送一个 **Parameter Problem** 消息。这个消息包含了有问题的 IP 头和一个指向出错字段的指针。
- **重定向 (Redirect)**：假设主机向路由器发送了一个包，而此路由器知道其他一些路由器能将分组更快地投递，为了方便以后路由，此路由器向主机发送一个 **Redirect** 消息。它通知主机其他路由器的位置，以及今后应当将具有相同目的地址的包发向那里。这就允许主机动态地更新它的路由表，更好地适应网络条件的变化。
- **源抑制 (Source Quench)**：当某个速率较高的源主机向另一个速率较慢的目的主机（或路由器）发送一连串的数据包时，就有可能使速率较慢的目的主机产生拥塞，因而不但不丢弃一些数据包。源主机通过高层协议得知丢失了一些数据包，就会不断地重发这些数据包，这就使得原本已经拥塞的目的主机更加拥塞。在这种情况下，目的主机就要向源主机发送 **ICMP Source Quench** 消息，使源站暂停发送。
- **超时 (Time Exceeded)**：当 IP 包中的 TTL 字段减到 0 或分片重组定时器到期时，此包或任何未重组的分片将从网络中被删除。删除分组的路由器接着向源发送一个 **Time Exceeded** 消息，说明分组未被投递。
- **时间戳请求和时间戳应答 (Timestamp Request and Timestamp Reply)**：时间戳分组使主机能估计它到另一个主机一次往返通信所需的时间。源主机创建并发送一个含有发送时刻（源时间戳）的 **Timestamp Request** 消息，目的主机收到分组后创建一个含有原时间戳和目的主机接收时间戳以及目的主机传输时间戳的 **Timestamp Reply** 消息。

当源主机收到 **Timestamp Reply** 时，它同时记录分组的到达时刻。这些时间戳使主机能够估计网络的 IP 包投送效率。

## 4.6 本章总结

### 课程总结

- 32位IP地址分为网络号和主机号两部分，用以标识网络和主机
- 主机将跨网段IP包交给默认网关，路由器负责跨网段转发数据包
- ARP协议用于把已知的IP地址解析为MAC地址
- RARP用于在数据链路层地址已知时解析IP地址
- ICMP定义了网络层控制和传递消息的功能

## 第5章 TCP 和 UDP 基本原理

TCP/IP 协议族的传输层协议主要包括 TCP（Transfer Control Protocol，传输控制协议）和 UDP（User Datagram Protocol，用户数据报协议）。TCP 是面向连接的可靠的传输层协议。它支持在并不可靠的网络上实现面向连接的可靠的数据传输。UDP 是无连接的传输协议，主要用于支持在较可靠的链路上的数据传输，或用于对延迟较敏感的应用。

### 5.1 本章目标

  
紫光集团 H3C  
核心企业 | 数字基础设施领导者

### 课程目标

学习完本课程，您应该能够：

- 了解TCP/UDP协议所提供的服务
- 了解TCP/UDP的报文结构
- 掌握TCP建立和拆除的过程
- 掌握TCP的滑动窗口机制
- 掌握TCP的可靠性技术



版权所有 2003-2021 新华三技术有限公司,保留一切权利

www.h3c.com


## 5.2 TCP/IP 传输层的作用

### 传输层的作用

- 提供面向连接或无连接的服务
- 维护连接状态
- 对应用层数据进行分段和封装
- 实现多路复用
- 可靠地传输数据
- 执行流量控制

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com



紫光集团 H3C  
核心企业 | 数字基础设施领导者

TCP/IP 的传输层位于应用层和网络层之间，为终端主机提供端到端的连接。TCP/IP 的传输层有 TCP（Transfer Control Protocol，传输控制协议）和 UDP（User Datagram Protocol，用户数据报协议）两种主要协议。TCP 和 UDP 都基于相同的网络层协议 IP。传输层协议的主要作用包括：

- 提供面向连接或无连接的服务：传输层协议定义了通信两端点之间是否需要建立可靠的连接关系。
- 维护连接状态：如果必须在通信前建立连接关系，传输层协议必须在其数据库中记录这种连接关系，并且通过某种机制维护连接关系，及时发现连接故障等。
- 对应用层数据进行分段和封装：应用层数据往往是大块的或持续的数据流，而网络只能发送长度有限的数据包，传输层协议必须在传输应用层数据之前将其划分成适当尺寸的段（segment），再交给 IP 协议发送。
- 实现多路复用（Multiplexing）：一个 IP 地址可以标识一个主机，一对“源-目的”IP 地址可以标识一对主机的通信关系，而一个主机上却可能同时有多个程序访问网络，因此传输层协议采用端口号（port number）来标识这些上层的应用程序，从而使这些程序可以复用网络通道。
- 可靠地传输数据：数据在跨网络传输过程中可能出现错误、丢失、乱序等种种问题，传输层协议必须能够检测并更正这些问题。
- 执行流量控制（flow control）：当发送方的发送速率超过接收方的接收速率时，或者当资源不足以支持数据的处理时，传输层负责将流量控制在合理的水平；反之，当资源允许时，传输层可以放开流量，使其增加到适当的水平。



## 5.3 TCP 协议基本原理

### 5.3.1 TCP 协议的特点

#### TCP的特点

- 三次握手
  - 建立可靠连接
- 端口号
  - 多路复用
- 完整性校验
  - 差错检测

- 确认机制
  - 应答接收
- 序列号
  - 丢失检测、乱序重排
- 窗口机制
  - 流量控制

紫光集团 H3C  
核心企业 | 数字技术生态领导者

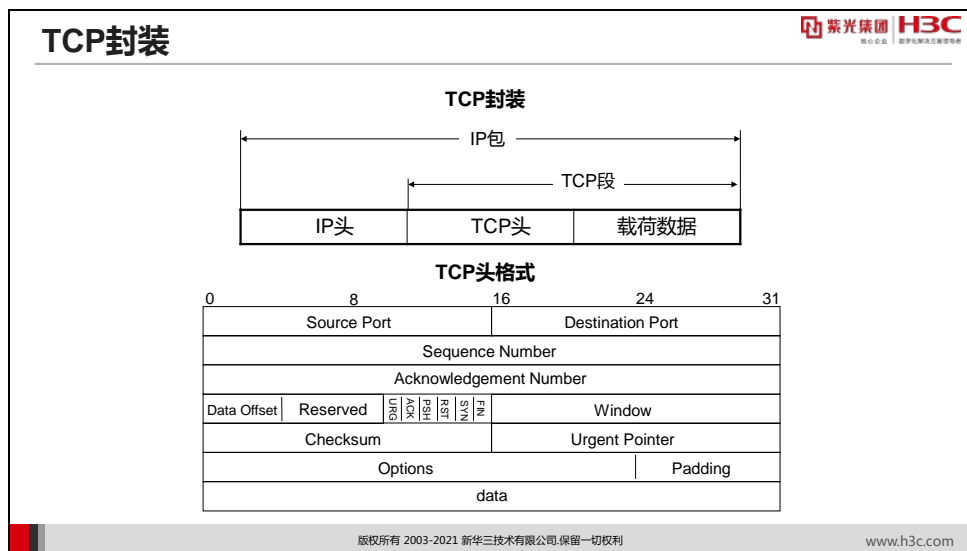
版权所有 2003-2021 新华三技术有限公司, 保留一切权利

www.h3c.com

RFC 793 定义的 TCP（Transmission Control Protocol，传输控制协议）是一种面向连接的、端到端的可靠传输协议。TCP 的主要特点包括：

- 三次握手（Three-Way Handshake）建立连接：确保连接建立的可靠性。
- 端口号：通过端口号标识上层协议和服务，实现了网络通道的多路复用。
- 完整性校验：通过对协议和载荷数据计算校验和（Checksum），保证了接收方能检测出传输过程中可能出现的差错。
- 确认机制：对于正确接收到的数据，接收方通过显式应答通告发送方，超出一定时间之后，发送方将重传没有被确认的段，确保传输的可靠性。
- 序列号：发送的所有数据都拥有唯一的序列号，这样不但唯一标识了每一个段（segment），而且明确了每个段在整个数据流中的位置，接收方可以利用这些信息实现确认、丢失检测、乱序重排等功能。
- 窗口机制：通过可调节的窗口，TCP 接收方可以通告期望的发送速度，从而控制数据的流量。

## 5.3.2 TCP 封装

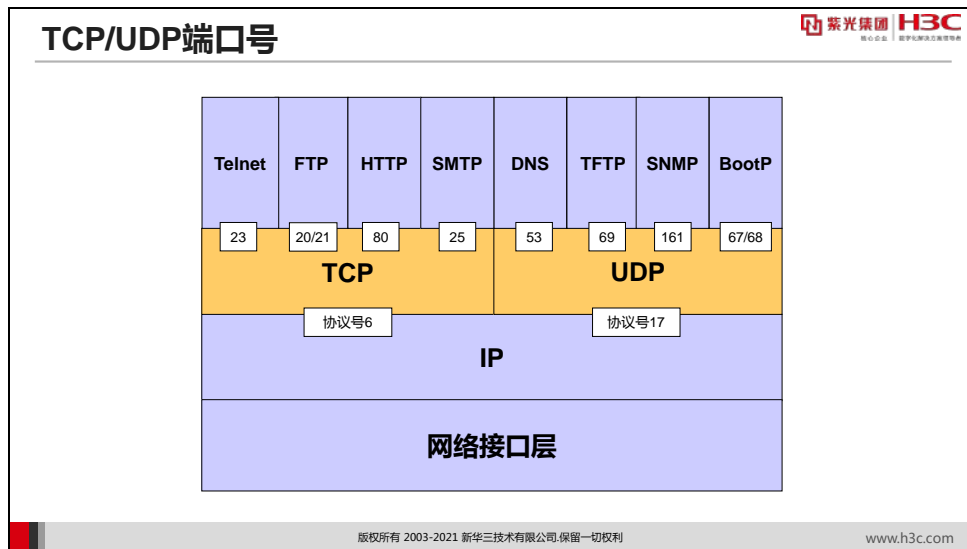


TCP 段的头格式如上图所示，其协议头最少 20 个字节。其中主要字段如下：

- **源端口 (Source Port)：**16 位的源端口字段包含初始化通信的端口号。源端口和源 IP 地址的作用是标识报文的返回地址。
- **目的端口 (Destination Port)：**16 位的目的端口字段定义传输的目的。这个端口指明接收方计算机上的应用程序接口。
- **序列号 (Sequence Number)：**该字段用来标识 TCP 源端设备向目的端设备发送的字节流，它表示在这个报文段中的第一个数据字节。如果将字节流看作在两个应用程序间的单向流动，则 TCP 用序列号对每个字节进行计数。序列号是一个 32 位的数。
- **确认号 (Acknowledgement Number)：**TCP 使用 32 位的确认号字段标识期望收到的下一个段的第一个字节，并声明此前的所有数据都已经正确无误地收到，因此，确认序号应该是上次已成功收到的数据字节序列号加 1。收到确认号的源计算机知道特定的段已经被收到。确认号的字段只在 ACK 标志被设置时才有效。
- **数据偏移 (Data Offset)：**这个 4 位字段包括 TCP 头大小，以 32 位数据结构（字）为单位。
- **保留 (Reserved)：**6 位置 0 的字段。为将来定义新的用途保留。
- **控制位 (Control Bits)：**共 6 位，每 1 位标志可以打开一个控制功能，这六个标志从左至右是 URG (Urgent Pointer field significant, 紧急指针字段标志)、ACK (Acknowledgment field significant, 确认字段标志)、PSH (Push Function, 推功能)、RST (Reset the connection, 重置连接)、SYN (Synchronize sequence numbers, 同步序列号)、FIN (No more data from sender, 数据传送完毕)。
- **窗口 (Window)：**目的主机使用 16 位的窗口字段告诉源主机它期望每次收到的数据的字节数。窗口字段是一个 16 位字段。

- **校验和 (Checksum):** TCP 头包括 16 位的校验和字段用于错误检查。源主机基于部分 IP 头信息、TCP 头和数据内容计算一个校验和，目的主机也要进行相同的计算，如果收到的内容没有错误过，两个计算结果应该完全一样，从而证明数据的有效性。
- **紧急指针 (Urgent Pointer):** 紧急指针字段是一个可选的 16 位指针，指向段内的最后一个字节位置，这个字段只在 URG 标志被设置时才有效。
- **选项 (Options):** 至少 1 字节的可变长字段，标识哪个选项（如果有的话）有效。如果没有选项，这个字节等于 0，说明选项字段的结束。这个字节等于 1 表示无需再有操作；等于 2 表示下四个字节包括源机器的最大段长度（Maximum Segment Size, MSS）。MSS 是数据字段中可包含的最大数据量，源和目的机器要对此达成一致。当一个 TCP 连接建立时，连接的双方都要通告各自的 MSS，协商可以传输的最大段长度。常见的 MSS 有 1024 字节，以太网可达 1460 字节。
- **数据 (Data):** 从技术上讲，它并不是 TCP 头的一部分，但应该了解到，数据字段位于紧急指针和/或选项字段之后，填充字段之前。字段的大小是最大的 MSS，MSS 可以在源和目的机器之间协商。数据段可能比 MSS 小，但却不能比 MSS 大。
- **填充 (Padding):** 这个字段中加入额外的零，以保证 TCP 头是 32 位的整数倍。

### 5.3.3 TCP/UDP 端口号



在 IP 网络中，一个 IP 地址可以唯一地标识一个主机。但一个主机上却可能同时有多个程序访问网络，要标识这些程序，只用 IP 地址就不够了。因此 TCP/UDP 采用端口号 (port number) 来标识这些上层的应用程序，从而使这些程序可以复用网络通道。而为了区分 TCP 和 UDP 协议，IP 用协议号 6 标识 TCP，用协议号 17 标识 UDP。

在实际的端到端通信中，通信的双方实际上是两个应用程序，这两个程序都需要用各自的端口号进行标识。所以，一个通信连接可以用双方的 IP 地址以及双方的端口号来标识，而每一

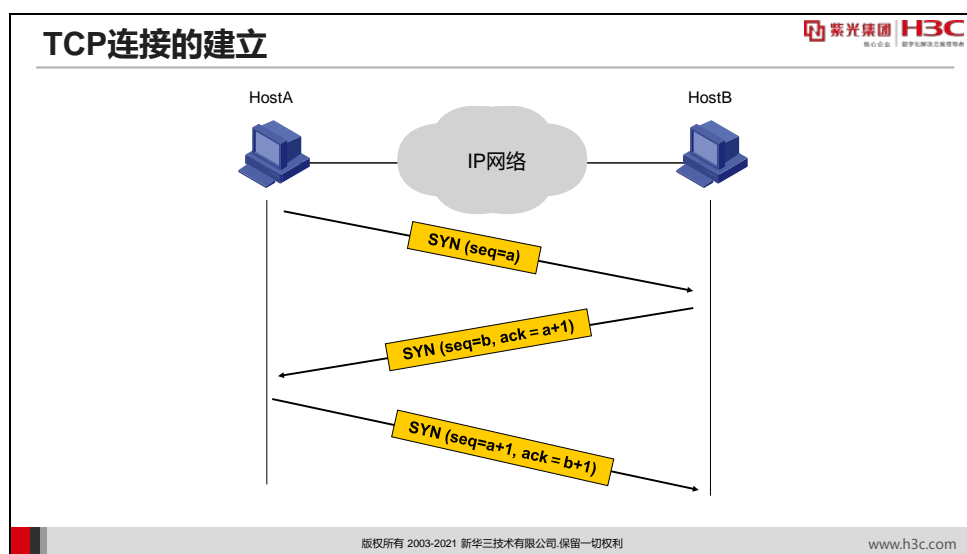
个数据报内也必须包含源 IP 地址、源端口、目的 IP 地址和目的端口。IP 地址在 IP 头中标出，而端口号在 TCP/UDP 头中标出。

TCP/UDP 的端口号是一个 16 位二进制数，即端口号范围可以为 0~65535。其中，端口 0~1023 由 IANA（Internet Assigned Numbers Authority，Internet 号码分配机构）统一管理，分配或保留给众所周知的服务使用，这些端口称为众所周知端口（Well-known port）。大于 1023 的端口号没有统一的管理，可以由应用程序任意使用。详细分配信息可参见 RFC 1700。

保留众所周知端口的必要性显而易见。例如，若 HTTP 服务的端口号是任意值，则用户在访问 Internet 网站时就会遇到麻烦，因为浏览器不知道目的网站所使用的端口号，用户就要自己输入端口号。但是这并不意味着众所周知的协议必须使用众所周知的端口号。例如管理员也可以为 HTTP 协议分配端口 8080，目的恰恰是避免任何人都能随意访问其网页。

### 5.3.4 TCP 连接的建立

TCP 是一个面向连接的可靠的传输控制协议，在每次数据传输之前首先需要建立连接，连接建立成功后才开始传输数据，数据传输结束后还要断开连接。



由于 TCP 使用的网络层协议 IP 只提供不可靠、无连接的传送服务，为确保连接的建立和终止都是可靠的，TCP 使用三次握手（Three-Way Handshake）的方式来建立可靠的连接。TCP 使用报头中的 SYN（Synchronization Segment，同步段）来描述创建一个连接的三次握手消息。另外，握手过程确保 TCP 只有在两端一致同意的情况下，才会打开一个连接。

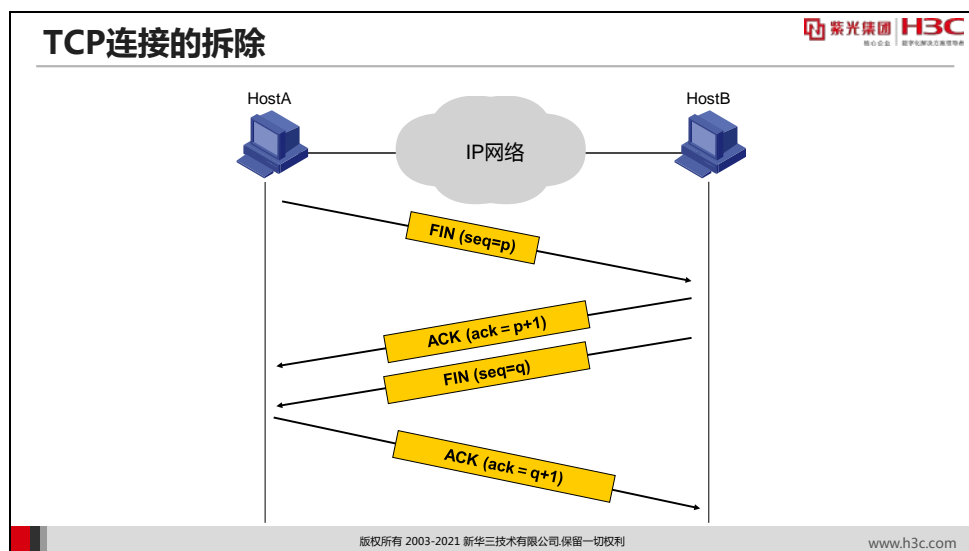
TCP 的三次握手建立连接的过程如下：

- 1) 由发起方 HostA 向被叫方 HostB 发出连接请求。将段的序列号标为 a，SYN 置位。由于是双方发的第一个包，ACK 无效。

- 2) HostB 收到连接请求后，读出序列号为  $a$ ，发送序列号为  $b$  的包，同时将 ACK 置为有效，将确认号置为  $a+1$ ，同时将 SYN 置位。
- 3) HostA 收到 HostB 的连接确认后，对该确认再次作确认。HostA 收到确认号为  $a+1$ 、序列号为  $b$  的包后，发送序列号为  $a+1$ 、确认号为  $b+1$  的段进行确认
- 4) HostB 收到确认报文后，连接建立。

这样，一个双向的 TCP 连接就建立好了，双方可以开始传输数据。

### 5.3.5 TCP 连接的拆除



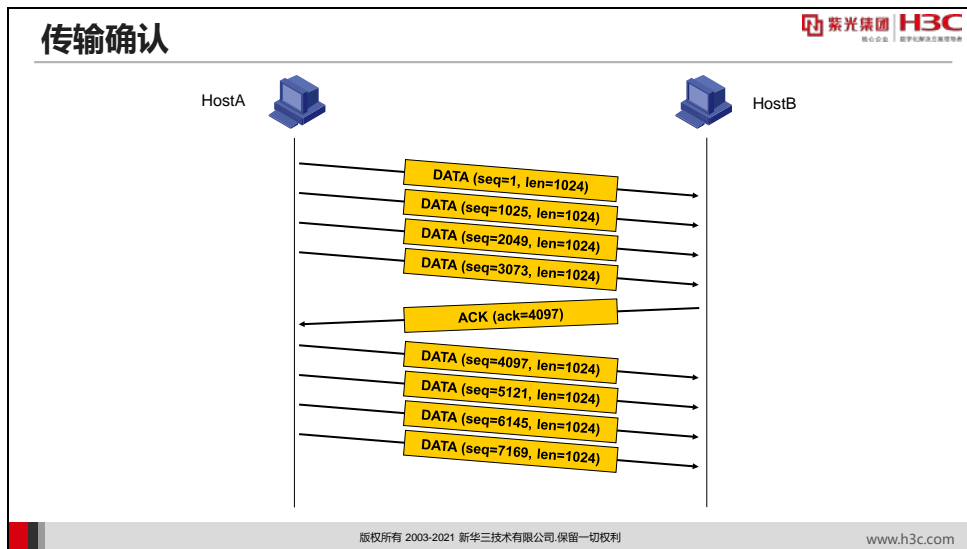
TCP 用 FIN（Finish Segment，结束段）来描述关闭一个连接的消息。

上图所示是一个常规的 TCP 连接终止过程。当数据传输结束后，需要断开连接，其过程描述如下：

- 1) HostA 要求终止连接，发送序列号为  $p$  的段，FIN 置为有效，同时确认此前刚收到的段。
- 2) HostB 收到 HostA 发送的段后，发送 ACK 段，确认号为  $p+1$ ，同时关闭连接。
- 3) HostB 发送序列号为  $q$  的段，FIN 置为有效，通知连接关闭。
- 4) HostA 收到 HostB 发送的段后，发送 ACK 段，确认号为  $q+1$ ，同时关闭连接。

TCP 连接至此终止。可见这是一个四次握手过程。

## 5.3.6 TCP 可靠传输机制



为保证数据传输的可靠性，TCP 要求对传输的数据进行确认。TCP 协议通过序列号和确认号来确保传输的可靠性。每一次传输数据时，TCP 都会标明该段的起始序列号，以便对方确认。在 TCP 协议中并不直接确认收到哪些段，而是通知发送方下一次该发送哪一个段，表示前面的段都已收到。序列号还可以帮助接收方对乱序到达的数据进行排序。

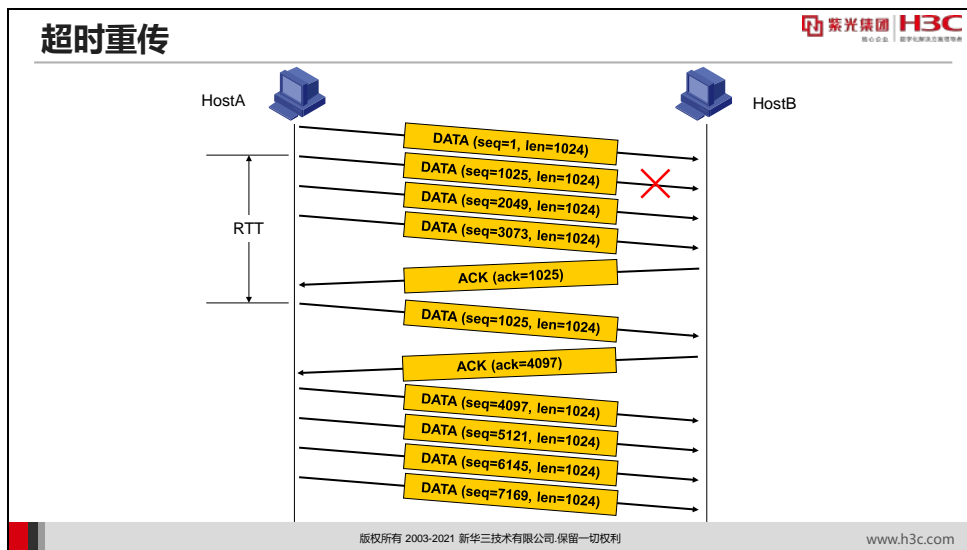
收到一个段确认一个段的方法虽然简单，但是会消耗网络资源较多。为了提高通信效率，TCP 采取了一些提高效率的方法。

首先，TCP 并不要求对每个段一对一地发送确认。接收端可以用一个 ACK 确认之前收到的所有数据。例如，接收到的确认序列号为  $N+1$  时，表示接收方对到  $N$  为止的所有数据全部正确接收。

另外，TCP 并不要求必须单独发送确认，而是允许将确认放在传输给对方的 TCP 数据段中。如果收到一个段后没有段要马上传到对方，TCP 通常会等待一个微小的延时，希望将确认与后续的数据段合并发出。

由于每个段都有唯一的编号，这样当对方收到了重复的段时容易发现，数据段丢失后也容易定位，乱序后也可以重新排列。在动态路由网络中，一些数据包很可能经过不同的路径，因此报文可能会乱序到达。32 位的序列号由接收端计算机用于把段的数据重组成最初形式。

上图给出了一个经过简化的 TCP 传输过程示例。为了便于理解，本例只关注从 HostA 到 HostB 的单向传输。假设 HostA 向 HostB 发送的初始序列号为 1，且发送窗口为 4096 字节，HostA 向 HostB 发送的每个段数据长度为 1024 字节，HostA 将一次性向 HostB 发送 4 个段。而 HostB 收到并校验了数据的正确性后，在回送确认时只需发送确认号  $4096+1=4097$ ，就可以表示 4096 之前的全部数据都已经正确接收，下一次期望接收从 4097 开始的数据。下一次，HostA 仍然一次发送总量为 4096 字节的 4 个段给 HostB。



上图给出了一个经过简化的典型的 TCP 重传过程示例。假设 HostA 向 HostB 发送的序列号为 1025 的第二个段在途中丢失。HostB 只对全部按序无错接收到的序列号最高的段给以确认，即 HostB 只以确认号 1025 向 HostA 确认第一个段已收到。

HostA 在收到这个确认时，并不能确定 HostB 没有收到第二个段，因为也许第二个段也许还没有到达 HostB，或者 HostB 发出的第二个确认可能被延迟了，因此，HostA 不能立即重传第二个段。只有在第二个段发出超过 RTT（Round Trip Time，往返时间）而仍没有收到确认时，HostA 才认为这个段已经丢失，并重传这个段。

HostB 收到重传的第二个段后，按序无误收到的最后一个段的序列号为 3073，因此向 HostA 发送确认号为 4097 的确认，表示之前的数据均正确无误地收到。

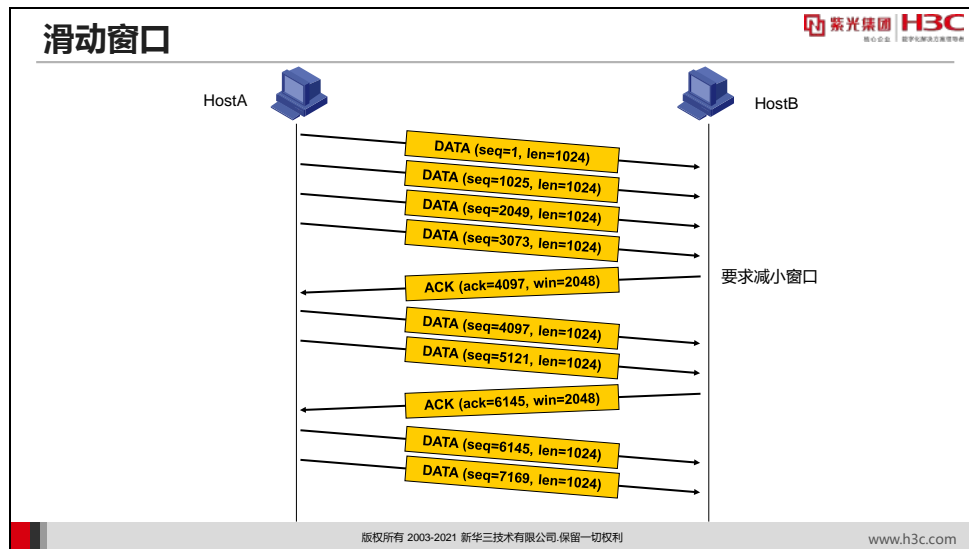
TCP 接收方并不通过“错误通知”告知发送方重传。如果 HostA 向 HostB 发送的序列号为 1025 的第二个段到达了 HostB，但被检查出校验错误，HostB 也不会向 HostA 发送“错误通知”要求重传。因此，RTA 仍然需要等待 RTT 时间之后再重传这个段。

必须考虑的另一种情况是，HostB 回送的确认段也同样可能在传输中丢失或出错。此时对 HostA 来说并不需要额外的机制，因为 HostA 面临的现象与此前的例子是完全相同的——没有收到确认。HostA 仍然用同样的超时重传机制来处理即可。而如果 HostB 回送的确认只是被延迟，则 HostA 在重传后就可能收到两个确认，此时 HostA 只需要忽略多余的确认即可。

因此，RTT 时间就成为一个非常重要的参数。过大的 RTT 导致 TCP 重传非常迟缓，可能会降低传输的速度；过小的 RTT 则会导致 TCP 频繁重传，同样降低资源的使用效率。在实际实现中，TCP 通过实时跟踪发送的段与其相应确认之间的时间间隔来动态调整 RTT 的数值。



## 5.3.7 滑动窗口



TCP 使用大小可变的滑动窗口，并定义了窗口尺寸的通告机制，以增强流量控制的功能。这些机制为 TCP 提供了在终端系统之间调整流量的动态方法。

TCP 滑动窗口尺寸的单位为字节，起始于确认字段指明的值，这个值是接收端正期望一次性接收的字节。窗口尺寸是一个 16 位字段，因而窗口最大为 65 535 字节。在 TCP 的传输过程中，双方通过交换窗口的大小来表达自己的剩余的缓冲区空间，以及下一次能够接受的最大的数据量，避免缓冲区的溢出。

上图仍然通过数据单向发送的简化示例，介绍 TCP 如何通过滑动窗口实现流量控制。

假定初始的发送窗口大小为 4096，每个段的数据为 1024 字节，则 HostA 每次发送 4 个段给 HostB。HostB 正确接收到这些数据后，应该以确认号 4097 进行确认。然而同时，HostB 由于缓存不足或处理能力有限，认为这个发送速度过快，并期望将窗口降低一半。此时 HostB 在回送的确认中将窗口尺寸降低到 2048，要求 HostA 每次只发送 2048 字节。HostA 收到这个确认后，便依照要求降低了发送窗口尺寸，也就降低了发送速度。

若接收方设备要求窗口大小为 0，表明接收方已经接收了全部数据，或者接收方应用程序没有时间读取数据，要求暂停发送。

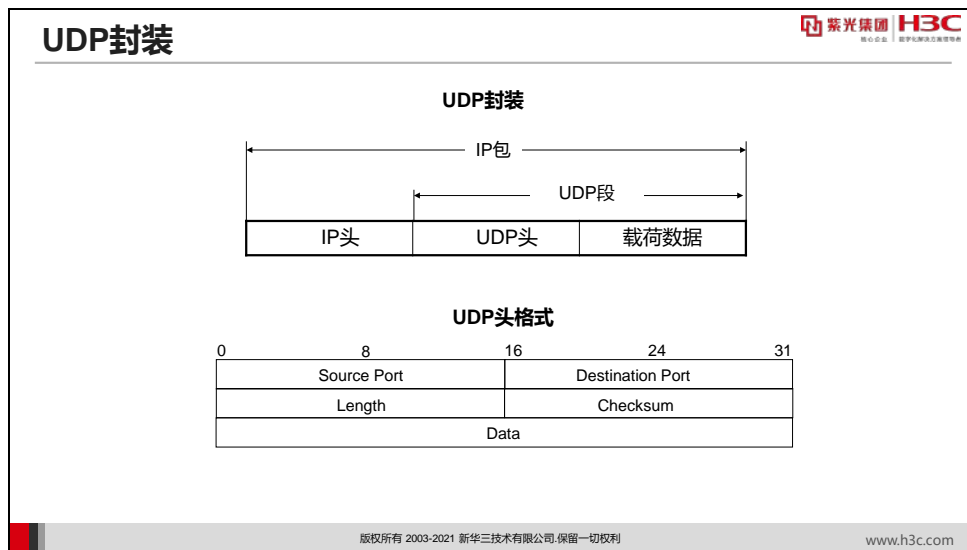
TCP 运行在全双工模式，所以发送者和接收者可能在相同的线路上同时发送数据，但发送的方向相反。这暗示着，每个终端系统对每个 TCP 连接包含两个窗口，一个用于发送，一个用于接收。

可变滑动窗口解决了端到端流量控制问题，但是无法干预网络。如果中间节点，例如路由器被阻塞，则没有任何机制可以通知 TCP。如果特定的 TCP 实现对超时设定和再传输具有抵抗性，则会极度增加网络的拥挤程度。



## 5.4 UDP协议基本原理

### 5.4.1 UDP 封装



RFC 768 定义的 UDP（User Datagram Protocol，用户数据报协议）以 IP 为下层协议。UDP 是为实现数据报（Datagram）模式的分组交换计算机网络通信而设计的。UDP 对应用程序提供了用最简化的机制向网络上的另一个应用程序发送消息的方法。UDP 提供无连接的、不可靠的数据报服务。

由于功能简单，UDP 头相对于 TCP 头简化了很多。UDP 头包含以下字段：

- 源端口（Source Port）：16 位的源端口号，含义与 TCP 相同。
- 目的端口（Destination Port）：16 位的目的端口号，含义与 TCP 相同。
- 长度（Length）：16 位的长度字段，表明包括 UDP 头和数据在内的整个 UDP 数据报的长度，单位为字节。
- 校验和（Checksum）：16 位的错误检查字段，基于部分 IP 头信息、UDP 头和载荷数据的内容计算得到，用于检测传输过程中出现的错误。

## 5.4.2 TCP 与 UDP 的对比

TCP与UDP的对比		
功能项	TCP	UDP
连接服务的类型	面向连接	无连接
维护连接状态	维持端到端的连接状态	不维护连接状态
对应用层数据的封装	对应用层数据进行分段和封装，用端口号标识应用层程序	对来自应用层数据直接封装为数据报，用端口号表示应用层程序
数据传输	通过序列号和应答机制确保可靠传输	不确保可靠传输
流量控制	使用滑动窗口机制控制流量	无流量控制机制


图示为 UDP 与 TCP 的功能对比。

UDP 报文没有序列号、确认、超时重传和滑动窗口，没有任何可靠性保证。因此基于 UDP 的应用和服务通常工作于可靠性较高的网络环境下。

当然，使用 UDP 作为传输层协议也有独特的优势：

- 实现简单，占用资源少：由于抛弃了复杂的机制，不需要维护连接状态，也省却了发送缓存，UDP 协议可以很容易地运行在处理能力低、资源少的节点上。例如，无盘 workstation 在获得 OS 软件之前不可能实现复杂的传输机制，但系统的传递恰恰需要基于传输层协议，这时就可以使用基于 UDP 的 BootP 获取引导信息。
- 带宽浪费小，传输效率高：UDP 头比 TCP 头的尺寸小，而且 UDP 节约了 TCP 用于确认的带宽消耗，因此提高了带宽利用率。
- 延迟小：由于不需要等待确认和超时，也不需要考虑窗口的大小，UDP 发送方可以持续而快速地发送数据。对于很多应用而言，特别是实时应用，重新传输实际上没有意义。例如对 VoIP 来说，如果丢失了一个语音包，通话质量立即会受到影响，但重新传递这个语音包也已经没有必要了，因为通话者不会等重建了语音之后再听。这种情况下 UDP 比 TCP 更加合适。

## 5.5 本章总结



紫光集团 H3C  
核心价值观：数字世界 创造价值

### 课程总结

- TCP和UDP通过端口号标识上层应用和服务
- TCP通过三次握手建立可靠连接
- TCP通过校验和进行差错校验，通过序列号、确认和超时重传机制实现可靠传输，通过滑动窗口实现流量控制
- UDP实现简单，资源占用少，实时性强，适用于可靠性高的网络和延迟敏感的应用

版权所有 2003-2021 新华三技术有限公司.保留一切权利

www.h3c.com